Balkan Journal of Philosophy

ISSN 1313 – 888X

Available online at: http://www.philosophybulgaria.org/BJP/index.htm

http://www.ceeol.com

BALKAN JOURNAL OF PHILOSOPHY

Vol. 2, Issue 2, 2010

CONTENTS

Articles

Book reviews

Information for contributors

IMAGINATION AND THE FORM OF LIFE TO COME*

Nathan Houser

Professor Emeritus of Philosophy
Senior Fellow, Institute for American Thought
Indiana University at Indianapolis
nhouser@iupui.edu

Abstract:
This is a reflective paper on the future of human life in the context of the finitude of our universe. A teleological approach is recommended and imagination is presumed to be the key to ushering in an advanced form of life that incorporates the advances of the new technological age we have begun to inhabit. The amazing possibilities of the coming posthuman age come with a great risk, however: we may be following seductive techno-visions that will bring an end to human life long before its natural demise. Reflecting on life along these lines leads one to a local and pragmatic perspective on human achievement and a more generalized understanding of consciousness.

Key words: C. S. Peirce, cyborg, finitude, form of life, future life, imagination, personhood, teleology.

1 Introduction

How can we see into the future? How are we able to wend our way through the maze of possibilities that lies before us? Imagination lights our way. Of course the journey from the present into the future is fraught with dangers, for we can never see fully ahead, but imagination is our only light and without it the future would be as dark as the blank night before the universe began. Imagination is fueled by experience—by what appears and by what is familiar and known. An important part of our increasingly common experience today is the remarkable advance in technology, especially computer technology, that has taken place during the last half century. Extrapolating from what we know about that advance, we can now begin to imagine possible worlds that promise incredible and wonderful new forms of life and intelligence—and opportunity. Yet some of us hesitate at the thought of what may be lost in the rush toward a radical unheard-of future, and we fear the as-of-yet-unimaginable horrors that our children and our children's children may have to face in consequence of our choices. The techno–visions that now fuel our imaginations are not rooted in the successes and failures of our forebears; they have not been tried by the fire of that analogue of natural selection that undergirds whatever we have of instinct and intuition. Today's visions may be guides to a new and better world, a posthuman word in which the lion lies down not only with the lamb, but with the human and the cyborg—or today's visions may be deadly fancies.

Perhaps, in passing, we should consider briefly whether it is naïve or delusional to suppose that humankind has any future at all, or much of one. Is the form of life to come ours to mould under the guidance of imagination or is it more likely that we have too little time, cosmically speaking, to worry about making our dreams come true? Even if they do come true, for how long will they last? I believe we must admit that the ultimate future of humankind is to have no future. And, of course, that is also the future of the universe.

In more primitive times, when individual lives were imagined to be destined for eternal survival, this would have been an awful thought, one that could only arise from the depths of pessimism or, perhaps, from the sorry hope of an evil or diseased mind. But, today, I suppose that most sane and educated men and woman hold little prospect, nor have much desire, for a history that will never come to an end. Imagination cannot prevent the death of our universe and it is unwise, and I believe unsound, to hold fast to imagined impossibilities.

If I may engage in just a little autobiography, I can tell you that these revelations first came to me during a dramatic personal moment in the mid-60s as I lay pensively regarding a shadeless light bulb hanging from the ceiling of the room I was renting in New York City's Lower East Side. I was reading John Steinbeck's *The Winter of Our Discontent* but had paused to settle my thoughts. Steinbeck had put me in a mood, and when my eyes fell on the barren

* This paper was written for presentation to the Academie du Midi, Alet-les-Bains, France, 12 June 2000.

bulb overhead, somehow it brought to mind the tenuousness of human achievement. Two or three years earlier I had heard Harlow Shapley lecture on the Big Bang theory and was impressed with the idea that eventually there might be a catastrophic implosion of the universe, after a prolonged period of contraction, and that that would be the end of everything that had been. So much for human accomplishment. Even if a new universe were to burst forth, as may have happened many times before, it would truly be new—the starting point for a new history that would be radically discontinuous with the erased history of the preceding age. Our universe would have been "laundered" in an extreme way.

Should the universe have too little mass to reverse its race outward so that it will never implode to erase forever all of its past—and this is what current science endorses—it will not deal any more gently with us. Some relics of dead civilizations might persist for eons upon eons, traveling aimlessly through the vast frigid reaches of a dead and thinning universe, but really that would be of no account. In any meaningful sense, history will have come to an end. When I was first thinking about beginning on this dismal note, I wondered if I should bother to make such obvious remarks and decided that it is important to remind ourselves that the future is finite and that, ultimately, nature is massively beyond our control. The power we have to make history, to be the authors of the future, while remarkable within a local perspective, is next to nothing in the long run. What Shakespeare's Macbeth said about human life applies just as well to all of human history: "Life's but a walking shadow, a poor player that struts and frets his hour upon the stage, and then is heard no more." The complete record of all human accomplishment, all that was and all that will be, is but a tale "full of sound and fury," ultimately "signifying nothing" (*Macbeth* Act 5, Scene 5).

Gloomy as this may seem, we must confront the finitude of human endeavor and achievement. The fiction of eternity should be abandoned—that fiction which is sometimes crudely dangled in front of us like a carrot on a stick, and sometimes, in a more sagely way, is hoisted as if on a banner over our undertakings to remind us that we should not limit our efforts to what can be achieved in one lifetime. Indeed, it is vital that we recognize that the greatest achievements are those that can only be carried out over generations by cooperating individuals whose personal goals are contributory to the achievement of a larger goal that can only be reached communally. I don't doubt that the traditionally religious perspective on human life, that there is a cosmic purpose which makes a mockery of death, has been beneficial for the survival of the species. It has served as a substitute for a more authentic social consciousness and has been a motivator for advancement. But to suppose that our lives and our work should be carried out "under the aspect of eternity" is to court failure. A far more sane perspective, certainly a more pragmatic one, is, I believe, the one I have just indicated. There is a limit to what we, as a people, as well as individually, can do, and to how long we have to get it done, so we cannot wait for evolution to bring about more satisfactory circumstances or for alternate possibilities to work themselves out through the mere action of chance. Intelligent life and thoughtful achievement must be contextualized and carried out within a shorter span. This, I believe, puts a greater demand on imagination and on deliberate action to actualize our plans and dreams.

I suppose it is evident by now that I am using the word "imagination" in a loose sense to encompass the breadth of meanings it has in natural language. Traditionally imagination was supposed to be either reproductive or productive, meaning that it either reproduces past perceptions more or less faithfully and is thus really only memory stimulated into action by some associative process, or else it creatively reassembles elements of experience into new combinations. C. S. Peirce wrote that "when a man desires ardently to know the truth, his first effort will be to imagine what that truth can be. He cannot prosecute his pursuit long without finding that imagination unbridled is sure to carry him off the track. Yet . . . nothing but imagination that can ever supply him an inkling of the truth. He can stare stupidly at phenomena; but in the absence of imagination they will not connect themselves together in any rational way."[1] So, as Peirce would have it, imagination is fundamental even to the proto-abductive underpinnings of perception. Imagination is often understood to be quite distinct from fancy, a mental activity characterized by playfulness with little concern for truthfulness or for its connections to past experiences; but "imagination," as I am using it, includes fancifulness

[1] Charles Sanders Peirce, *Collected Papers of Charles Sanders Peirce*, Volume I, paragraph 46 (Cambridge: Harvard University Press, 1960), p. 20.

and playfulness. We must bear in mind, however, as Peirce indicated, that imagination should not be carelessly trusted.

Since this is a short paper, I have the advantage of being able to beg off giving an in-depth etiological account of the link between imagined outcomes, intentional acts, and future states. This is a good thing for me, because I'm not yet able to deliver a convincing argument for teleology except to the already convinced. Indeed, what I am claiming is that there is, or can be, a sort of final causation at work in the universe—a designed and engineered teleological current to history. According to Peirce, the control we can exert over future events, which are, as we know, only potential, is by means of general dispositions to act in certain ways under circumstances of certain kinds. Some readers will recognize this as Peirce's account of belief. Now, on Peirce's view, belief-habits do not necessarily have to be formed through the repetition of similar acts under similar conditions, but they may be formed in the imagination. Peirce says that "a belief-habit formed in the imagination . . . , as when [we] consider how [we] ought to act under imaginary circumstances, will equally affect [our] real action[s] should those circumstances be realized."[2] Today we recognize this as a kind of conditioning or, better, programming. This may at least provide a suggestion as to how one might begin to develop a causal account of the way imagination helps shape future events and outcomes.

Though believing, as I have confessed, in the finitude of man, of course I admit that, until catastrophe strikes, we do have a future. And I admit that prospects for our future extend far enough ahead to justify hope and perhaps even some mild exuberance. When I consider what this future may be like, I find that I cannot accept as a serious possibility that our form of life will evolve absent the guiding influence of human or extra human purposes. Not only *can* we influence, and to some extent direct, the march of history, of course we *will*. But there is certainly no guarantee that the power we have over the future will be spent to improve life in general, or that the form of life to come will even be an explicit issue of concern. This has rarely been the case in the past and it will doubtless be the same in the future. We will probably just let life develop as it will after the model of a market economy—guided willy-nilly by the invisible hand of competing forces. Probably those competing forces will mainly be multinational corporations or other giant institutions with visions of a future that will keep their coffers full. The form of life to come will then be a sideshow, at best. No doubt this is what we can expect—everything points in that direction. But maybe, just maybe, as the united peoples of the world, we can somehow find our way to a common guiding vision of the future, one that is more concerned with the conditions of life than with the bottom line. Maybe, though, we will be better off without these grand visions. History seems to bear out that grand visions generally result in terrible disasters. In any case, it will be imagination that will illuminate the path, or the various paths that are taken. By putting imagination into action we *make* history and exert real control, either directly or indirectly, over the form of life to come. To a great extent, the future will be shaped by how we conceive it.

How *do* we conceive of the future? Or, more to the point, what are the conceptions of worthwhile ends that we may suppose are likely to guide us into the future? On Sunday, 5 March 2000, while the composition of this paper was underway, I noticed this headline in the *Indianapolis Star:* "Robots get training to work in teams." The subtitle predicted that "Programmed platoons may one day search for distant planets or duel World Cup champions." This is an indication of what we have come to expect and of what we are aiming for. Does any informed person really doubt the eventual success of the artificial intelligence research program that began in the imaginations of a few philosophers and computer scientists? Do you doubt that in the not too distant future there will be environment-sensitive robots with the computer analogues of central nervous systems—robots that can in some rather meaningful way *really* communicate with humans as well as with fellow robots? At any rate, this is what we have imagined and have set our sights on, and we have already come a long way toward bringing this castle in the air down to earth.

Are robots, then, one of the forms that life will take in the future, or do you think that robots are too mechanical to count as life? What if robots can be made to become independent? To be able to learn and perhaps even to make plans? What if robots can be taught to repair themselves by making use of resources in their environments? What if robots can learn to reproduce themselves by constructing new robots following their own patterns, or perhaps more

[2] *Ibid.*, Volume 2, paragraph 148, p. 79.

creatively, or at least with intended or even random variations for the sake of difference? Would they then be living? What if robots were constructed with flesh-like bodies to have not only artificial intelligence but also artificial metabolisms? Would that make the crucial difference? Perhaps not. But, if not, I believe we should expect that one day there will be life-like processes that will serve all of the same purposes as life itself, and probably, in many cases, more efficiently and effectively. And I predict that eventually we will extend our conception of life to include such processes.

One reason for believing that our conception of life will have to widen is a change in the other direction that is under way and gaining momentum—what might be called the mechanization of the human body. We have begun to fill our bodies with mechanical parts and with regulating computers, and even with artificial organs. There is every reason to believe that this practice will accelerate and that the idea of the Six Million Dollar Man will become commonplace. Irony of ironies: as robots become more like humans, we become more like robots. If we cannot say that robots, as I imagine them, are alive, because they are too mechanical, must we say that at some point in the mechanization of the human body the person ceases to live—even though he or she continues to function in all of the relevant ways? Does a person whose brain survives alone in the proverbial vat continue to live? If so, then it must not matter how mechanical the rest of a body becomes as long as the brain survives and continues to function. But what if the brain is enhanced with computer memory chips; perhaps with some added programming? Surely this will happen—perhaps it already has on a small scale. So you see, the form of life to come must not be supposed to be simply the ways of naturally evolved creatures but something broader that includes the ways of extended, mechanically and computationally enhanced, creatively designed, and even altogether artificial life-like beings.

Let me get just a bit more fanciful with my speculations, although I believe not unduly so. I think we can assume that future persons will be sentient embodied minds, but I don't think we can stipulate what sort of embodiments they will have, except to say that their embodiments will have a sufficient degree of durability, complexity, and so forth. Their minds will have to be self-programmable and modifiable and have a capacity for development (we might say growth). There will obviously have to be a pretty large capacity for information processing and for operational control of the relevant embodiments. Some future persons will have mechanical (must we say artificial) bodies and thus any kind of "reproduction" would essentially amount to program duplication, although I suppose some mechanically embodied minds might want to replicate their line or model of embodiment. For such persons there would be no reproductive reason to differentiate sexes so, presumably, there would be no gender and no artificial sexual apparatus; but I suppose something of that sort could be added were it desired for esthetic or recreational purposes.

There is a lot of excitement today over the prospect of being able to download human memory into computing devices which can then be accessed directly via a datajack implanted in one's skull. But not just one's own memory; might we not some day be able to have the same kind of direct access to stored information from external sources—information that derives from the experiences and knowledge of others. What's more, why couldn't we switch circuits between brains so that you can see with someone else's eyes, hear with their ears, feel with their body, and have all of their sensations? And they yours. Perhaps two, or maybe three or four, of us could feel simultaneously with each other's bodies.

This is a tiny sampling of what I believe to be our agenda for the future. These are some of the ideas that are motivating the powerful among us—those whose dreams we inevitably enter into and help fulfill. What if this all comes to pass, as I believe it will? Will these remarkable changes in our form of life, in what we count as a person, in how we access and transmit information, and so on—will these profound changes radically alter our senses of ourselves? Is our individuality seriously threatened? Although it may seem so, and may even be so, I am not so sure.

I used to have the honor and the pleasure of working next door to the office of the great Peirce scholar, Max H. Fisch, a man born in the final year of the 19th century. Sometimes I would walk into Professor Fisch's office to ask what he thought about some pressing matter having to do with the Indianapolis edition of Peirce's writings. Usually, on those occasions, he would look up at me, reflect for a moment, and then say, "Let me see." Then he would push back from his desk, rise from his chair, and walk to his filing cabinets and begin searching through files. Finally he would announce what he thought about the question at hand—he had found what he thought in his files. To remember, in this case, meant to look it up in his files.

I do not see any fundamental difference, beyond that having to do with the medium, between Professor Fisch's procedure and my plugging a computer module into my brain to access my stored memory. Nor do I see a great difference in kind between access to the experiences and knowledge of others through the books they write or the lectures they give and having the same information digitally transmitted directly into one's brain without the mediation of our usual sensory mechanisms. Of course the semiosis that connects the transmitted information (signs) with the received information (interpretants) in these different cases will be enacted by remarkably unlike means, but the logic of it, or the semiotic of it, is the same. You might think that the calculating power of a computer will so remarkably supplement an attached brain that the resulting "cooperative" reasoning will be of a wholly new order, or that super servo-sensory mechanisms will enable us to perceive in incredibly new ways, but are these possibilities really so distinct from the way our present thought interacts with calculators and computing devices, though less directly, or the way we enhance our perceptual powers by using reading glasses, or telescopes, or video cameras, or other instruments designed to let us perceive what we cannot perceive with our unaided senses?

Well, you might say, perhaps these technological developments that I predict do not offer such radically new kinds of thinking and perceiving after all. Maybe even cyborgs are not wholly new creatures; perhaps in a sense Max Fisch was a cyborg in that some of his memory was stored in data banks outside of his body—data banks he could access more or less on demand. But surely the likelihood that we will soon be living in a world with artificially intelligent machines, really conscious machines, will be a whole new ball game. What will *that* mean for humankind? Maybe I'm just stubborn, but frankly I cannot even accept on its face that even that would be such a different world. Those of you who are reading this paper are intelligent beings external to me and I cannot see why, logically speaking, I would find a room filled with intelligent machines (who might also read this paper) so radically different from a room filled with my human readers. Sure, there is the matter of common experiences, of similar ways of life, and so on. But in a room filled with human mortgage brokers I would probably feel far less comfortable, far less at home, than in a room full of robot philosophers.

It is not that I believe that nothing remarkable is happening; on the contrary, I think we may be on the threshold of a radically new world—maybe even a truly posthuman world. It would not surprise me if the computer revolution has a more profound impact on our form of life than the combined changes wrought by the cartographic revolution, the Protestant reformation, and the cosmological revolution as those great events moved us into the modern age. But it is important to recognize that today's prophets of the new great paradigm shift are to some extent selling us old goods, however unrecognizable they may be in their new garb. Let's face it, the computer revolution has come about in part because of the duplicity and gullibility of CEOs, politicians, university presidents, library directors, and so forth, and because so many of us are ready at a moment's notice to jump on the latest bandwagon. The cost of moving into what is now called "the information age" is staggering, both in money and time, and it seems that for every step forward there is a corresponding step backwards. For example, as university libraries manage to put more of their resources on line, large segments of their budgets have to be diverted to computer infrastructure and software acquisition, and the research and development of new information resources takes a back seat. Information resources, however important, are coming to be ignored if they are not available on line. The entire management structures of organization after organization have been reconceived and revamped, often at astonishing cost, to accommodate the capabilities of computer programs. Often there is no longer any choice in the matter—the modern way of running an organization, of doing business, requires that we buy into "the information age"; by now, this goes without saying. Every manager and director who is forced by these circumstances to shift limited resources to computer infrastructure, endless software upgrades and licenses, archiving systems, salaries for data managers, systems analysts, and so on, can tell you that the costs have been astronomical. We have paid and are paying dearly to change our world to accommodate the visions of Steve Jobs, Bill Gates, *et al.* But there is now no turning back. We have swallowed the bait, the hook has been set, and we are being reeled in by the masters of the future.

Fortunately, even though the price is high, the possibilities, as I have already indicated, are amazing. Are we, perhaps, attracted by the idea of real group consciousness, a sort of symphony of feeling—a feat to be achieved, I suppose, by interconnecting the brains of a group of individuals in the right way? Surely this will be possible. Perhaps we would like to have telepathic powers, at any rate something far nearer the real thing than we can now achieve by means of the telephone or email (which are, in themselves, pretty amazing as media for the almost instant transmission of thoughts across many miles from mind to mind)? I believe this could be brought about by training our brains to use implanted transmitters to emit appropriate signals which are automatically sent by satellite to designated receptor implants in targeted brains. Maybe you would like to integrate your thought processes with the inner workings of a computer by interfacing your brain with the computer? Probably this will come to pass. Instead of doing data searches and financial calculations by the old method of typing information into a computer system and reading results off a terminal screen, you could somehow attach the computer to your brain and access the computer directly. Perhaps your consciousness would extend to the computer itself. Maybe while you were attached you and your computer would *feel* as though you were one, a single-minded self, however temporary. The possibilities are only limited by our imaginations and the physical resources we have to work with. What we choose from our stock of imaginings will set our course into the future and determine the form of life to come.

What's wrong with this picture? I don't mean to be asking how I've gone wrong in characterizing the little I have of what can be achieved through computer and biotechnologies. If I've gone wrong here and there, I doubt that I've gone far wrong; and my guess is that far more remarkable and even astonishing developments are on the horizon than any I've previewed. What is wrong with the picture I've given is that the future I've tried to look into and reveal, even if not very attractive to some of us, is still too good to be true. For one thing, I may have led you to believe that I think that everyone will benefit more or less equally from the achievements to come. That's very doubtful. John Dewey, for one, imagined a democratic world in which everyone would be well educated and a participant in public affairs. The great socialists have envisioned a world in which everyone shares more or less equally in material goods. These visions have had a great impact, and continue to, but they certainly have not won the day. Those who through their wealth or position have great power are generally not very supportive of visions of equality. The greatest benefits will, as always, go to the most advantaged in society. New life forms will probably have the worst of it. I would guess that there will be great pressure to deny personhood to robots, no matter how lifelike and capable they become, in part because of their potential for cheap labor or even servitude.

A second thing that is wrong with my picture of the future is that I have ignored the great risk of disaster. It is commonplace that great successes often can be achieved only at the risk of great failures. I believe that is precisely our situation today. As I indicated at the start, the imaginings that are now leading us into the future derive from experiences and situations that are generally too new to have been tested either by the unforgiving demands of natural selection or even by the quicker test of pragmatic workability. It is likely that the secret of long-lived ecosystems is historical continuity. Open futures are attractive and probably essential for growth and progress but developments that do not derive from past successes are as likely as not to be deadly. When imagination works from past experiences, when the future grows out of the soil of the past, as it were, there is a certain control over the path taken by virtue of selection processes. But when history moves too quickly and imagination builds its castles, not from the building blocks of tested experience, but from the promises of new technologies and mere wishes, the light of imagination dims and we travel more blindly into the future. And when one of the constituents of that future is sure to be manufactured self-replicating systems with some level of intelligence, we are indeed rolling the dice. If we can beat the odds our success may be astounding. But if the fruits of our imagination are such that either we or they must be selected out of existence, then we may bring about our annihilation ahead of schedule.

Whatever happens, we are surely in for a marvelous ride. And there should be ample opportunity for rapid philosophical advances in consequence of our widespread experimentation with intelligent systems. The philosophy of mind has been an active field for many years, yet our understanding of mind and consciousness leaves much to be desired. We may learn much in the coming years as we observe how consciousness arises in different kinds of sensory and information-processing systems and how consciousness changes in amalgamated intelligent systems (such as with group consciousness or with human-computer combinations)—and perhaps also as we notice the loss of self-consciousness as individual stable selves become less common. The nature of individuals and persons should take on a new interest. The extension of personhood to techno-devices and mechanisms should focus attention on integrated, interrelated systems and information networks instead of on the physically separate or on psychological histories. Peirce was adamant that we are in mind and that mind is not in us, and it is possible that such a conception of mind will gain respectability in consequence of reflections on the variety and flexibility of intelligent systems in the future. Philosophy of life is bound to be in for a great renewal. And clearly there will be difficult work ahead for philosophers who specialize in the theory of rationality or in ethics. It would be well to prepare for the future by directing even more philosophy students into ethics (I ignore the question of university employment) for there will be serious issues to contend with—like attempts to overcome the present inefficient and unpredictable system for programming new minds (i.e. schools) with more invasive programming methods (i.e. modular implants), and perhaps a new try at eugenics as we learn more about how to manipulate the human gene? We will have to be able to answer the questions: Why shouldn't we manipulate gene strings to customizing new life? This is a vital question that calls for hard philosophical thought. And what about privacy? It seems inevitable that what we count as public will grow larger and larger and that the desire for privacy may one day be viewed as an old-fashioned phobia. Old fashioned though it may be, I rue the day. Yet I do see some advantages to living in full view. Pretense and deception, both of others and of one's self, would probably become less prevalent, and in some ways it might be a relief to be an open book. But quite clearly there may be dire psychological and legal consequences to the loss of privacy. These and many other concerns will furnish ethics with an abundance of difficult and urgent problems in years to come. As for esthetics, I will merely point out that our entire future depends on the choices we make as to ends.

I have pointed out already that not many, if any, of these issues had to wait for the information age to make their appearance. Max Fisch's filing cabinets or even Charles Peirce's pen (for he said that sometimes he could not think without it) might have raised many of the same problems. But there is something about the pervasiveness (or invasiveness) of, and even the idea of, the computer revolution that gives urgency to these philosophical questions. In recent years, any number of philosophers from the field of Artificial Intelligence have tried to bring these issues into focus by constructing elaborate science fiction examples. Putnam and Dennett, for instance, have given us nice stories to help illustrate and resolve problems with our conception of mind, self, and personhood. Interest among philosophers soon flagged, but new circumstances will rekindle it. Consider the case of Cousin Bob, who is furious that Aunt Emma left her fortune to Jim the Robot. It does not matter to Cousin Bob that Jim is more intelligent, that he has a kinder disposition, or that Aunt Emma loved Jim as she never loved him. Nor does it matter that Jim really needs the money. All that matters to Cousin Bob is that Jim is a "machine" while he is flesh and blood—and the closest blood relative. You can be sure that Cousin Bob will mount a legal challenge to Aunt Emma's will, and the courts will need expert witnesses. I think it will only be after life has evolved to a point where the techno-achievements I have previewed become commonplace, or until it is not unusual to confront ethical, legal, and practical issues of import arising from these developments, that philosophers will be compelled to sort out these issues and to work hard for consensus. Only then will interest become widespread outside of philosophy. This is what seems to have happened in biomedical ethics.

As I worked on this paper, I often thought of Sarte's famous dictum: "existence precedes essence." In one sense it might seem that I am claiming the reverse. We imagine possibilities, essences of a sort, and then do what we can to actualize them. But the crucial point is that *we* imagine the possibilities that *will become* our actualities; *we* decide what we will become, what kind of world we will live in. Not that we can have whatever we want. Of course there are limitations and unexpected consequences—possibly disastrous ones. But the form of life to come will be a consequence of choices we make from the futures we can imagine—or, more accurately, from the choices *someone* makes from the futures *they* imagine.

As I said near the beginning of my talk, it is the larger-than-life goals, the plans and dreams that can only be achieved communally over the course of generations, that reveal so well our transcendence from a brute state of nature. But all too often we have fallen into lock-step behind dreamers whose grand visions have no strong bridges of continuity to the past, nor the authorization of rigorous esthetic investigation or of well-considered ethical judgment. Under those conditions, imagination can only see dimly into the future and it becomes an unsafe guide. I believe that is the state of things today and that the course we are taking into the future may translate us into forms of life we never asked for and should never have hoped for.

On the other hand, we may have good luck, and for the little time that will still be ours, our coming form of life may be the grandest achievement ever to grace the cosmos. Ultimately, though, no one will ever know

References

Charles Sanders Peirce (1960) *Collected Papers of Charles Sanders Peirce*, Volume I, Volume 2. Cambridge: Harvard University Press.

BALKAN JOURNAL OF PHILOSOPHY. Vol. 2, Issue 2, 2010

THE LOGIC OF INTENTIONAL CONSTRUCTIONS: ITS PHILOSOPHICAL SIGNIFICANCE

Paul Gochet

University of Liège
pgochet@ulg.ac.be

Abstract:
The logic of intentional verbs and constructions has been intensively studied over the last twenty years in different logical fields : pure logic, philosophy of logic, logic for AI and logic computer science. There is little interaction among these scientific communities. In this paper I shall bring together the more significant results which have been obtained and offer an up to date status quaestionis. I will reduce logical technicalities to a minimum. and lay emphasis on what is philosophically and ethically significant in the recent formalizations of intentional verb and adverbs.

Key words: intentionality, closure, implication, semantics, *stit* logic, neighbourhood.

1. A characteristic feature of intentional verbs: the lack of closure under logical implication.

In *Towards Non-Being The logic and Metaphysics of Intentionality*, Graham Priest claims that generally intentional verbs are not closed under *logical implication*. He provides a formal semantics which accounts for this lack of closure. Priest's semantics illuminates the logical behavior of typical intentional verbs like «believe, desire, intend to». It should however be supplemented by a different semantics for grammatical constructions which also fall under the concept of intentional construction such as «see to it that.». I shall briefly compare the two semantics and move on to an inquiry into the lack of closure of intentional verbs under *causal implication* and *believed causal implication* as these are other anomalies that have to be reckoned with if we aim at giving a general characterization of the logic of intentional verbs. Let us first discuss the lack of closure under logical implication. An operator Θ is closed under logical implication if it satisfies the following schema:

$\models \mathrm{A} \supset \mathrm{B}$

$\models \Theta\mathrm{A} \supset \Theta\mathrm{B}$

The *modal operator* «necessarily» ($\Box$) fulfils this condition.

Here is an example:

$\models (p \supset r) \supset ((p \wedge q) \supset r$

$\models \Box\ (p \supset r) \supset \Box\ ((p \wedge q) \supset r)$

Intentional operators which Priest represents by $t\ \Psi$ do not. In other words we can have

$\models \mathrm{A} \supset \mathrm{B}$

and not $\models t\ \Psi\ \mathrm{A} \supset t\Psi\ \mathrm{B}$.

Here is Priest's example

If I eat my cake, I do not have my cake anymore.

But it does not follow that

If I desire to eat my cake, I desire not to have my cake any more. People, Priest says, want to eat their cake and have it.

One might object that other inferences of the same form as Priest's example seem to display closure under logical implication. Here is a tentative example of closure.

If I die painlessly, then I die

Hence if I desire to die painlessly, I desire to die.

Admittedly the example becomes questionable because of contextual and pragmatic accretions. It may be the case that I desire to die painlessly if death is imminent and unavoidable without for that matter desiring to die. I will no try to settle this issue here. As Quine observed in his *Elementary logic*, ascribing a logical form to an inference in natural language is not a matter of calculation but of translation. Priest circumvents the translation problem by providing an abstract example which will be examined later.

2. A model theory for pure intentionality

The interpretation of the modal operator in possible worlds semantics has become standard since Kripke's paper [Kripke 1963]. It consists in a clause which gives the truthconditions of «necessarily A» evaluated at world w in terms of those of «A» evaluated at world w'. related to w by the accessibility relation R. The gist of this method is lucidly rendered by Hintikka in this passage: «Putting the main point very briefly and somewhat crudely by stepping from a world to its alternatives, we can reduce the truth-conditions of modal statements to truth-conditions of nonmodal statements [Hintikka 1973, 198].

The interpretation of the intentional operator $t\ \Psi$ follows the same pattern. The main difference lies in the model used. Besides *possible worlds* (in which the actual world is enclosed), Priest's model accommodates *logically impossible worlds* (in which *some* molecular formulas can be treated as atoms) and *open worlds* with respect to which all molecular formulas can be treated as atoms. All these worlds share the same non-empty domain of individuals D. Moreover the model involves an accessibility relation R^{Ψ} which is defined over the disjoint union of the three kinds of worlds just mentioned.

The interpretation of the formula $t\ \Psi$ A is given by the following clause:

«$t\ \Psi$ A» is true in w if and only if for all $w' \in W$ such that $w\ R^{\Psi}\ w$, «A» is true in w'

We are now ready to examine the *abstract example* designed by Priest to show the lack of closure of the intentional operator $t\ \Psi$ under logical implication. The example reads as follows:

$\models (Pa \wedge Qa) \supset Pa$ but $\not\models t\ \Psi\ Pa \wedge Qa) \supset t\ \Psi\ Pa$

Formula $(Pa \wedge Qa) \supset Pa$ is true by hypothesis. We have to find an interpretation $\Im$ in which $t\ \Psi$ (Pa $\wedge$ Qa) is true and $t\ \Psi$ Pa is false. For that purpose, Priest builds up a relational structure containing one closed world only, namely @ (the actual world) and one open world only, namely w. The accessibility relation leads from @ to w and nowhere else.

$$@ \ \bullet \rightarrow \bullet \ w$$

Figure 1

In virtue of the clause defining the intentional operator $t\ \Psi$, showing that $t\ \Psi$ (Pa $\wedge$ Qa) is true in @, boils down to showing that (Pa $\wedge$ Qa) is true in w. Analogously showing that $t\ \Psi$ Pa is false in @ reduces to showing that Pa is false in w.

Since w is an *open world*, formula (Pa $\wedge$ Qa) can be treated as *atomic*. Let us take advantage of this licence and treat (Pa $\wedge$ Qa) as if it has the form Rab and give it an interpretation which makes it true. For that purpose, $Px_1 \wedge Qx_2$ (i.e to Rx_1x_2) will be assigned the extension D×D and singular terms «a» and «b» will be assigned individuals which are members of D. Under this interpretation $t\ \Psi$ (Pa $\wedge$ Qa) is true.

Finally we have to give an interpretation to Pa which makes it false in the open worlds. Priest assigns the extension $\varnothing$ to Px_1 and assigns individuals of the domain D to «a». Clearly, under that interpretation, Pa is false.

One might be tempted to object that once the formula (Pa $\wedge$ Qa) is treated as atomic , it ceases to qualify as the antecedent of $\models (Pa \wedge Qa) \supset Pa$. This objection however is misguided. Priest does not change the *syntactic form* of the formulas under consideration *but assigns them a non compositional semantics* which makes it possible to construct the countermodel he needs. [I owe this crucial point to private communication with Prof. Shahid Rahman].

However *contrived* Priest's formal semantics may be due to its non-compositionality, it fully succeeds in establishing the lack of closure of the intentional operator $t\ \Psi$ under logical implication. Priest should not be blamed for the abstract nature of his example. He has set himself the task of characterizing *pure* intentionality, defined *a priori* as failing to be closed under logical implication.

We have focused on Priest's model theory for intentionality and said nothing about his proof theory. There is however a proof theory for that logic [Priest 2001] but, as expected, it is relatively poor as Priest himself recognized.

Priest's model theory for intentionality is original and deep, but it does not fit all cases of lack of closure under logical implication. Intriguing instances of such a lack of closure have been discovered by logicians and philosophers exploring the logic of agency (*stit logic*). These happen to be highly philosophically significant. We shall examine them in the next section.

3. A concrete example of lack of closure under logical implication

In *Facing the future, Agents and Choices in Our Indeterminist World*, Nuel Belnap, Michael Perloff and Min Xu offer the following example of lack of closure under logical implication: «If there is at least one injured man who is bandaged, then there is a least one injured man» does not entail «If *you see to it that* there is at least on injured man who is bandaged, then *you see to it that* there is at least one injured man». Formally this failing inference reads as follows :

$\models (A \wedge B) \supset A$ but $\not\models Stit\ (A \wedge B) \supset Stit\ A$

The above-mentioned authors stress that this is neither a paradoxical nor funny logic that is at work here, nor a grammatical subtlety or a grammatical curiosity, but something which is both «typical» and «deeply built into the real-choice-based idea». [Belnap *et al.* 2001, 40]

Most philosophers dealing with ontology worry about *what there is* [Quine 1960] Belnap and his co-authors are concerned about *what people do* [Antoniol 1998]. They contributed both a model theory and a proof theory for the *stit* logic which will prove to be closely linked to the logic of intentionality.

Here again, a model theory has to be built which explains the lack of closure of the operator *stit* over logical implication. Before delving into a model theory for *stit* logic, let me recall the natural rendering of the *stit* operator, or rather that of the *astit* operator (achievement *stit* operator). For Belnap and his co-authors, [α *astit*: A] means «the present momentary fact that A is guaranteed by a prior choice of the agent α» [Belnap Ibid. 33].

To capture the meaning of the *achievement stit*, they built a *frame*, i.e. a tuple, $\langle A, T, I, \leq, C\rangle$ whose first element is a set of *agents* thought of as making choices in time. The second element is a *tree*, whose branches represent the different courses of events (called histories) which can realize or not, depending on the choices made by that the agents. The third one is a set of *instants*. The fourth one is the *relation* «earlier than or simultaneous with» defined over I. An instant represents a set of contemporaneous moments from each of the various histories. The fifth one is the function *choice*. It maps each agent α and moment m into a partition of the histories $H_{(m)}$ through m. The agent has to choose between equivalence classes of histories. Belnap and his co-authors claim that the structures exhibited by the semantics of *stit* "are not just mathematical curiosities, but describe – up to a legitimate idealization – the world in which agents act" [Belnap *et al. Ibid.*36].

To turn the frame into a *model*, we need an interpretation function $\Im$ mapping each atomic formula into a set of *m*/*h* pairs (pairs whose first term is a moment and whose second term is the history to which the moment belongs).

To obtain a semantics, we still need to provide a recursive definition of truth. The novelty lies in the clause defining the truth-conditions for the formula *stit A* evaluated at a given pair *m*/*h*. This clause involves two requirements (1) the truth of *A* is guaranteed now as a result of a previous choice of the agent (positive requirement), (2) the truth of *A* was not already settled at the previous moment.

The model sketched above and the two requirements just stated *explain why* the

inference quoted in the beginning of this section breaks down. Requirement 2 has been violated. One cannot bring about that there is an injured man at time t+1 if the man was already injured at time t. Let us observe that the explanation of the lack of closure under logical implication requires a very peculiar temporal structure. This casts some doubt on Priest's claim that lack of closure under logical implication is a general characteristic of intentional constructions.

4. Logical implication, causal implication and believed causal implication

If it is debatable whether typical intentional constructions fail to be closed under *logical implication*, it is unquestionable that they fail to be closed under *causal implication* and *believed causal implication*, a fact that has been established about twenty years ago by logicians belonging to the A. I. Community.

Let me first recall the main difference between logical and causal implication. Logical implication, as we understand it here, is another name for a truth-functionnally valid conditional [Quine 1982] i.e. for a formula that is valid in virtue of its truth-functional components. The formula below exhibits that kind of validity:

$$(p \supset r) \supset ((p \wedge q) \supset r)$$

Yet the inference below is obviously not valid [I thank Joost Joosten and Frank Veltman for pointing this problem to me]:

If I absorb poison I die, Hence if I absorb poison and absorb counter poison I die.

The problem arises because we have implicitly construed the two causal conditionals occurring in the premise and in the conclusion as truth-functional conditionals. But this is fallacy of equivocation. Causal conditionals are not truth-functional. As opposed to truthfunctional conditionals, causal conditionals are *not monotonic*. To avoid confusion, we shall express the *causal conditional* by «inevitable $\varphi \supset \psi$».

It is still incumbent upon us to define *believed causal implication*. Following common practice, I shall say that *believed causal implication* is simply causal implication prefixed by a belief operator: BEL (inevitable $\varphi \supset \psi$).

In a famous paper published in *Artificial Intelligence*, Philippe Cohen and Hector Levesque produced an example which shows that the intentional verb «intends to» is not closed under *believed causal implication*. The example has become part of the folklore of the subject and we quote it *in extenso*: «[…] an agent intended to have his teeth filled. Not knowing about anaesthetic (one could assume this took place just when they were being first used in dentistry), he believed that it was always the case that if one's teeth are filled, one will feel pain. One could even say that surely the agent *chose* to undergo pain. Nevertheless one would not like to say that he intended to undergo pain [Cohen & Levesque 1990, 251]».

5.The BDI architecture

Cohen and Levesque's formalism is that of first-order logic enriched with modal operators. Yet beliefs, goals and intentions are still treated as *predicate constants*. A year later however, two logicians of the AI Community, Anand Rao and Michael Georgeff worked out a formalism called the *Belief-Desire-Intention architecture* set up for the study of the interrelations between these three intentional notions.

As opposed to Cohen and Levesque's formalism, Rao and Georgeff's formalism is a fully fledged first-order modal logic embodying the temporal logic CTL*. Beliefs, goals and intentions become modal operators (respectively «BEL», «GOAL» and «INTEND»)

Instead of treating each possible world as a *sequence of events*, Rao and Georgeff consider each possible world to be a *time tree*. They interpret modal operators in terms of accessibility relations: $\mathbf{R_B}$ for «BEL», $\mathbf{R_G}$ for «GOAL», $\mathbf{R_I}$ for «INTEND». The treatment of belief is standard: a formula is said to be believed at time *t* by an agent if and only if it is true in all belief-accessible worlds. [Gochet & Gribomont 2006], Rao and Gerogeff follow the same pattern to provide a semantics for «GOAL». As they put it, «the goal-accessibility relation specifies the situations that the agent *desires* to be in [...]. An agent has a goal φ at time *t* if and only if φ is true in all the goals-accessible worlds of the agent at time *t* .[Rao & Georgeff 1991, 477]».

Rao and Georgeff spell out axioms which define the relationships between the intentional operators. Let us focus on the first one: *For every belief-accessible world, there must be a goal-accessible world.* To borrow the author's example, if the agent believes that three courses of events can be realized (1 visiting dentist A, having one's teeth filled and suffering pain, 2 go shopping, not having one's teeth filled and not suffering pain, 3 visiting dentist B, having one's teeth filled and suffering pain), he or she may select the first and the third as goals.

The converse of the italicized statement does not hold As Rao and Georgeff notice, even if the agent believes that certain facts are inevitable, he or she is not forced to adopt them as goals. Hence goals are not closed under the beliefs of the agent .The following state of affairs can occur:

GOAL $\varphi \wedge$ BEL (inevitable $\varphi \supset \psi$) $\wedge \neg$ GOAL ψ

Analogous reasoning holds for intentions with respect to beliefs.

6. A combination of BDI architecture and *stit* logic

Several proposals have been made for combining BDI architecture with *stit* logic. I shall first examine Caroline Semmling and Heinrich Wansing's work and show that what they say about the intentional verb «desire» holds for the deontic operator «it is obligatory that». This will support some ethical conclusions about the act of causing collateral damages.

Semmling and Wansing have bought out an important distinguishing feature of the intentional operator «desires» which had passed unnoticed. They have observed that although desires cannot be *incoherent* – *pace* Priest who claims that people want to eat their cake and have it – desires can be *conflicting*. I both *desire* to donate one of my kidneys to save my brother, and *not desire* to donate it as it will affect my physical integrity. Yet I do not entertain the incoherent desire of *donating* and *not donating* one kidney.

To capture this situation we need a modal logic in which ($\Box$ p $\wedge$ $\Box$ q) $\supset$ $\Box$ (p $\wedge$ q) does not hold. Such a logic exists, but it require of us that we switch from Kripke's *relational semantics* for modal logic to Montague Scott's *neighbourhood semantics* [Chellas 1980, 210, Straetmans1991] Semmling and Wansing take advantage of this possibility. They adopt a different semantics for belief (relational semantics) and desire (neighboorhood semantics).

Besides *desires*, *duties* can conflict. .Here is an example due to the casuist Raymond of Penaford quoted by Frédéric Sart: «The casuist imagines a householder who is asked by people bent on taking the life of someone hiding in the house whether she is in. Replying «Yes» is obligatory by virtue of the duty to tell the truth, but at the same time it is forbidden by virtue of the duty to save a life [Sart, 2009, 132]».

Suppose we adopt the law ($\Box$ p $\wedge$ $\Box$ q) $\supset$ $\Box$ (p $\wedge$ q). We can reason as follow: If in a given situation it is obligatory to tell the truth and forbidden to tell the truth (i.e. obligatory to lie), then it is obligatory to tell the truth and to lie. But telling the truth and lying is impossible. Since ought implies can, doing what is impossible (telling the truth and lying) is not obligatory. By contraposition, it is permitted either to lie (and violate the duty to tell the truth) or to tell the truth (and tviolate the duty of saving a life). According to that reasoning, if we comply with one of the two conflicting duties, whichever we chose, we are morally safe.

This is a very superficial ethical response. Considering *conflicts of duties*, Nicolai Hartmann takes the opposite view in his *Ethik* . He holds that whatever we do in such a case we are guilty. As he puts it «The alternative is here not between wrong and right, but between wrong and wrong. Either way a value is violated and, again, either way a value is fulfilled. Whoever stands in such a predicament– and life continually places man in such – cannot escape without offence [Hartmann 1926, transl. 1932, 1950, 301].

To block the chain of inferences which leads to the counterintuitive ethical conclusion denounced by Hartmann, we need again to adopt a semantics which deprives the modal formula quoted above from its validity. This can be done for free by using again the neighbourhood semantics that was adopted to cope with the logic of desire.

7. A combination of *stit* logic with epistemic logic

I shall draw this paper to a close by briefly examining Jan's Broersen's recent findings in the logical behaviour of the intentional constructions «knowingly doing» and «intentionally doing». Up to now we have first presented and discussed lack of closure of intentional constructions under *logical implication*. Next we examined lack of closure under *believed causal implication*. Now we shall exhibit a case of lack of closure under *causal implication*.

Broersen spells out a formal language which contains the following constructions besides the usual connectives:

The operator $\Box\varphi$ expresses *historical necessity*.

The operator $X\varphi$ expresses that *in the next state* φ.

The operator [*A xstit*] φ means agents *A* jointly *see to it that* in the next state φ

The operator $Ka\varphi$ is the standard *knowledge operator*.

The operator [*a xint*] φ expresses that agent *a* performs action that will be realized in the next state. It can be read: «agent a *intentionally* performs the action bringing about the state of affairs φ».

A frame is provided in which the *domain* is a set of dynamic states, i.e. tuples $\langle s,h \rangle$ where *s* is a member of a set of states (*S*), h is a member of a set of histories (*H*) and *s* is a member of *h*. *Accessibility relations* are used to interpret the modal operators and an axiomatic system (correct and complete) is given. I will not dwell on these technical details. I will focus on the philosophical significance of the new formal semantics put forward by Broersen.

Intentions are no longer treated as *mental states*. They are treated as *modes of action* and this will contribute to explain why intentions are not closed under side effects of action. Moreover Broersen produces a new operator by combining the *K*-operator with the *stit* operator: the resulting operator *Ka*[*a xstit*] φ . means that agent *a knowingly* performs the action which brings about the state of affairs φ.

The following relationship holds: intentionally doing φ logically implies knowingly doing φ but not conversely. As the author observes, an agent killing in self-defense kills knowingly, but does not kill intentionally.

Knowingly doing is *closed under causal implication*, but *intentionally doing* is not, as the following examples borrowed from the author shows. If an agent *knowingly* stabs another person *x* out of self defense and if stabbing *x causally implies* killing *x*, then the agent *knowingly* kills *x*. Yet from the fact that the agent *intentionally* stabs *x* out of self-defense and that stabbing *x causally implies* killing *x*, it does not follow that the agent *intentionally* kills *x*.

By exploiting this crucial difference between *knowingly doing* and *intentionally doing*, a satisfactory solution to the classical dentist' problem can be given. Brorsen sums it up nicely in the following passage: «[…] (1) the agent intentionally visits the dentist, (2) knowingly visits the dentist in a way that causes him pain, and (3) does not know a way of visiting the dentist without having pain» .From these premises, we can derive that the agent knowingly sees to it that he has pain, but we cannot derive that the agent intentionally sees to it that he has pain [Broersen, forthcoming].

I emphasize that a satisfactory solution to the dentist problem requires of us that we take into account not only *logical implication* but also *causal implication* (*believed or not*) when we explore the peculiarities of intentional constructions.

Let me mention that the combination of *Stit logic* with *epistemic logic* has shed light on the logic of the intentional verb «regret» in a paper due to Emiliano Lorini and François Schwartzentruber [Lorini and Schartzentruber, forthcoming]. This problem however falls outside of the theme of this essay and we shall not discuss it here.

References

Antoniol Lucie (1998). *Things People Do* PhD. Stirling University

Belnap Nuel, Perloff Michael, Xu Ming (2001) O.U.P. *Facing the Future*, Oxford

Blackburn Patrick, de Rijke Maarten, Venema Yde (2001). *Modal Logic*, Cambridge, C.U.P.

Broersen Jan (2009) «A complete STIT Logic for Knowledge and Action, and Some of Its Applications», M.Baldoni et al (Eds) *Lecture Notes in Artificial Intelligence* 5397, 47–59

Broersen Jan «An xstit-Logic Analysis of Intentional Action» forthcoming in *The Journal of Philosophical Logic*.

Chellas Brian (1980). *Modal Logic, an Introduction*, Cambridge University Press;

Cohen Philippe & Levesque Hector (1990). «Intension Is Choice With Commitment» *Artificial Intelligence* 42, 213–261.

Emiliano Lorini and François Schwarzentruber (forthcoming) «A Logic for Reasoning about Counterfactual Emotions »

Gochet Paul et Gribomont Pascal (2006) «Epistemic Logic» Dov Gabbay & John Woods (Eds) *Handbook of the History of Logic*. Vol. 7, Elsevier Amsterdam, 99–195.

Hartmann Nicolai (1926), transl. 1932, 1950, Ethics, vol. 1, George Allen and Unwin, London.

Hintikka Jaakko (1973) «Grammar and Logic: Some Borderline Problems» in J. Hintikka, J. M. E. Moravcsik, and P. Suppes, *Approaches to Natural Language*, Dordrecht, D.Reidel,197–214

Kripke Saul (1963, 1971) "Semantical Considerations on Modal Logic" in Leonard Linsky (Ed.) *Reference and Modality*, Oxford University Press, 62–72.

Priest Graham (2001). *An Introduction to Non-Classical Logic*, Cambridge, Cambridge University Press.

Priest Graham (2005). *Towards Non-Being, The logic and metaphysics of Intentionality*, Oxford, O.U.P.

Priest Graham (2008). *An Introduction to Non-Classical Logic*, second extended edition Cambridge, Cambridge University Press.

Quine Willard Van Orman (1941, 1965). *Elementary logic*, Harper, Harvard.

Quine Willard Van Orman (1950). 4th edition 1982, *Methods of Logic*. Harvard University Press, Cambridge, Mass.

Quine Willard Van Orman (1960). *Word and Object*, MIT Press, Cambridge Mass.

Rao Anand S. & Georgeff Michael P. (1991). «Modeling Rational Agents within a B.D.I. Architecture» in James Allen, Richard Fikes, Erik Sandewall Eds. *Principles of Knowledge, Representation and Reasoning*. San Mateo Morgan Kaufman, 473–484.

Sart Frédéric (2009). "A purely combinatorial Approach to Deontic Logic" *Logique et Analyse*, 206, 131–138.

Semmling Caroline et Wansing Heinrich (1998). «From BDI AND stit TO bdi-stit LOGIC» *Logic and Logical Philosophy*, 17, 189–211.

Semmling Caroline and Wansing Heinrich (2009). «A sound and complete system of bdi-stit logic» in M.Perlis (Ed.) *Logical Yearbook* 2008, College Publications, London, 193–210.

Straetmans Marina (1991) *Logiques Modales Non Normales et Logiques Modales Non Monotones*, M.A. thesis, Liège.

Acknowledgments

Earlier versions of this paper were read at Prof.Dr.Shahid Rahman's Seminar (University of Lille 3), at Prof.Dr. Angelo Nepomuceno's Seminar (University of Sevilla) and at the «Séminaire sur l'Intentionnalité» organized by my colleagues Denis Seron and Arnaud Dewalque in the Department of Philosophy of the University of Liège. thank them all for their comments, criticisms and suggestions for improvement. I also thank researchers and senior researchers who helped me a great deal :Jan Broersen, Hans van Ditmarsch, Andreas Herzig, Joost Joosten. I also benefited from corrrections due to Pascal Gribomont and Manuel Rebuschi. Parts of his essay will come out in French in the electronic journal *Bulletin d'Analyse Phénoménologique*. I thank the editors Denis Seron and Arnaud Dewalque for giving me permission to use them here in a modified form.

BALKAN JOURNAL OF PHILOSOPHY. Vol. 2, Issue 2, 2010

HILBERT'S PROGRAM: THE TRANSCENDENTAL ROOTS OF MATHEMATICAL KNOWLEDGE

Rosen Lutskanov

Institute for Investigation of Societies and Knowledge, Bulgarian Academy of Sciences, Sofia, Bulgaria
rosen.lutskanov@gmail.com

Abstract:

The design of the following paper is to establish an interpretative link between Kant's transcendental philosophy and Hilbert's foundational program. Through a regressive reading of Kant's Critique of Pure Reason (1781), we can see the motivation of his philosophical project as bound with the task to expose the a priori presuppositions which are the grounds for the possibility of actual knowledge claims. Moreover, according to him the sole justification for such procedure is the (informal) proof of consistency and (architectonical) completeness. Hilbert tried to strip Kant's philosophy of its last anthropomorphic vestiges which led to the formulation of his "finite standpoint" and the proof-theoretical methods for axiomatic reconstruction of classical mathematics. Therefore, contrary to the received view, the proofs of consistency and completeness which were envisaged as part of his metamathematical program were not conceived as a means to secure to epistemic basis of mathematical knowledge. Accordingly, the program itself was not confuted by Gödel's theorems and remains as viable as ever.

Key words: philosophy of mathematics, transcendental philosophy, Hilbert's program.

Today, nearly a century after its original formulation, the question about the viability of Hilbert's foundational program seems settled once and for all. Almost all researchers in the field of foundational studies have acknowledged that Gödel's incompleteness theorems delivered a deadly blow in its very heart (Simpson 1988: 349) or even "refuted it" (Sieg 1988: 338). Recently, the reign of consensus was broken by Michael Detlefsen who debated the verdict. According to his formal reconstruction, Hilbert did not endorse the requirement that the ideal (transfinite) theory is to conserve extension of its finite part and this obviously makes the incompleteness theorems much less devastating (Raatikainen 2003: 162). His treatment of Hilbert's program was questioned, so the issue remains unsettled. That's why I propose a different way to tackle the question: I'll try to show that in fact Hilbert's program was not and could not be disproved by the incompleteness theorems, because its aim was not to lay once and for all the unshakable foundations of mathematical knowledge but to expose the transcendental presuppositions that were at the basis in the mathematical practice of his day. To this end, I intend to accomplish three interconnected tasks: (1) to display the influence of Kantian philosophy on Hilbert's mental development; (2) to fix the requisite definition of "transcendental method" as set forth by Kant in his "Critique of Pure Reason"; (3) to show that in fact Hilbert's program was modeled after this philosophical paradigm.

First of all, I would like to stress the point that such comparison is only seemingly artificial, mostly because of the peculiar character of the German educational system and the unusual role that mathematics played in it at the turn of the twentieth century. As Felix Klein noted in his Evanston lectures in 1893, while in France mathematics became applied science due to the reform that led to the establishing of Ecole polytechnique, in Germany it remained a "pure academic discipline closely connected with the theoretical physics via the widespread neo-Kantian philosophy" (Klein 1894: 100). Moreover, in the context of the Prussian "neo-humanistic" educational reform, mathematics was awkwardly classified (along with philology and history) as a science of antiquity (*Altertumwissenschaft*): the reason for this decision was probably that it comprises truths (as the theorems of Euclidean geometry) that were established in the distant past (Hoyrup 2000: 150–151). This makes comprehensible the fact that in the end of nineteenth century in Germany the mathematical departments were still part of the philosophical faculties with all dreadful consequences of this fact: in Königsberg, for example, there were no special lecture halls and equipment for instruction in mathematics, the mathematical library was housed in the university's detention quarters and had only twenty books at its disposal (Rowe 2003: 48). As a matter of fact, Hilbert's academic career started in Königsberg and we may justifiably maintain that despite this terrible setting he appreciated Kant's teachings because we are told that he "selected as one of the two theses which doctoral candidates were required to defend in public the proposition that the objections to Kant's theory of the a priori nature of arithmetical judgments are unfounded" (Shanker 1988: 246–7). About forty years later, his career ended in Königsberg where he delivered the famous radio-lecture in which he cited with approval Kant's dictum that "in any particular natural science one encounters genuine scientific substance only to the extent that mathematics is present" (Vinnikov 1999: 43).

Moving from the biographical to the conceptual, we may find that the aforementioned thesis is echoed in Hilbert's own contention that "everything whatsoever, that can be the object of scientific thought is subject, as soon as it is ripe for the formation of a theory, to the axiomatic method and thereby to mathematics" (Bernays 2003: 10). Generally speaking, almost everywhere in his writings we can locate the vestiges of Kant's influence: for example, in the statement that "we must make the concept of specific mathematical proof itself the object of an investigation, just as the astronomer takes into account the motion of his location, the physicist concerns himself with the theory of his apparatus, and the philosopher criticizes reason itself" (Bernays 2003: 14), or in the suggestive way his assistant Bernays epitomized their joint epoch-making foundational program: "the task falling to metamathematics vis-à-vis the system of mathematics is analogous to the one which Kant ascribed to the critique of reason vis-à-vis the system of philosophy" (Bernays 2004: 38). Finally, it is appropriate to mention here the fact that Hilbert defined as early as 1899 the task of mathematics, in completely Kantian fashion, as "logical analysis of our faculty of intuition (*Anschauungsvermögen*)" (Corry 2006: 141). This shows clearly that the transcendental background was already there when the fundamentals of his program were conceived.

If we admit that the link between Kant's critical philosophy and Hilbert's axiomatic mathematics may be established by such historical considerations, we have to locate the particular interpretation of Kant's legacy that suits our current task. Here I'll rely upon the recently published book *Interpreting Kant's Critiques* (2003), in which Karl Ameriks challenged the so-called "received interpretation" of Kant's transcendental deduction of the categories. He insists that its common construal as an argument "basically aiming to establish objectivity, i.e. to prove that there is an external and at least partially lawful world, a set of items distinct from one's awareness, and to do this from the minimal premise that one is self-conscious" (Ameriks 2003: 54) is inherently flawed. According to his reading of the relevant passages of the "Transcendental Analytic," "the overwhelming evidence is that Kant did not put forth the transcendental deduction as having the aim … to give the sufficient conditions of empirical knowledge or a proof that there is an objective world. What Kant says he is doing is providing a deduction of the categories, a proof of their objective validity, which, since they are pure concepts, can be done only by showing their a priori applicability to experience, i.e. by showing that they are part of the (necessary) conditions for the possibility of experience, where 'experience' is defined as empirical knowledge" (Ameriks 2003: 58–9). In fact, this specific rendering of the transcendental deduction was formulated much earlier in Norman Kemp Smith's *Commentary on Kant's Critique of Pure Reason* (1918) where it is asserted that "though the proof of the a priori is not empirical in the sense of being inductive, neither is it logical in the sense of being deduced from necessities of thought. Its "transcendental" proof can be executed only so long as experience is granted as actual" (Kemp Smith 1918: xxxv-xxxvi). According to Ameriks, this is equivalent to the contention that Kant's *Critique of Pure Reason* must be read not progressively (as descending from pure a priori principles to the possibility of empirical knowledge) but regressively (as ascending from empirical knowledge to the principles that make it possible). As is well known, this particular distinction was introduced by Kant himself, who in his *Prolegomena* noted that the theoretical position of critical philosophy must be prepared by means of the analytical method which "signifies only that we start from what is sought, as if it were given, and ascend to the only conditions under which it is possible. In this method we often use nothing but synthetical propositions, as in mathematical analysis, and it were better to term it the regressive method, in contradistinction to the synthetic or progressive" (Kant 1902: §4). As was noted in the same passage, Kant elaborates the synthetic or progressive method in his *Critique of Pure Reason* but this does not imply that there he intends to accomplish the opposite: namely, to deduce the objectivity of factual knowledge from the structure of the pure understanding which determines the conditions for its possibility.

But the preconditions for the possibility are obviously relative: their significance is correlative to the object whose possibility they are determining. This relative character of the a priori principles is clearly detectable in one remark at the end of the section of "Transcendental Analytic" named "A refutation of idealism" which reads as follows: "that more than one all-embracing experience is possible, cannot be inferred from what is given; and still less can any such inference be drawn independently of anything being given since without material [*Stoff*] nothing whatsoever can be thought. What is possible only under conditions which themselves are merely possible is not in all respects possible" (Kant 1929: A232–3; B 284–5). In a nutshell, this means that the regressive reading of the transcendental deduction establishes the relative character of the a priori principles. Due to Volker Peckhaus we have at our disposal significant indications that this particular interpretation of Kantian philosophy was familiar to Hilbert. In his paper "Über den Satz von der Gleichheit der Basiswinkel im gleichschenkligen Dreieck" ("On the Theorem of the Equality of the Base Angles in the Equilateral Triangle") Hilbert defined the axiomatic method as follows: "I understand under the axiomatic exploration of a mathematical truth [or theorem] an investigation which does not aim at finding new or more general theorems being connected to this truth, but to determine the position of this theorem within the system of known truths in such a way that it can be clearly said which conditions are necessary and sufficient for giving a foundation of this truth" (Peckhaus 2000b: 10). In other words, at the turn of the century "Hilbert defined the axiomatic method as a procedure of finding for a given proposition the premises from which it follows" (Peckhaus 2000b: 5). Almost two decades later, in his lecture course "Nature and Mathematical Knowledge" (1919/1920) he stressed that what was called before "regressive method" "finds its perfect expression in what is called today "axiomatic method." This is a general method of scientific research as such" (Peckhaus 2000b: 11). In the meantime, this correlation between the regressive justification of knowledge claims and the axiomatic exposition of mathematical theories was fully developed by the neo-Kantian Leonard Nelson, who declared himself "Hilbert's pupil" and in his article "Über das sogenannte Erkenntnisproblem" (1908) identified the axiomatic method with the de-compositional procedure that Kant examined in the already adduced passage from the fourth section of his *Prolegomena* (Peckhaus 2000b: 11). According to Peckhaus, this juxtaposition makes perfectly visible the parallelism between Nelson's "critical philosophy centered on the regressive method" and "Hilbert's axiomatic method of discovering mathematical axioms" (Peckhaus 2000a: 11–2).

Now I come to the main point: the particular way in which Kant's transformation of metaphysics into rigorous science was mirrored in Hilbert's foundational program. Here the correspondence is more than obvious: in the preface to the first edition of his Critique Kant characterized the metaphysics of his day as a "battle-field of endless controversies" (Kant 1929: A viii) and tried to relieve her from this bitter fate. To this end, he tried to re-establish it as a science which according to him is the same as to reconstruct it in the form of an architectonically structured system: as far as "systematic unity is what first raises ordinary knowledge to the rank of science, that is, makes a system out of a mere aggregate of knowledge, architectonic is the doctrine of the scientific in our knowledge" (Kant 1929: A832, B860). In Kant's own words, the system of knowledge is like "an animal body," it is "organized unity, not an aggregate" which grows from within, not from without (Kant 1929: ibid). Accordingly, metaphysics must be replaced by transcendental philosophy which is the "idea of a science, for which the critique of pure reason has to lay down the complete architectonic plan. That is to say, it has to guarantee, as following from principles, the completeness and certainty of the structure in all its parts" (Kant 1929: A13, B27).

Ironically, a whole century later Hilbert found himself in pretty much the same situation and offered precisely the same solution: because of the rise of non-Euclidean geometries, the controversies about the foundations of analysis and (last but not least) the discovery of the paradoxes in Cantorian set theory, mathematics fell into unprecedented disrepute. The quarrel around the new geometries even brought into existence the term "meta-geometry" (with a built-in analogy to "meta-physics") which stigmatized the systems of Riemann and von Helmholz as creations of "mathematical mysticism" (Grattan-Guinness 2000: 121). That is why at the turn of the century among the mathematicians was wide spread the belief that the recent developments show only the inevitability of a reform that would expel such "unscientific theories" and re-establish mathematics as the queen of the sciences (Shanker 1988: 187). Correspondingly, the primary task of Hilbert's program was to reconstruct axiomatically classical mathematics since the axiomatic method has been proved to be the proper tool to remove present disagreements. This very idea was expressed by Carlo Cellucci in his article "The Growth of Mathematical Knowledge" where it is said that Kant "supports a hologenetic view of mathematical theories according to which, just like a seed from the very beginning contains the plant that will grow out of it, with all its differentiations, the principle of mathematical theory will contain all the subsequent developments of the theory. This … closed world view is further developed by Hilbert, and indeed through him it has become the official viewpoint of mathematical logic" (Cellucci 2000: 4). This interpretation is supported by the text of Hilbert's fabulous Paris lecture, in which he reiterates the Kantian organic metaphor: "the mathematical science is indivisible whole, an organism whose vitality is based on the mutual connections of its parts" (Hilbert 1902: 478). But according to Kant the characteristic trait of an organized unity is its completeness, the fact that it "grows from within," which means that we are free to impose this requirement on any axiomatic presentation of a body of knowledge. This is the basis of the so-called "axiom of completeness" (*Vollständigkeitsaxiom*) which was introduced by Hilbert in the following way: "the elements of geometry form a system that can not be expanded if the axioms are to remain valid" (Zach 2004: 326). In other words, the axiomatic explication of the science of geometry is possible precisely because and only if its objects form an organic whole which is complete in itself.

But, as Kant demonstrated in his *Critique*, the formal requirement for systematic unification may be met only if we restrict our claims for knowledge to objects that are capable of presentation in the forms of sense-intuition (*Anschauung*). Accordingly, Hilbert focuses his investigations on such mathematical operations which may be represented by manipulations of sense-objects which we treat as signifying the corresponding mathematical objects. This is not an epistemological innovation at all: he just replaced "Kant's a priori intuitions of space and time, which he viewed as so much 'anthropological garbage' with a single intuition which was taken to provide a framework of shapes or forms in which our experience of concrete signs was embedded" (Detlefsen 2005: 117). In perfect coordination with this line of argument, Hilbert adduces the Kantian thesis that mathematics cannot be reduced to logic because it has fixed content that is determined by experience: the precondition for the application of logic and the accomplishment of logical operations is that something is given in our sense intuition (namely, the signs of the formula-language of classical mathematics). This conception forces Hilbert to face one serious obstacle: due to the work of Cantor, mathematics was gradually transformed into a science of the trans-finite; but the trans-finite is supposed to be incapable of sensuous presentation; so, after all, how is it possible for mathematics to become a real science?

In his own words, Hilbert's answer sounds something like this: the infinite (which allegedly can not be effectively represented by sense-objects) must be reduced to an "ideal element" or idea if, following Kant, we mean by this a notion of reason which transcends experience. This vague suggestion is explicated by Stephan Körner who points that Hilbert's

> way of reconciling concrete, finite mathematics with the abstract and transfinite theory of Cantor's is something Hilbert again owes to Kant. It was not, indeed, in the philosophy of mathematics that Kant employed the principle on which Hilbert's reconciliation proceeds. Kant employed it in a part of philosophy which for him was much more important - the reconciliation of moral freedom and religious faith with natural necessity. Arguing in this context, Kant first pointed out that the notion of moral freedom (and some other notions, including that of actual infinity) were Ideas of Reason which were unrelated to perception … He then argued that any system containing notions applicable primarily to concrete objects (such as mathematics and physics of his day) could indeed be amplified by Ideas, but only provided the amplified system could be shown to be consistent. Proving consistency … was Kant's way as he himself put it of making room for faith. In quite similar fashion Hilbert distinguishes between the concrete or real propositions of finite mathematics and the ideal notions (Ideas) of transfi-

> nite mathematics. In order to justify the adjunction of Ideal notions to the real, he too requires a proof that the system is consistent (Körner 1971: 73–4).

This means that Hilbert's quest for consistency proof was not intended to provide epistemic justification for classical mathematics but only to outline the ontological demarcation between "real" and "ideal" objects of mathematical knowledge which is part of the legacy of Kant's transcendental philosophy.

At the bedrock of his Kantian approach to mathematical knowledge Hilbert placed the so-called "axiom of thought" which was introduced in his lecture course "Logische Principien des mathematisches Denkens" (1905) in the following way: "I have the ability to think of things and designate them by simple symbols (a, b, … , X, Y , …) in such a perfectly characteristic way, that I can always recognize them again unambiguously; my thought operates with these things under this symbolization in certain ways according to definite laws and I am able by introspection (*Selbstbeobachtung*) to know and completely describe these laws" (Webb 1997: 3). This axiom, to which he also refers as "the axiom of the existence of intelligence" and "the apriori of the philosophers," is the fundamental warranty for the feasibility of this "enumeration of the rules which in fact govern our thinking" which exhausts the fundamental idea of Hilbert's proof theory (*Beweistheorie*). Accordingly, he views his own formalist epistemology as stripping of the transcendental method from its last "anthropomorphic vestiges" (Detlefsen 2003: 324).

According to Hilbert's point of view, the process of thinking is effected "in the consecutive production of propositions" which receive verbal or written expression and, respectively, the whole conceptual content of the mathematics is manifested in "a game with non-interpreted formulae." This interpretation of mathematical knowledge is deeply rooted in Hilbert's modification of transcendental philosophy which implies that

> the general idea of the Kantian epistemology is concerned with the determination of the fundamental a priori and intuitive foundations of the conceptual knowledge which are examined in view of the conditions for their possibility. I contend that this task is accomplished in my investigations related to the principles of mathematics. They show that the a priori is nothing more than a fundamental presupposition (*Grundeinstellung*) or expression of some unconditioned conditions (*unerläßliche Vorbedingungen*) of thinking and experience (Detlefsen 2003: 325).

This passage from Hilbert's "The grounding of elementary number theory" (1931), is consistent with the idea that the mentioned "unconditioned condition" is not constructability in the pure forms of sense-intuition but formal expressibility in the language of classical mathematics. The justification for this contention was reconstructed by Michael Detlefsen who shows that the formalizability condition guarantees the applicability of public standards in the assessment of individual knowledge claims (Detlefsen 2003: 330). In short, if universal validity is the characteristic mark of a priori knowledge and communicability by the already adopted language is the only way the universal validity may be recognized, then a priori valid are precisely those elementary combinatory facts that are expressible in our language and accordingly are capable to be intersubjectively acknowledged (Detlefsen 2003: 330).

As far as Hilbert's program does not pretend to answer the question "what is mathematics?" but the related question "how is the existing mathematics possible as a science?" it simply cannot be viewed as a doctrine whose factual claims are disproved by Gödel's incompleteness theorems. Gödel has shown that the classical mathematics cannot be presented as a complete axiomatic system which is formalizable in the language of classical mathematics and is provably conservative in respect to its properly finitistic fragment. This means, that Hilbert's program is inadequate if we interpret it as a foundational enterprise whose sole aim is to provide classical mathematics with secure epistemic foundations. But this has nothing to do with Hilbert's original idea which as we have seen was to analyze what we mean when we grant to classical mathematics the status of a priori synthetic knowledge. As Volker Peckhaus has put it, if Hilbert has at all "any idea of final foundation (*letzt begründung*), in the background, it has only a heuristic function, that is to say, it served as a regulative idea in the Kantian sense" (Peckhaus 2000a: 8).

In my opinion, the moral of this story is that without a profound study of the history of philosophy we may be unable to locate the precise meaning of the fundamental scientific theories which form the invisible background of present-day scientific practice and to assess impartially their vices and virtues. When we remove the present-day theoretical demarcations which prohibit any comparison between Kant (the philosopher) and Hilbert (the mathematician) we can see the development of his metamathematical program as one of the most exciting events in the history of post-Kantian philosophy.

References

Ameriks, K. (2003). *Interpreting Kant's Critiques*. New York: Oxford University Press.

Bernays, P. (2004). *The philosophy of mathematics and Hilbert's proof theory*. (The Bernays Project: www.phil.cmu.edu/projects/bernays)

_____. (2003). *Hilbert's significance for the philosophy of mathematics*. (The Bernays Project: www.phil.cmu.edu/projects/bernays)

Cellucci, C. (2000). The growth of mathematical knowledge: an open world view. In: Grosholz, E. et al (eds.) *The growth of mathematical knowledge*. Dordrecht: Kluwer Academic Publishers.

Corry, L. (2006). Axiomatics, empiricism, and Anschauung in Hilbert's conception of geometry: between arithmetic and general relativity. In: Ferreiros, J. et al. (eds). *The architecture of modern mathematics*. Oxford: Oxford University Press.

Detlefsen, M. (2005). Philosophy of mathematics in the twentieth century. In: Shanker, St. (ed.) *Philosophy of Science, Logic and Mathematics in the Twentieth Century*. London: Routledge.

_____. (2003). Constructive Existence Claims. In: Schirn, M. (ed.) *The philosophy of mathematics today*. Oxford: Oxford University Press.

Grattan-Guinness, I. (2000). *The Search for Mathematical Roots*. New Jersey: Princeton University Press.

Hilbert, D. (1902). *Mathematical Problems*. Bulletin of the AMS, 8.

Hoyrup, J. (2000). *Human sciences: reappraising the humanities through history and philosophy*. Albany: State University of New York Press.

Kant, I. (1902). *Prolegomena to any future metaphysics*. London: Cambridge University Press.

_____. (1929). *Critique of Pure Reason*. London: Macmillan and Co.

Kemp Smith, Norman. 1918. *A commentary to Kant's Critique of Pure Reason*. London: Macmillan & Co.

Klein, F. (1894). *Lectures on Mathematics*. London: Macmillan & Co.

Körner, St. (1971). *The philosophy of mathematics*. London: Hutchinson.

Peckhaus, V. (2000a). *The Pragmatism of Hilbert's Programme*. (Lecture delivered at the "History of Logic" conference at the University of Helsinki in June 2000)

_____. (2000b). *Regressive analysis*. (Paper presented on June 5, 2000, at the Bled Conference in Philosophy on Philosophical Analysis)

Raatikainen, P. (2003). Hilbert's program revisited. *Synthese*, 137.

Rowe, D. (2003). From Königsberg to Göttingen: a sketch of Hilbert's early career. *Mathematical Intelligencer*, 25: 2.

Sieg, W. (1988). Hilbert's program sixty years later. *Journal of Symbolic Logic*, 53: 2.

Simpson, St. (1988). Partial realizations of Hilbert's program. *Journal of Symbolic Logic*, 53: 2.

Shanker, St. (1988). Wittgenstein's remarks on the significance of Gödel's theorem. In. Shanker, St. (ed.) *Gödel's theorem in focus*. London: Croom Helm.

Vinnikov, V. (1999). We shall know: Hilbert's apology. *Mathematical Intelligencer*, 21: 1.

Webb, J. (1997). Hilbert's formalism and arithmetization of mathematics. *Synthese*, 110.

Zach, R. et al. (2009). The development of mathematical logic from Russell to Tarski: 1900-1935. In: Haaparanta, Leila (ed.) *The development of modern logic*. Oxford: Oxford University Press.

BALKAN JOURNAL OF PHILOSOPHY. Vol. 2, Issue 2, 2010

"DIAGNOSTIC HEDONISM" AND THE ROLE OF INCOMMENSURABILITY IN PLATO'S *PROTAGORAS*

Tea Logar

Primorska Institute for Natural Sciences and Technology, University of Primorska, Slovenia
tea.logar@upr.si, http://tealogar.com/

Abstract:
The dispute over Socrates' apparent endorsement of hedonism in the Protagoras has persisted for ages among scholars and students of Plato's work. The solution to the query concerning the seriousness and sincerity of Socrates' argument from hedonism established in the dialogue is of considerable importance for the interpretation of Plato's overall moral theory, considering how blatantly irreconcilable the defense of this doctrine is with Plato's other early dialogues. In his earlier works, Socrates puts supreme importance on virtue and perfection of the soul, so the puzzle apparent in the Protagoras merits a thorough examination.

Several scholars have argued that, since Socrates' defense of hedonism in this work clashes significantly with his views on morality in other dialogues, Socrates must only have been defending hedonism ironically, or with the intention of "diagnosing" his opponent's point of view. In this paper, I examine the approaches according to which Socrates didn't in fact mean to defend hedonism, but merely used it as a diagnostic tool; I argue that there is no compelling evidence for this resolution of Socrates' defense of hedonism, and that the views that attempt to defend it really have no convincing grounds for it apart from the desire to reconcile the Protagoras with other Socratic dialogues.

Key words: Socrates, Protagoras, Hedonism, Diagnostic Tool, Incommensurability.

1. Introduction

In Plato's *Protagoras*,[1] the interlocutors are trying to resolve the dilemma regarding the possibility of teaching of virtues, and are primarily concerned with the question whether knowledge can be sufficient for virtue. Socrates' main weapon in trying to persuade Protagoras that nothing but knowledge is required for achieving virtue is his equation of good with pleasure. Based on this purely hedonistic premise, Socrates defends his denial of incontinence (*akrasia*), as well as his "unity of virtue" thesis. Since Socrates compellingly rejects any form of hedonism in Plato's other dialogues (notably in *Apology*, *Crito*, and especially *Gorgias*), and even directly attacks hedonism held by his opponents, this ostensible endorsement of the hedonistic doctrine seems particularly problematic. Because of this discord when compared to other dialogues, several attempts have been made to bring the *Protagoras* in harmony with the ideas expressed in other works of Plato. Donald Zeyl constructed the interpretation which allows for the possibility that Socrates' endorsement of hedonistic premises in the dialogue might be merely apparent. According to Zeyl, Socrates is not in fact approving of the hedonist position, but is only "diagnosing" the view Protagoras and others are really in favor of, although they may not be aware of it initially. According to this interpretation, Socrates is using hedonistic premises not because he takes them to be true, but in order to show his opponents that they, being hedonists themselves, are nevertheless committed to the same principles he is (denial of weakness of will, the "unity of virtue" thesis, and the view that knowledge is sufficient for virtue), even though Socrates reached his conclusions through a different doctrine, namely eudaimonism, rather than hedonism.

I argue that Zeyl's argument could be made plausible only if it could be shown that Protagoras was actually in favor of the same form of hedonism Socrates is putting forward. If it is to be contended that Socrates was merely diagnosing Protagoras' view without actually holding it himself, then the form of hedonism the two interlocutors are proposing should be of the same kind. My thesis in this paper is that Zeyl's argument cannot work, because Protagoras' ultimate account of hedonism entails incommensurability, while Socrates' does not, which in turn demonstrates that the latter cannot be "diagnostic" of the former. After summarizing Socrates' argument from hedonism, I explain its importance in terms of wider Plato's moral theory; I proceed with

[1] All the references to *Protagoras* are from John M. Cooper's *Plato - Complete Works* (Indianapolis: Hackett, 1997), pages 746–790.

assessing Protagoras' and Socrates' accounts of hedonism as they appear in the dialogue separately, concluding that Protagoras' account necessarily involves incommensurability. I then examine Henry Richardson's argument that Socrates' account of hedonism in the *Protagoras* does not entail commensurability, and that therefore Zeyl's thesis about Socrates' hedonism being merely diagnostic could nevertheless be sustained. I argue, however, that Richardson's interpretation is farfetched, and I conclude that Zeyl's argument of Socrates' hedonism as purely diagnostic inevitably fails. Nevertheless, I still allow for the interpretation that Socrates is being ironic in the *Protagoras*, and that hedonism is therefore possibly not the doctrine he actually endorses – but I maintain that this cannot be explained through the claim that Socrates is *diagnosing* Protagoras' view, since Protagoras does not agree with Socrates' idea of hedonism in the first place, and only seems to be "tricked" into accepting it later on.

2. The Argument from Hedonism

The section of the *Protagoras* that ultimately concerns hedonism is initiated through the disagreement between Socrates and Protagoras on the subject of whether virtue can be taught. While Protagoras claims that he is teaching virtue, Socrates is expressing doubts about this being possible, questioning whether and how it can be done. However, at certain point the two interlocutors seem to exchange positions: Socrates eventually maintains that knowledge is sufficient for virtue, implying that it can be taught after all, while Protagoras argues against this claim, pointing to particular virtues the practice of which seems to depend on more than mere knowledge of what they consist in. Because of this change in their positions, Socrates notes in the end that the argument he just had with Protagoras would ridicule them if it could speak in its own voice.

Approaching the passage that concerns hedonism, Socrates argues in support of his thesis regarding the unity of virtue by claiming that knowledge is sufficient for virtue, and since all the virtues can be identified with the knowledge of the good, it presumably follows that all the names of the virtues are really names of one and the same thing. Protagoras disagrees with Socrates' view by pointing out that certain virtues defy this definition, since knowledge is not sufficient for achieving them: courage, for example, comes not only from knowledge, but also from "nature and the proper nurture of the soul" (351a-b). If Protagoras is successful in showing that at least some virtues require non-cognitive components and not only knowledge, Socrates would most likely have to reject the unity of virtue thesis. In trying to answer this challenge, Socrates turns to hedonism. While Protagoras and the majority of those present reject the doctrine at first, Socrates succeeds in convincing his opponents that they are actually hedonists without realizing it, and since they are all committed to equating good with pleasure, they are inevitably committed to accepting Socrates' denial of weakness of will,[2] as well as his unity of virtue thesis.

The passage we should focus on in order to learn whether Socrates' defense of hedonism is serious, and which form of hedonism he thinks everyone should accept, we should focus on the section of the *Protagoras* from 351b to 358d, where Socrates is primarily concerned with denying incontinence. Socrates begins his argument by claiming that, since to live pleasantly is good, and to live unpleasantly bad, it has to follow that pleasant things are good, and unpleasant bad. In his opinion, if someone truly knows what is good and what bad, there is nothing that could force them to do the bad thing, and not the good one – hence, knowledge is sufficient for virtue, and there is no such thing as the weakness of will.

Protagoras and the majority disagree: not all pleasures are necessarily good, they say, such as the pleasures of eating and drinking too much, and even though people in general know that these are likely to be bad for them, this knowledge alone does not necessarily stop them: people are often overcome by immediate pleasure, and they enjoy these pleasant things nevertheless, while knowing they are bad in the long run. Accordingly, there are inevitably some bad pleasures, as well as there are many good, yet unpleasant, things: undergoing a painful medical procedure, physical exercise, going to war, etc. People thus often show weakness of will, since they sometimes cannot resist bad pleasures, or, conversely, they refuse to suffer certain pains despite knowing that undergoing them would be ultimately good for them. In Protagoras' own words, "…people are unwilling to do what is best, even though they know what it is and are able to do it" (352d).

Socrates, on the other hand, finds this view absurd. According to him, good is defined solely through pleasure, and it therefore doesn't make

[2] The terms "incontinence," "akrasia", and "weakness of will" are used interchangeably throughout the paper.

sense to speak of bad pleasures and good pains. It is true that some things Protagoras mentioned may result in immediate pleasure and yet end up being ruinous and harmful for us; but Socrates' point is that it is wrong to call them "pleasures" at all in such cases: since we call these things bad precisely because they bring about pain and deprive us of pleasure, we should thus call them pains, and certainly not pleasures. Socrates thus asks Protagoras:

> These things [athletics, military training, and treatments by doctors such as surgery, medicines, and starvation diets] are good only because they result in pleasure and in the relief and avoidance of pain? Or do you have some other criterion in view, other than pleasure and pain, on the basis of which you would call these things good? (354b-c)

It turns out that neither Protagoras nor other people present can think of any other criteria for determining what is good, so it presumably becomes self-evident that the proportion of pleasure and pain that is ultimately produced by an action is the only criterion for determining whether the action is good or bad. Good clearly equals pleasure, says Socrates: "the good and the bad are that which result in pleasure and pain" (355b), and that is all there is to their definitions. Based on their own beliefs, Protagoras and everyone else are hence committed to equating good with pleasure, accepting the hedonist position, and consequently agreeing with Socrates' denial of incontinence, as well as accepting his "unity of virtue" thesis.

In defense of his denial of weakness of will, Socrates explains that the description of incontinence becomes absurd once 'good' has been replaced with "pleasant," and "bad" with "painful." It is simply ridiculous to say that one decides to do something bad because he finds himself overcome by the good (or "pleasant") in it. Since pleasure cannot outweigh pain in any other way then by relative excess or deficiency (356a), a thing simply cannot be bad and pleasant at the same time: ultimately, after all the measuring of pleasure against pain has been done, it is either one or the other. So the presumed "bad pleasures" are not really pleasures at all, since their resulting pain outweighs the pleasure they produce immediately. If the painful exceeds the pleasant in certain action, then it is bad, and people should refrain from performing it for their own good.

If Socrates is right, however, then how can we then explain that people keep acting in self-defeating manner by choosing immediate pleasure despite being aware of the pain it will bring in the future, seemingly demonstrating weakness of will? According to Socrates, it is simply because people are not able to tell correctly whether future pain will exceed present pleasure. Just as things closer to us appear larger than things of the same actual size that are positioned farther away from us in space, the same thing happens with things positioned in time. Things in the future appear "smaller," and we are thus unable to measure present pleasures against future pains (or the other way around) accurately. "The art of measurement, in contrast, would make the appearances lose their power by showing us the truth" (356e), says Socrates, implying that, were we able to tell exactly whether certain action will produce more pain altogether, we would without doubt refrain from performing it, and nobody would be considered "incontinent."

Ultimately, it is the lack of knowledge of the accurate measurement, and not the weakness of will, that results in people doing things that are immediately pleasant, but produce more pain on the whole. "Being overcome by pleasure" is simply ignorance – only if one has a false belief about matters of the amount of pleasure and pain will he choose to perform an action that will ultimately result in more pain than pleasure. Knowledge of measuring the accurate amount of pleasure and pain is thus sufficient for virtue, which further supports Socrates' "unity of virtue" thesis, grounded in his claim that one can reach conclusions about each virtue based entirely on the (epistemological) measuring science of pleasure and pains, and not on any further ethical judgments.

3. The Importance of Socrates' Defense of Hedonism, and the "Diagnostic" Thesis

The difficulty this passage raises is of much larger importance than Socrates' seemingly problematic denial of incontinence. As already noted, the *Protagoras* leaves the reader with a very strong impression that Socrates is defending hedonism, especially since some conclusions that are otherwise crucial for Plato's moral theory of the time seem to end up relying entirely on the equation of good with pleasure. Socrates' view of hedonism in this passage is relevant for the interpretation of Plato's overall ethical theory, given that Socrates vigorously denies hedonism in both earlier and later dialogues: *Apology*, *Crito*, and *Gorgias*, for example, all appeal to the supreme importance of virtue and are inconsistent with the hedonistic equation of good with pleasure. Although the *Protagoras* itself gives no indication that Socrates is not in fact committed to the doctrine he uses in support of his conclusions (i.e., hedonism), some authors have tried to defend the view that Socrates is here merely using hedonism as a convenient *tool* for convincing his audience to accept some of his crucial moral principles, but without in fact committing to hedonism himself. The main motivation behind this interpretation is the one of justifying and making sense of Socrates' appealing to a questionable moral doctrine which he otherwise rejects fiercely and often.

Donald Zeyl claims that Socrates' questions in *Protagoras* should be interpreted as merely diagnostic of the standpoint of his interlocutor and the audience, and not necessarily as expressing Socrates' own views.[3] To the question of why else would Socrates be persuading his audience to accept the hedonistic position, if he did not believe it to be true himself, Zeyl answers that Socrates is not persuading his interlocutors to *become* hedonists – he is merely pointing out that they already *are*, without realizing it themselves.[4] On Zeyl's view, Socrates could have just as well provided sufficient evidence for the unity thesis and denial of *akrasia* through eudaimonism – the theory that he in fact endorses – and this would have made the dialogue much more plausible in the view of Plato's overall moral theory and generally compatible with other Socratic dialogues. But, Zeyl argues, such a course of action would have been much more difficult for Socrates to take considering the circumstances, since he would first have to convince his audience to accept eudaimonism over their own view of the good (which is presumably hedonism), and only after they have accepted this view could Socrates show them that this commits them to denial of the weakness of will, as well as of the unity of virtue thesis.

According to Zeyl, then, since Protagoras and the rest are already committed to hedonism, Socrates would have only lost the strategic advantage of his position if he openly questioned the doctrine; the great advantage which Socrates gains from the *use* of hedonism in arguing against incontinence therefore doesn't imply that he actually *accepts* the doctrine. Zeyl refers to his position regarding Socrates' apparent hedonism as that of *antihedonism* (since he *denies* that Socrates in fact endorses hedonism), and concludes that "the weight of evidence is heavily in favor of the antihedonist case," considering that such an interpretation is much more intelligible in the light of Socrates' aims; Zeyl continues that the *pro-hedonist* account (according to which Socrates *is* endorsing hedonism in this text) actually creates more problems than it solves, as opposed to Zeyl's own antihedonist view.[5]

Terence Irwin, however, objects to this type of *ad hominem* argument in support of Socrates' antihedonist position. Irwin argues that, since Zeyl's conclusion is based on an assumption that hedonist premises are sufficient, but not necessary, for a valid argument in support of the unity of virtue thesis and denial of incontinence, there should be another plausible path available for Socrates to argue for these doctrines, without appealing to hedonism, but preferably to premises he actually accepts.[6] According to Zeyl, this path is available to Socrates: he could have, theoretically, used eudaimonism instead of hedonism; he only chose not to do so because this path was easier, strategically advantageous, and less time-consuming. But Irwin insists that Socrates could not have used eudaimonism in order to achieve his intended goals; according to Irwin, the common-sense belief in incontinence constitutes a strong objection to Socrates' psychological eudaimonism, just as it does to hedonism. The fact that Socrates chooses to fight the incontinence objection by relying on hedonism instead of eudaimonism seems to present quite a strong case against those who claim that Socrates is only using, but not actually endorsing, hedonist premises in the *Protagoras*. "The hedonist argument does what Socrates needs it to do only if he believes it," claims Irwin.[7]

This conclusion seems quite plausible when one takes into account the fact that Protagoras and the rest initially resist Socrates' hedonist premises. Since Zeyl claims that the only reason why Socrates is not using eudaimonism is simply because he would then lose the advantage of his audience agreeing with his initial premises, one would expect that the audience actually accepts the hedonist premises right from the beginning. The fact is, however, that Protagoras resists Socrates' claims about the equation of good and pleasant quite vigorously at the beginning of the passage. When Socrates asks Protagoras whether he accepts the premise that pleasant

[3] Donald J. Zeyl, "Socrates and Hedonism: Protagoras 351b – 358d" in *Phronesis* 25 (1980): 250–269.
[4] Ibid., p. 261.
[5] Ibid., p. 264.
[6] Terence Irwin, *Plato's Ethics* (New York: Oxford University Press, 1995); p. 86.
[7] Ibid., p. 86.

things are good only insofar as they are pleasant, while painful things are bad only insofar that they are painful, Protagoras is very hesitant. He responds: "I don't know, Socrates, if I should answer as simply as you put the question," and further states that "there are pleasurable things which are not good, and on the other hand, there are painful things which are not bad…" (351d), thus obviously denying Socrates' idea of hedonism. It seems that Socrates does not really have the advantage of the use of hedonism in the form Zeyl is proposing: Socrates first has to assert hedonism, then convince his interlocutors that they should be committed to it as well, and only then can he proceed with support of his further theses. It is unclear why Socrates wouldn't choose to use the same strategy to persuade his audience of the moral principles he actually accepts.

It does seem, however, that Zeyl could still vindicate his view by arguing that Socrates is only so successful in convincing his audience to accept the hedonistic premises precisely because they are already committed to them, although without realizing that this is so before Socrates' intervention. As already noted, Zeyl's claim is that Socrates has to assert hedonism in order to show the audience that they *already* accept it, not in order to *convince* them to accept it. If one supposes that Irwin's claim about Socrates' inability to prove his doctrines through eudaimonism is exaggerated, and that all that Socrates is really trying to do is show his opponents that they are committed to the same ends he is, despite employing different means, then Zeyl's interpretation certainly gains plausibility. Zeyl's conclusion obviously relies on the interpretation that Protagoras and the rest are already accepting the form of hedonism proposed by Socrates, and are thus committed to his conclusions regarding *akrasia* and unity of virtue. The question I would like to investigate more thoroughly is whether there is enough evidence in the *Protagoras* to support the interpretation that Protagoras is actually committed to the form of hedonism Socrates is presumably merely "diagnosing," rather than endorsing. In order to examine this approach as presented by Zeyl, I will have to compare the form of hedonism Protagoras seems to be voicing against the form seemingly advocated by Socrates.

4. The Form of Hedonism Accepted by Protagoras

As already mentioned above, Protagoras is hesitant to accept the equation of good with pleasure which Socrates initiates in the beginning of the passage. When Socrates contends that to live pleasantly is good, and to live unpleasantly bad, Protagoras replies that this is true only insofar as the life in question was lived taking pleasure in *honorable* (praiseworthy) things (351c). This response seems to imply that Protagoras has a standard of good in mind that is different from the narrow criterion that only refers to mere amount of pleasure – and this despite the fact that later, when asked to provide a different criterion of good, Protagoras cannot think of any (354c). The idea that good depends not only on quantity, but also on *quality* of pleasures, however, differs significantly from the criterion Socrates seems to be endorsing in this dialogue. If something can be pleasant, as Protagoras contends, but yet not honorable, and thus presumably bad, then this surely entails that not everything pleasant is also necessarily good; in other words, it appears that something can be bad even when its resulting pleasure eventually outweighs its resulting pain.

The statement about honorable pleasures, combined with Protagoras' claim about pleasant things that are bad, and painful things that are good (351d) is probably the single most important sentence in the passage that shows how strongly Protagoras is initially opposed to the principle that good can be defined and explained solely through the amount of pleasure involved. The fact that Protagoras endorses the thesis according to which the quality of actions seems to depend on more than mere pleasure possibly entails a certain kind of incommensurability in his account of the good. Socrates, however, seems surprised at receiving such a response: he accuses Protagoras of being like most people (in the pejorative sense, to be sure) who call some pleasant things bad, and some unpleasant things good (351c-d). But once Protagoras' claim about honorable pleasures is taken into account, it seems that his view of the relation between good and pleasant cannot entail the same form of hedonism that Socrates needs in order for his conclusions about the denial of incontinence and unity of virtue to follow – even if one accepts Zeyl's argument of Protagoras being initially committed to hedonism. As Henry Richardson puts it,

Since much of Zeyl's otherwise convincing and thorough argument depends upon his insistence that Protagoras is committed to hedonism despite himself, his theses that Socrates is not committed to hedonism in the *Protagoras* will stand of fall with the validity of supposing that the commensurability of pleasures and pains – if not indeed of all goods – is an implicit assumption of the opponents.[8]

But the commensurability of pleasures does not seem to be Protagoras' implicit assumption at all: he introduces qualitative distinctions between pleasures that Socrates' simplistic picture of "relative excess and deficiency" cannot seem to incorporate. Zeyl's argument therefore seems determined to fail: Protagoras' mention of honorable pleasures, even if understood hedonistically in broader context, implies certain qualitative differences between various sorts of pleasures, while Socrates seems to need some kind of a strictly commensurable form of hedonism, where the *amount* (and not the *quality*) of pleasures versus pains seems to be the only acceptable criterion for distinguishing between good and bad. If this assumption is true, then it seems that Zeyl's conclusion about Protagoras and the many already accepting the form of hedonism that commits them to Socrates' conclusions must be mistaken: Protagoras initially rejects Socrates' claim that the amount of pleasure versus pain is the only thing needed for measurement of good, and he seems to be practically tricked into accepting it later by Socrates. It should therefore follow that Protagoras' initial view of the good commits him neither to denial of incontinence, nor to the "unity of virtue" thesis.[9] If this is true, then Zeyl's argument for Socrates' merely using, but not accepting hedonistic premises loses its grounding force: if Protagoras is committed neither to eudaimonism, nor to Socrates' form of hedonism, then why would Socrates choose to employ hedonism, the doctrine that he presumably rejects, over eudaimonism, which he accepts, as a tool for persuading his opponents?

My aim in this paper is to show that Socrates' hedonism as presented in this dialogue differs from Protagoras' understanding of the relation between good and pleasure, with the result that Zeyl's claim about Socrates' hedonistic premises being merely diagnostic of his interlocutors' beliefs loses its persuasiveness. I will show that, even if one assumes that Protagoras is accepting *some* form of hedonism, this form is significantly different from the one Socrates is endorsing, and because Protagoras' hedonism necessarily includes some incommensurability, while Socrates' pleasures and pains have to be mutually commensurable, the latter account cannot be diagnostic of the former. In order to show exactly where the difference between the accounts lies, we must explore Socrates' own understanding and implications of hedonism as he presents them in the *Protagoras*.

5. The Form of Hedonism Endorsed by Socrates

The claim made above, namely that Socrates' view of hedonism entails some kind of "Benthamian" picture of pleasures and pains which are all mutually commensurable, and that therefore this view differs significantly from Protagoras' account, is taken into question by Richardson. According to him, "there is nothing in the dialogue that requires us to see commensurability as part of its hedonism."[10] While Richardson refrains from directly addressing the question whether incommensurability of goods apparently present in Socrates' view of hedonism implies that he is using the doctrine merely as a diagnostic tool in this dialogue, Richardson does maintain that such an interpretation at any rate brings us closer to Socrates' otherwise accepted beliefs regarding good and pleasure. On the one hand, this interpretation eases Socrates' problematic endorsement of the doctrine in the *Protagoras*, since it allows for the interpretation that Socrates' hedonism is only diagnostic of Protagoras' view, but it differs from Zeyl's argument in that it allows for both accounts to contain incommensurability.

According to Richardson, many authors such as Terence Irwin, Martha Nussbaum, and also Donald Zeyl are mistaken in interpreting Socrates' story in the *Protagoras* as necessarily endorsing commensurability of value which results from the singleness of pleasure as the ultimate end. Richardson first points out that hedonism does not automatically imply commensurability, as these authors seem to assume. For example, argues Richardson, the form of very weak hedonism – namely *coextensive hedonism* – allows for qualitative differences and conflicts among different kinds of pleasures and pains, and is as such compatible with incommensurability of

[8] Henry Richardson, "Measurement, Pleasure, and Practical Science in Plato's Protagoras," in *Journal of the History of Philosophy* 28 (1990): 7–32; p. 14.

[9] The claim that "incommensurable" hedonism which Protagoras seems to be endorsing cannot entail denial of incontinence and the unity of virtue thesis is predominantly Terence Irwin's conclusion. I am here more concerned with the difference between Socrates' and Protagoras' accounts of hedonism than with their implications considering incontinence and the unity thesis.

[10] Richardson, p. 8.

pleasures. Coextensive hedonism simply maintains that all and only pleasures are good, and all and only pains are bad, without further implying that it is easy, or indeed possible, to measure these pleasures against each other. While Richardson recognizes that Socrates' hedonism in *Protagoras* does go beyond this depiction of the doctrine, namely by the means of describing pleasure as the *single* good, he maintains that "…the singleness of pleasure might merely be a matter of there being a univocal definition of pleasure. This does not imply, however, that there are no pleasures of such radically different types as to become incommensurable."[11]

First, Richardson assesses Zeyl's assumption that Socrates' view implies commensurability, in order to show that Zeyl's interpretation of Socrates' hedonism as *ad hominem* can be preserved without assuming commensurability in his view. Zeyl argues that "if being a pleasant life is a sufficient condition for being a good life, it will follow that a predominance of pleasure in any life […] suffices to qualify that life as a good life."[12] Based on this, Zeyl interprets Socrates' claim that "living a predominantly pleasant life is good" as indicating that the use of the term "predominantly" implies that living pleasantly is a matter of *degree*, which would subsequently allow for a commensurability between degrees of living well. Richardson, however, sees no grounds for this conclusion. While Socrates without a doubt voices a version of coextensive hedonism according to Richardson, he maintains that "Plato has [Socrates] devote extra words for keeping the door open for the incommensurability of pleasures and pains."[13] In support of this claim, Richardson calls our attention to Socrates' naming of three standards for weighing pleasures against pains in the form of three disjuncts: larger or smaller, more or fewer, and to greater or lesser degree (355d-356a). One possible reading of this picture is, of course, that Plato is setting out details for how different pleasures and pains might be measured against each other, thus endorsing commensurability. Socrates' exclamation that there is no other way for pleasure outweighing pain except in relative excess or deficiency seems to confirm this interpretation. Richardson agrees that this understanding seems plausible at first: "On this reading, the tripartite standard may be taken to refer to the number of pleasures, their intensity, and their duration."[14]

Richardson makes the point, however, that this is only one possible interpretation, and that we have no reason to assume that the disjuncts were actually not meant to preserve incommensurability of value. He argues that Zeyl presumes homogeneity of pleasure which is in reality not suggested by Socrates, who actually speaks of the many "pleasant things."[15] Richardson further explains that it is actually difficult to make sense of Socrates' disjunctive standards of choice without backing off from the homogeneity of pleasure, and also that it is easier to explain them on the assumption that Socrates is speaking of incommensurably different types of pleasure and pain. His ultimate claim is, however, not that Socrates' account necessarily endorses incommensurability, but merely that Plato leaves the possibility for it open – Socrates' account of hedonism does apparently not exclude it.

If Richardson is right in arguing that Socrates' account of hedonism need not entail commensurability, then it seems that the proposed argument about the failure of Zeyl's account seems to lose its force. My claim was namely that Protagoras' hedonism entails incommensurability of some sort, while Socrates' view endorses strict commensurability between pleasures and pains. Based on this, I concluded that Socrates' hedonism cannot be merely diagnostic, since it fails in "diagnosing" Protagoras' account accurately, and because of this Zeyl's argument cannot hold. In order to see whether Richardson's claim about Socrates' account possibly involving incommensurability actually succeeds in preserving Zeyl's argument about Socrates' hedonism in the *Protagoras* being merely diagnostic of his opponents' view, we must explore Richardson's argument in support of incommensurability further.

6. Place for Incommensurability in Socrates' Hedonism

As explained in the previous section, Richardson draws his conclusions about the possibility of Socrates' hedonism containing incommensurability from two sources. The first one is the fact that Socrates consistently uses plural when referring to pleasures and pains. While this can be explained by him simply referring to different *instances* of pleasure and pain, Richardson calls our attention to the

[11] Ibid., p. 11.
[12] Zeyl, p. 252.
[13] Richardson, p. 15.
[14] Ibid., p. 16.
[15] Ibid., 13.

possibility that Socrates is actually talking about different, possibly incomparable, *types* of pleasure and pain.[16] Since this is a possible interpretation, and it would allow for incommensurability in Socrates' account of hedonism, we can move further to see if there is any more compelling evidence in support of this suggestion. The main reason why Richardson believes the door for incommensurability is not only open, but that such an interpretation actually makes more sense, is Socrates' introduction of standards of weighing pleasures against pains. According to Richardson, Socrates offers three distinct disjunctive standards when he exclaims: "But how else does pleasure outweigh pain, except in relative excess or deficiency? Isn't it a matter (to use other terms) of larger and smaller, more or fewer, greater or lesser degree?" (356a) Richardson examines the option that these three disjuncts stand for a tripartite standard that could be illustrated by the analogy of the pan balance, and the standards correspond to the number, density, and size of the weights.[17] This interpretation of Socrates' words seems to be intuitively right. If I have all the needed knowledge about exactly how much pleasure and pain a given action will produce, then I will be able to measure whether the action in question will yield more pleasure or pain, and, accordingly, whether it is good or bad. I will figure this out by using the appropriate measurement, either when trying to resolve whether an action x produces more pain or pleasure, or when trying to compare two actions against each other. If I ask myself whether I should perform action x or y, I put all the pleasures and pains of x in one pan, and all the pleasures and pains of y in the other, and then compare which pan turns out to be 'heavier.' The pan containing action x might end up being heavier either because it contains more units of pleasure, or because its units are more "intense," or because they are larger – the balance will figure out the exact combined result. The similar thing should happen if I am trying to assess whether a certain action is good or bad, by putting all its resulting pleasures in one pan, and all its pains in the other. According to Socrates this is what we do all the time: this is the criterion of how we in fact assess actions. The main reason why we sometimes obtain wrong results is that we attribute less weight to the future pleasures or pains, because we think they are smaller than they really are.

What could be wrong with such an interpretation of Socrates' argument, which seems to rely on a somewhat complex understanding of commensurability among pleasures and pains? Richardson believes that it doesn't really make sense, once examined more closely. He points out that this picture fails to explain the place it gives to the "number" of pleasures as one of the independent measuring factors. He shows through the means of a practical example that, once the other two factors (intensity and duration) have been incorporated when measuring two actions against each other, the standard of "more or less" becomes redundant.[18] Its role seems to be merely to ensure that all relevant sources of pleasure were counted, and not to actually make a difference when assessing actions. Besides that, even if one chooses to interpret "more or less" as indicating not the number of pleasures, but the temporal dimension Socrates gives such importance to, the interpretation fails nevertheless, since Socrates is merely saying that the factors of the temporal nearness and remoteness must be put into balance, and not that they require a distinct standard.[19]

Since none of the interpretations of the "more or less" standard seem to make sense when commensurability is assumed with Socrates' hedonism, Richardson proposes a different interpretation: if one assumes that the pleasures being weighed here are actually of incomparably different types, the disjunctive standards finally start to make sense. He incorporates Ralph Barton Perry's mutually irreducible standards of intensity, preference, and inclusiveness. Richardson supplements the proposed system with the dominance rule, the principle that combines elements of intensity and inclusiveness. If we then interpret Socrates' disjunctive standards through this system, we get the following picture: "larger or smaller" corresponds to the judgment of intensity, "more or fewer" to counting terms, and "greater or less" to the dominance rule.[20] With such a picture in mind, one can make sense of the three disjunctive standards proposed by Socrates, assuming, of course, that his hedonism incorporates incommensurability.

I will refrain from providing all the details of Richardson's argument, since I will at this point assume that his incorporation of Perry's standards can be sustained, and that a system incorporating

[16] Ibid., p. 15.
[17] Ibid., p. 16.

[18] Ibid., pp.16–17.
[19] Ibid., p. 17.
[20] Ibid., p. 18.

these features and assuming incommensurability can be actually made more plausible than the previously proposed "pan balance" standard. However, I intend to question Richardson's assumption that incommensurability can be hypothesized within Socrates' view of hedonism in the first place.

7. Reply to Richardson

There are two ways in which one can address Richardson's interpretation: first, it is not clear why one should assume that Socrates had such an elaborated account (the "pan balance" analogy) in mind in the first place. Socrates explicitly says that "larger and smaller, more or fewer, greater and lesser degree" are merely *other terms* one can use to describe what he means by excess and deficiency (356a) – it does not follow that these are three distinct standards that correspond to intensity, duration, and number. By introducing the three disjuncts, Socrates seems to be merely explaining in what senses the quantity of pleasures can be understood; he does not seem to be introducing distinct standards, each of which needs to be employed every time one tries to assess an action, or compare one action against another, as Richardson takes them to be. Socrates is merely explaining what one means when saying that one is overcome by pleasure: "you mean getting more bad things for the sake of fewer good things," he says first (355e). Then he apparently realizes that the quantity of pleasure might be understood differently than in the number of units ("more or fewer") contained in certain instances, so he uses other terms that can be applied depending on the specific occasion, such as larger or smaller, and greater or lesser degree. Sometimes comparing the number of pleasures will suffice; at other times relying on the number alone could be misleading, so one needs to consider their intensity instead (or as well), and so on. The reason why Socrates mentions these disjuncts may therefore not be that he wants to introduce *different* standards, but rather to illustrate that *no other* standards may be used. Socrates may be plausibly interpreted as saying that any standard that can be employed must necessarily be of a *quantitative* nature. There is no room in this account for such absurdities as "good pains" and "bad pleasures;" there is no such thing as wondering about the "quality of pleasure," as Protagoras is suggesting, since pleasure can only be good; there is only *quantity of pleasure*, which determines the *quality of an action*. Having said this, it seems erroneous to understand the disjunctive forms Socrates uses as implying different standards: the interpretation that he is merely pointing out that whatever standard we will use in order to evaluate actions, this standard will necessarily have to be quantitative, is much more consistent with his other remarks about Protagoras' reasoning.

Moreover, even if one assumes that Socrates had the exact "pan-balance" model in mind, one can still conclude that he did simply not realize that the "more or fewer" standard would not make sense once thought about more thoroughly, as Richardson does.[21] The fact that a thorough examination of Socrates' account might prove to feature redundant principles hardly gives us a reason to suspect that this is not an account he had in mind. Socrates' claims about the "quality of pleasure" talk being absurd, as well as his trust in the possibility of exact measurement, all seem to imply that he strongly believed that all the pleasures and pains are mutually commensurable. Therefore, the reason why it seems intuitively right to interpret Socrates as endorsing commensurability in the *Protagoras* is simply that he objects to any qualitative differences between pleasures right from the beginning – recall his arrogance when Protagoras mentions honorable pleasures. Socrates' message is: there is no other way to define the quality of an action except through the quantity of pleasure and pain it ultimately produces. And because quantity is the only criterion of value, we cannot know the value of an action (i.e., whether it is good or bad) through any other means but by comparing the amount of pleasure to the amount of pain it produces – and in order for an accurate measurement to be possible, pleasures, as well as pains, must be mutually comparable and commensurable. If they were not, there would have been no use in claiming that knowledge of appropriate measurement is sufficient for knowing what is right – one would also have to know how to distinguish between different *qualities* of pleasures, and the whole point of Socrates' speech seems to be that this is not only unnecessary, but that it also makes no sense: it is ridiculous to claim that a pleasure as such can be anything but good, and pain anything but bad – it is their sum total that makes a particular action good or bad. Claiming that one pleasure is worse or better

[21] I suppose that Jeremy Bentham's utilitarian account could have been criticized along the lines that Richardson uses to criticize Socrates' hedonism. But the possibility that Bentham's account cannot be sustained gives us no reason to infer that commensurability is not what Bentham had in mind.

than any other pleasure is clearly considered absurd by Socrates, since one would then have to claim, after substituting "pleasure" for 'good,' that one good is better than another good, despite there being no difference in quantity between them. Only if one assumes that differences between pleasures can be of no other than quantitative can one make sense of Socrates' conclusion that knowledge of accurate measurement is sufficient for virtue.

8. *Diagnostic Hedonism*?

My thesis in this paper is that Donald Zeyl's interpretation of Socrates' use of hedonism in the *Protagoras* being merely diagnostic of his interlocutor's views can only hold if both Socrates' and Protagoras' accounts were sufficiently similar. Zeyl assumed that both accounts involve commensurability, and that it is therefore plausible to assume that Socrates was merely expressing his opponent's views when seemingly endorsing hedonism, with the intention to show his opponents that they are committed to the same conclusions about virtue that he himself reached through eudaimonism. It has been established, however, that Protagoras could not have been endorsing the kind of commensurability that Socrates argues for, which is evident from Protagoras' remarks about honorable pleasures, as well as his initial denial of the amount of pleasure being the only criterion for establishing what is good. Because it seems that Protagoras was embracing incommensurability of values, while Socrates kept defending their commensurability, I conclude that Zeyl's argument must fail, since Socrates' account differed from Protagoras' in too great a measure to support the thesis that the former is merely diagnostic of the latter.

Regarding Richardson's argument that Socrates' account of hedonism actually makes more sense if interpreted as endorsing incommensurability, and that Zeyl's argument about Socrates' hedonism being diagnostic of Protagoras' could work after all, I have argued that, while the suggestion is plausible in and of itself, we really have no good reason to suspect that Socrates was endorsing incommensurability in his account, since the evidence for such an interpretation seems highly questionable and farfetched. It is evident that the assumption that there is place for incommensurability in Socrates' account would not have even arisen if his endorsement of hedonism didn't look so strange in the light of other Socratic dialogues. It cannot be denied that no reader who has read the *Protagoras*, but has never heard of other Plato's writings, would have even considered the possibility that Socrates is not endorsing hedonism in this dialogue. Consider the facts: Protagoras rejects Socrates' initial hedonistic claims, but Socrates works hard enough throughout the dialogue in order to ultimately persuade Protagoras to accept a different point of view which is vigorously endorsed by Socrates. Why would one even suspect that the view finally agreed on has in fact been accepted all the time from the party who initially denied it, but really denounced by the party who worked so hard to make his opponent accept it? The possibility that Socrates is merely "diagnosing" his opponent's convictions seems to be an exceptionally farfetched attempt to try to restore harmony among Socratic dialogues, especially when one considers the differences between Socrates' and Protagoras' thoughts on the issue of commensurability of value.

My conclusion is that the only reason one could have for accepting the "diagnostic" interpretation Zeyl proposes is in order to make sense of the whole Socratic theory of virtue. I am not denying the possibility that the interpretation of Socrates' using hedonism ironically is implausible on the whole, but I maintain that no evidence for this claim can be found in the *Protagoras* itself.

References

Irwin, Terence (1995). *Plato's Ethics*. New York: Oxford University Press,

Plato. Protagoras (1997). In *Plato – Complete Works*, by John M. Cooper, 746-790. Indianapolis: Hackett.

Richardson, Henry (1990). "Measurement, Pleasure, and Practical Science in Plato's Protagoras." *Journal of the History of Philosophy* 28: 7–32.

Zeyl, Donald J. (1980). "Socrates and Hedonism: Protagoras 351b – 358d." *Phronesis* 25: 250–269.

BALKAN JOURNAL OF PHILOSOPHY. Vol. 2, Issue 2, 2010

ONTOLOGY AND FUNCTIONALISM IN HEGEL'S *PHENOMENOLOGY OF SPIRIT*

Claudiu Baciu

Institute of Philosophy and Psychology "Constantin Radulescu-Mortu" of the Romanian Academy
cbaciuro@yahoo.com

Abstract:

The paper starts by describing the distinction between substantialist thought and functionalist thought. According to philosophical functionalism the object of human knowledge is always a result of an inner construction, and not a passive reflection of the outer world. The Critique of Pure Reason was the first modern materialization of this new philosophical program. One of the most important results of the Kantian criticism was the reconstruction of the concept of ontology. The possibility of ontology, according to its new concept, was given by the idea of identity between the possibility of reality, pertaining to human thinking and knowledge, and that which we call "reality". The entire German Idealism took over this new concept of ontology. The specific Hegelian undertaking in the Phenomenology of Spirit was to show that this identity is not posited at the beginning, but is described in its development for consciousness. The study tries to show that this undertaking is a functionalistic one because here reality, as an object of consciousness, receives its specific configuration on the ground of the logical unity of the moments belonging to each form of consciousness and to their totality.

Key words: ontology, function, science, phenomenon.

1. Introduction

According to Aristotle, whose ontological conception has dominated western thought until the end of the Middle Ages, our concepts reflect real substances, those substances which are effectively *in re*. Between knowledge and reality there is no insurmountable wall. This understanding of knowledge has ensured that ancient and medieval science has had mainly a qualitative character, unlike modern science, which had a quantitative one. Ancient and medieval science tends to grasp the inherent substance of things and not the relations between them. Moreover, mathematics, for us the science of these relations, was considered by Aristotle an obstacle to such a grasping. On the other hand, as Cassirer states (1910, 15), in mathematics and geometry we deal with a different type of concept building. Mathematical concepts do not reflect outer realities, so numbers and geometric forms cannot be considered such reflections. On the contrary, in mathematics we encounter a real *Setzung*, a creation of concepts. And the same thing happens in the whole modern science of nature. This science, which separates itself only with difficulty from the substantialist ideas of the philosophical tradition, founds its theoretical undertaking on certain conceptual constructions, by means of which it can proceed to investigate reality. On account of these fundamental concepts elaborated by scientists, one makes a *selection* within the multitude of the aspects of reality. We find such a selection in every science. This selection takes place even on the level of our common perception. Hence, we encounter reality only while starting from a previous concept about the thing that we shall encounter (Cassirer 1910, 225). This is the modern *functional* model of knowledge, a model in which reality is given to us according to the concepts which are already possessed by us.

This change from the ancient and medieval paradigm of substantialism to modern functionalism takes place also with regard to the understanding of the human subjectivity. If in the case of the ancient way of understanding this subjectivity, the human self is interpreted as starting from the world – understood as a cosmic order, as a world which is at the same time a "creation of God," where natural things are in harmony with each other and have a finality (even if not always understandable by the human being). The human self is thought here as a "substance" (after the model of the external things), an entity, which, therefore, is conceived as coming in this world and going out of it, without changing its essence. In the case of the second way of thinking, we have an interpretation of this self on the ground of an inner principle, which is the self-certainty (Taylor 1975, 6). Even if now this self remains also in correlation with the interpretation of the outer world (an interpretation whose fundament is given now not by the Aristotelian science, but by the modern Galilean science of nature), due to the new content of this science the self is no longer seen as a reflection of the macrocosm, but as a principle of unity, a principle of construction. Modern science no longer works with analogies or final causes, but with principles and deductions, with experiments and laws. The self and human consciousness are conceived starting from the model of this science, as generative principles. As one knows, the intention of this science, and in general that of all modern knowledge, is to dominate nature (knowledge is power, for modern thinkers), a nature, which is no longer perceived as a seat of a "cosmic order," but as a seat of processes which have no universal logic, i.e. no universal *ontological meaning*.

This new sensibility from which modern science has emerged appeared at the end of the Middle Ages, when, in the "quarrel of the universals," nominalism defeats realism. According to this nominalism, man and human intellect cannot understand the reasons of divine will. If in the past the human intellect was interpreted as participating in the divine, now at the centre of theology is the postulate of divine omnipotence, which replaces that of divine love. The consequence of this shift is that God, in his omnipotence, cannot be thought as being limited, even if this limitation be his love. Therefore one cannot give any rational reason for the fact that the natural things are as they are and not different, since God can transform them anytime and can create them differently. Our knowledge cannot, therefore, be a knowledge of the substance of reality, of that what lies at its ground, but is only a knowledge of the way in which this reality appears to us, of its *phenomenal character*.

Another consequence of this perspective is the rejection of Aristotelian finality: modern scientists do not accept that natural things evolve by being moved by an inner force, which we have to assume in order to understand reality. For all natural processes have an accidental (that is, phenomenal) character, which (as far as we can establish through experience) we can reproduce and manipulate for our purposes. This is one of the theological reasons which have led to the modern scientific revolution.

The modern thinker who has developed a functionalistic conception of the human self is Immanuel Kant. One knows that the Kantian philosophy distinguishes between the thing-in-itself and the phenomenon, and that due to this distinction the concept of knowledge with which Kant operates is one which is valid only for phenomena: we know the phenomena, says Kant, but not also the thing in itself. Our knowledge presupposes a matter of knowledge (sensations) and a form of knowledge (concepts). In the most general meaning, conceptual activity is seen by Kant as an ordering activity, whose object is the "manifold" of intuitions. But this ordering activity presupposes a criterion, an instrument of operating which ensures that the undetermined manifold of intuitions can be arranged in representations, namely it can build a *unity*; certainly, our representations do not appear chaotically in our consciousness. The "meaning" of our representations is the concept, and this is actually a *function*, i. e. the result of the activity of bringing a multitude of representation under one common representation:

> All intuitions, writes Kant, being sensuous, depend on affections, concepts on functions. By this function, I therefore mean the unity of the act of arranging different representations under one common representation (Kant 2, 1881, 60).

This concept of knowledge excludes therefore the idea of a simple mirroring of nature; it excludes the understanding of knowledge as an activity of reflecting the outside of human being in his inner consciousness. Human knowledge is not a result of reflecting, but a *construction*. However, it is not an arbitrary construction.

Without any doubt, in the discourse of the *Critique of Pure Reason* mathematical science of nature occupies a privileged place, for good reasons. Comparing metaphysics with other disciplines which claim to offer knowledge of reality, Kant observes that in metaphysics we have to do with a disagreement between authors, while in the sciences, on the contrary, we encounter an agreement of researchers who work in those sciences. Being himself a great admirer of metaphysics, Kant's intention is actually to discover what makes this agreement in science possible, in order to apply subsequently that condition of possibility also in the case of metaphysics. This agreement means that all scientific sentences are "universal" (i. e. they are admitted by all members of scientific community) and "necessary" (i. e. their admission as being true is the result of a rational constraint, which nobody can elude). The Kantian construction starts therefore from the *fact* of the Newtonian science, associated, of course, with mathematics and logic that have encountered no major transformations through their history. According to Kant, knowledge becomes true knowledge only when it succeeds "to enter on the secure path of a science" (Kant 1, 1881, 369). But such disciplines were in his time only Newtonian physics, mathematics, and logic. Their scientific character consisted in their deductive character. This deductive character is explained by Kant as being grounded in a transcendental structure of the human intellect.

The fact that logic has not encountered any major modifications from the antiquity on made Kant believe that the supreme functions of the intellect – functions which were identified by him with the categories – are given by the types of judgements that were discovered already by Aristotle. The reason for this is that in these judgements we have to do with a synthesis of several representations under a common concept. Kant affirms that there is an unchanging structure of knowledge, a pure and universal reason, which is present in all human beings. The self-knowledge of this reason is accomplished not directly, but through its activity: reason discovers itself in its true identity on the one hand on account of its errors, and on the other hand on account of the sciences that are developed by it during the history. Only by reflecting on these sciences does understanding understand itself, that it is an a priori faculty, which possesses certain "innate" principles. These principles permit understanding to unite (to synthesize) all contents that come from outside, all that is given through sensibility. Only because there exists such "innate" or a priori principles, which determine the synthesis of the manifold of intuition, can we explain, affirms Kant, the possibility of scientific predictions: these predictions are in fact exactly the recognizing of the synthesis which is permanently operated by the intellect and imagination, only on an unconscious level. All scientific judgements express explicitly the (earlier) unknown (i. e. unconscious) operations of the human spirit. And the discovery of "the secure path of a science," namely of these operations of the human spirit, is accomplished in the course of history, i. e. by means of attempts, successes and failures, as with all other human actions. The historicity of man is in this sense an important component of the Kantian philosophy.

If, according to Kant, philosophy, as a science, meant the self-knowledge of reason – philosophy being able to become a science as far as it succeeded to offer an exact and complete description of the human transcendental reason – later, with Fichte, the science of philosophy receives a new meaning: certainly, we are deal here with the self-knowledge of "reason," but now this reason is no more the human and finite one, but infinite reason, which takes the activity form of the "I." At Fichte we encounter a "transcendental deduction," which exposes, as by Kant, the way in which the fundamental structures of empirical consciousness derive, are deduced, or have at their ground the manifestation of the *I*. This *I* of Fichte is a pure activity, which lacks at the beginning of any reflection and thus of any object. Only when the *I* is splitting itself, when the *I* puts from itself a *Non-I* in front of it, in order to retreat in itself and thus to disclose the *Non-I*, only then the object appears as an object of consciousness. All movements, through which the *I*, as an absolute activity, comes to the empirical, common consciousness, were described by Fichte in the theory that he has called *The science of knowledge* (*Die Wissenschaftslehre*).

One of the criticisms brought to Fichte is that the original identity between the empirical I and the absolute *I* (an identity which was also present at Kant, as an identity between empirical consciousness and transcendental apperception), represents a theoretical, pure constructive presupposition, therefore having only a subjective character (in this sense, essentially opposed to science).

The idea that philosophy must have the form of the science is thus a dominant idea of this epoch. But what was to be the object of this science, was, in a certain sense, an open problem. In the Kantian project of metaphysics as science, ontology (as far as it was the discipline where the conditions of the possibility of nature in general were given) was itself a science. And that, because, in the form of an "Analytic," it studied only the human thought (intellect), whose domain was a restricted one and which therefore can be known "once for all." The post-Kantian idealism takes over the ideal of this new type of ontology – and *this* type of ontology is meant by the thinkers of this epoch when they speak about the philosophy as a *Wissenschaft*. In his *Wissenschaft der Logik* Hegel also has set up an ontology, but also in his *Phenomenology of Spirit*, which is and will to be an "introduction" to the *Wissenschaft*, i.e to the ontology.

The object of Hegelian *Phenomonology of Spirit*, is, unlike the "subjective" and "assuring" idealist approach of Fichte, to describe the road which leads from empirical consciousness to universal Spirit and thus to "scientifically" (that is undoubtedly) certify the identity between these two. Empirical consciousness is unable to immediately discover this identity, it cannot understand this identity through its simple affirmation, but has to go through all the configurations of consciousness, in order to understand how this identity appears. Only later, in *Wissenschaft der Logik*, this identity was considered as being known and being able to be developed as ontology.

Thus Hegel does not return to the pre-Kantian metaphysical tradition. The idea of the evolution of consciousness raises the problem of history. History cannot be anymore a simple succession of events, which take place because of accidental causes, but it becomes a series of necessary events having an inner finality. The causality can be accepted only in the (Newtonian) science of nature, where one can distinguish between a cause and an effect, which are completely independent from each other. Actually, this is the fundamental principle of modern mechanical conception, which founds modern physics: objects are material objects (from this reason, modern science is closely related to the materialistic conception), i. e. they have no spontaneity and are completely inert. They can only *react*. When such a causal conception regarding consciousness is abandoned, in other words, when consciousness is no more conceived as a *thinking* and the relation of consciousness to reality is no longer a relation between two things – a relation which presupposes a causal acting and reacting – then the inner evolution of consciousness becomes a finalistic development, where the contents of each stage of consciousness are permanently a purpose of an earlier development, they emerge from this evolution and are no more a local and causal product of some events with which they are no more connected. That is the reason why Hegel insists so much on the concept of "purpose" (Germ. *Zweck*), and is constrained in a certain measure to take over the finalistic conception of Aristotle. However, this finalism is not completely a biological one, because Hegel no longer thinks in a substantialist manner, as Aristotle. Substantialism admits the existence of certain substances of reality, of certain *objective* essences, whereas for Hegel, as for the entire German Idealism, such an objectivity no longer exists. To say solely that Hegel transfers the biological finalism from the sphere of reality into the sphere of consciousness, is, certainly, true, but it remains an "abstract" truth, because it does not explain the possibility of this transfer and does not clarify the reasons that permit or rather constrain Hegel to make such a transfer. When we try to answer these questions we come to a deadlock. For when we try to explain this transfer, as a simple methodological transfer from an objective reality into a subjective one, we fail, and our failure is determined by the fact that we further on conceive consciousness as a substance, i.e. as a thing which exists in a given environment, even if we do not realize that we use substantialist terms in our thought. The concept of substance implies the idea of multiplicity, of plurality of substances, a substance having from this point of view always a *real* exteriority. But at the beginning of the nineteenth century consciousness can no longer be interpreted in this way: after Kant and Fichte, consciousness cannot be understood as having such an exterior environment, because, if one admits this presupposition, one arrives again at the contradictions which the post-Kantian thinkers have emphasized regarding the Kantian conception about the thing in itself. Thus, the substantialism of consciousness must be abandoned. In fact, Kant has already abandoned this substantialism in his *Critique of pure reason*. There he founds the entire thought and pure reason on the concept of *function*. But this concept of function does not permit exterior elements in the field of its values, as sensations are understood as being results of an influence coming from the thing in itself. The concept of function, as Kant uses it, as an act of bringing a "manifold" of representations under a common one, is actually equivalent to the mathematical concept of function, where a given variable receives, due to a certain function, different values. When at Kant several representations are brought under a common one, they all receive, actually, the content of that common representation, they becomes "moments" of the realization of that function. Thus, all representations belonging to consciousness must be thought of in fact as being *exclusively* values posed by the intellect, that is by the transcendental apperception. Certainly, Kant goes not so far in this direction, but Fichte does. Here we encounter a system which is grounded in a single principle, i. e. in a single function, the *I*. Here there is no rest that would not belong to the functional field of the spirit.

The function creates its own field of meanings. These meanings (concepts) cannot be thought of as a result of a passive mirroring of an outer reality, but are the results of an inner and immanent process of generation. And they cannot be thought of as being such a reflecting because of the inner logical unity of the field. In other words, if they would be only reflections of an outer reality, we could not have the inner unity and necessity of our knowledge; that is, we could not speak of an *a priori* content of our sciences, but only of an a posteriori one.

As far as the field of the values of the function "grows" from itself (from within), the function may be actually compared with a biological process – that is the reason why later H. Rombach have described the structures which are grounded on functions as "life" (Rombach, 1971, 85). But the growing in the case of knowledge and of the contents of consciousness is in no way a biological one, because this growing is a growing in "meanings," produces senses and concepts. That is the type of growing that we see at work in the Hegelian phenomenology. All forms, all "configurations" (*Gestalten*) of consciousness are actually generative centres. Around them the entire activity of consciousness on that stage develops. Besides, these forms are, as totalities, each of them, a result of such a generation, as far as they are values which realize the universal function, i. e. the universal Spirit.

Thus we see that an absolute idealism can be built only as a functionalism. And that, because – as the Hegelian model shows impressively –idealism means not only that reality is "in itself" something "ideal," "spiritual," or a "representation" (even if all these predicates are true). The very essence of Hegelian absolute idealism means that the contents of the world do not chaotically lie next to each other, but they are bound together in an inner *necessary* order, an order designated by Hegel as a "rational" one. And the necessity of these contents emerges, in a functional way, from the fact that they receive a meaning. Because, in the case of functionalism, is not important how things are generated, but how these (real or ideal) things emerge as *meaningful* things. Thus, German Idealism (we may say that our sentence counts also for the entire epoch, if we think about Humboldt's idea that every language has a particular *Weltanschauung*, or at Herder's interpretation of language as a totality, whose components express all this totality) represents a movement of thought which stresses that reality must not be understood only in its "substance" (in the sense that this reality is something ideal, perhaps a representation, or, on the contrary, a matter or sensation), but it must be thought of especially in its inner articulation, which is first of all a product of the reason. This articulation of the field of meanings is a process which emerges from itself and to which we cannot attribute any external causes. Therefore we see Hegel permanently being interested in the meanings, in the logical content, in the "Concept," and not in the inner dynamic of the reality, in the forces which move the reality and in its Urge. This interest shows up in the case of Schopenhauer, who therefore arrives not by accident to an irrationalism. Later on, Max Scheler, because he admits an ontological duality formed by Spirit and Urge (in fact separating that what Hegel has called "universal Spirit" in two aspects: the sphere of meanings, the Spirit, and the domain of that what brings about the transformation of reality, the Urge). In the case of Scheler, this Urge has an ontological reality, which is totally *independent from the human consciousness* (due to this conception Scheler leaves the phenomenological approach and exposes himself, with good reason, to the criticism of M. Heidegger). This Urge does not produce meanings, it is a blind force, as Schopenhauer's Will (which also does not create meanings, being on this ground opposed to Hegelian Reason). Hegel does not explain a phenomenon, a fact, in the sense that it is caused by another phenomenon, but he explains it in the sense that it is a *logical* consequence of the former phenomenon, their connection being a logical and not a causal one.

> Consciousness, writes Hegel, is Spirit as a concrete knowing, a knowing too, in which externality is involved; but the development of this object, like the development of all natural and spiritual life, rests solely on the nature of the pure essentialities which constitute the content of logic(Hegel 1969, §10).

Therefore all history and reality are for Hegel logical and rational, which means that the human being can understand them. And only this understanding leads to an immanent purpose. By describing reality as rational, Hegel exposes himself thus to all sorts of criticisms. But these criticisms do not grasp the inner ground of the Hegelian assertion, because they understand Reason in the sense of a moral desideratum, e. g. of the Good. But the main content of the Hegelian Reason is not a moral one, but a logical and ontological one.

Even if the Hegelian discourse is full of biological terms, as "*instinct* of Reason," "Purpose" (*Zweck*), "Form" (*Gestalt*) or even "Cunning of Reason," terms that try to explain the *factual* appearance of the New, the action of the universal Spirit is not a material one, an action which brings about from outside the historical events, but remains a logical action, that is to say it brings them about from inside. Therefore the Hegelian *meta*-physics (i. e. the place where one would expect to see the universal Spirit "in action", independently from the "physical" things, on whom it acts) is called, so paradoxically, *The science of logic*.

The logical dimension of the Hegelian project can be seen in the fact that the process, as Hegel understands it, has a purpose which exists beforehand and it remains present during the entire development. Its presence has not only an orientative character (which would orient from within the movement of this process), but is precisely the function from which the logical content of each stage emerges and which becomes manifest at the end, when the entire logical field of these meanings becomes actual and can thus be known. However, if we do not interpret the Hegelian development in the logical context, but in the sense of something real, objective, the Hegelian project would be only a return to the pre-Kantian stage of philosophy, where one could talk, without any inner critical limitation, about what lies beyond the human reason, i. e. about that what is transcendent in rapport to the human reason.

From this point of view one can understand why the series of the forms of the consciousness is not a *deductive series* in the classical meaning, but it is, paradoxically in rapport with such an understanding, an examining of that what is happened and was thought of in the course of the human history, of the development of the "universal individual" (Hegel 1977, 16) as Hegel says in the *Preface* to his *Phenomenology*. During this examination, the different stages are not only registered, but are actually *endowed with meaning*, are described as sequences of a subjacent sense. What was a real history becomes a history of consciousness, i. e. an *unfolding of meanings* which the "bearers" of these meanings (to use a term from the philosophy of Max Scheler) only "illustrate". They become "moments," as Hegel says:

> In this way the arrangement of the 'shapes' which have hitherto appeared differs from the way they appeared in their own order. On this point we shall briefly observe at the beginning what is necessary. In the series we considered, each moment, exploring its own depths, formed itself into a totality within its own peculiar principle; and cognition was the depth, or the Spirit, wherein the moments which have no other subsistence of their own possessed their substance. But this substance is now manifest; it is the depth of Spirit that is certain of itself, which does not allow the principle of each individual moment to become isolated and to make itself a totality within itself; on the contrary, gathering and holding together all these moments within itself, it advances within this total wealth of its actual Spirit, and all its particular moments take and receive in common into themselves the like determinateness of the whole (Hegel 1977, 414).

Those concrete bearers, which illustrate the "moments," do not have for Hegel any reality, the concrete things lacking of reality, as Hegel points out in the chapter about *Sense-Certainty* of his *Phenomenology*: realitys belong only to *the universal* (*das Allgemeine*), which expresses itself through them:

> In this respect, we can tell those who assert the truth and certainty of the reality of sense-objects that they should go back to the most elementary school of wisdom, viz. the ancient Eleusinian Mysteries of Ceres and Bacchus, and that they have still to learn the secret meaning of the eating of bread and the drinking of wine. For he who is initiated into these Mysteries not only comes to doubt the being of sensuous things, but to despair of it; in part he brings about the nothingness of such things himself in his dealings with them, and in part he see them reduce themselves to nothingness. Even the animals are not shut out from this wisdom but, on the contrary, show themselves to be most profoundly initiated into it; for they do not just stand idly in front of sensuous things as if these possessed intrinsic being, but, despairing of their reality, and completely assured of their nothingness, they fall to without ceremony and eat them up. And all Nature, like the animals, celebrates these open Mysteries which teach the truth about sensuous things (Hegel 1977, 65).

This series has also a circular character: "Through this movement the pure thoughts become *Notions*, and are only now what they are in truth, self-movements, circles, spiritual essences, which is what their substance is" (Hegel 1977, 20).

This circularity means precisely the generation from a common centre of all meanings belonging to the semantic field. However, the functional circularity is manifest also in the Hegelian procedure of formulating a thesis whose truth cannot be seen from the beginning, but only after the unfolding of the entire project. The thesis is here the very principle of selection, which confers a meaning to a totality, it synthesizes it on the ground of a function (in the Kantian sense of the term).

References

Cassirer, E. (1910). *Substanzbegriff und Funktionsbegriff*. Untersuchungen über die Grundfragen der Erkenntniskritik, Verlag von Bruno Cassirer, Berlin.

Hegel, G. W. Fr. (1969). *The Science of Logic*, Preface to the First Edition, translated by A. V. Miller, George Allen & Unwin.

Hegel, G. W. Fr. (1977). *Phenomenology of Spirit*, translated by A. V. Miller, Oxford University Press, 1977.

Kant, Im. (1881). *Critique of Pure Reason*. First Part. Trans. F. Max Müller, London, Macmillan and Co., London.

Kant, Im. (1881). *Critique of Pure Reason*. Second Part. Trans. F. Max Müller, Macmillan and Co., London.

Rombach, H. (1971). *Strukturontologie*, Verlag K. Alber, Freiburg/München.

Taylor Ch. (1975). *Hegel*, Cambridge University Press, Cambridge, London, Melbourne.

BALKAN JOURNAL OF PHILOSOPHY. Vol. 2, Issue 2, 2010

FREUD AND RELIGION

Christfried Tögel

SALUS-Institute, Sigmund-Freud-Centre, Seepark 5, D-39116 Magdeburg
c.toegel@salus-institut.de

Abstract:
The paper deals with Freud's attitude towards religion. It developed from a rather ironic and anecdotal criticism in his youth to fundamental writings about the origin and the future of religion in his later life. Freud believed that "the voice of the intellect is a soft one, but it does not rest till it has gained a hearing" and would overcome religion. Because of the "undeniable conflict" between religion and science Freud expressed his strong opinion, that all believers should be prevented from attending university.

Key words: Freud, Religion, Science.

At the age of 18, Freud wrote to a schoolmate:

> People are wrong to reproach religion for being of a metaphysical nature and for lacking the certainty of sensory perceptions. Rather, religion addresses the senses alone, and even the God-denier who is fortunate enough to belong to a tolerably pious family cannot deny the holiday when he puts a New Year's Day morsel to his lips. One might say that religion, consumed in moderation, stimulates the digestion, but that taken in excess it harms it ... one's stomach is upset. It is most remarkable how certain holidays are distinguished by a very special religious effect on the abdominal organs. Thus the Passover bas a constipating effect due to unleavened bread and hardboiled eggs. Yom Kippur is so lugubrious a day not so much through God's wrath as through the plum jam and the evacuation it stimulates.[1]

This kind of ironic criticism of religion marks Freud's letters and works until the end of his life. With advancing age, Freud's critique becomes ever more serious and calls for consequences. Two years before his death, aged 81, in a letter to Theodor Reik, Freud expressed his strong opinion, that all believers should be prevented from attending university.[2] What influenced Freud to develop this radical position, and how did his negative attitude towards religion shape his life?

Freud's first encounters with religion occurred thanks to his Catholic nanny, "an ugly, elderly although clever woman, who told me a great deal about God Almighty and hell." She brought him to all churches of his hometown and when they came back, little Sigmund "preached and told what the Lord did." [3] This did not embarrass his Jewish parents at all, and Sigmund's early childhood was not dominated by religious conflicts. Later on, things changed. In his book "The Interpretation of Dreams," Freud reported the following story:

> I may have been ten or twelve years old, when my father began to take me with him on his walks and reveal to me in his talk his views upon things in the world we live in. Thus it was, on one such occasion, that he told me a story to show me how much better things were now than they had been in his days. 'When I was a young man', he said, 'I went for a walk one Saturday in the streets of your birthplace; I was well dressed, and had a new fur cap on my head. A Christian came up to me and with a single blow knocked off my cap into the mud and shouted: 'Jew! get off the pavement!' 'And what did you do?' I asked. 'I went into the roadway and picked up my cap', was his quiet reply. This struck me as unheroic conduct on the part of the big, strong man who was holding the little boy by the hand. I contrasted this situation with another which fitted my feelings better: the scene in which Hannibal's father, Hamilcar Barca, made his boy swear before the household altar to take vengeance on the Romans. Ever since that

[1] Freud 1990, p. 62–63.

[2] Reik 1956, p. 118.

[3] Freud 1985, p. 268–269.

> time Hannibal had had a place in my phantasies.[4]

At the time of this conversation with his father, Freud was about 10 years old. The situation described by the father provoked strong emotions. But, of course, at that time Freud was far from theoretical considerations about religions and their relation to each other. 65 years later, however, he used the occasion of Arnold Zweig's first visit to Palestine to wrote to him:

> ... but Palestine has never produced anything but religions, sacred frenzies, presumptuous attempts to overcome the outer world of appearance by means of the inner world of wishful thinking…[5]

But the road to psychoanalytic interpretations of religion was long. Freud's theoretical interest in God, faith and religion began in his second year of studies. He began reading Ludwig Feuerbach and took part in a seminar on the "existence of God" by Franz Brentano. At this point, Freud was still very impressed by the philosopher and wrote to his friend Eduard Silberstein: "Prof. Brentano … is a splendid man, a scholar and philosopher, even though he deems it necessary to support this airy existence of God with his own expositions."[6]

Under the influence of Brentano, Freud considered switching from the medical to the philosophy department. Freud and his friend Joseph Paneth wrote to Brentano and were invited to his home and on walks. On these occasions, they mainly discussed the existence of God. Freud could not escape Brentano's argumentation and admired his skills to discuss theistic arguments with great precision. Under Brentano's influence, Freud "was forced" to become a theist, but only out of helplessness in the face of Brentano's arguments. After several weeks of intense confrontation with the problem of the existence of God, they realized that Brentano had a knack for playing with words and proving the unimportant, while clouding the important with the use of sophistry and paralogisms. Aged 19, Freud summarized that "the existence of God cannot be settled by union debates, parliamentary speeches or speculation, but only by logical and psychological studies, for which not everyone has a taste, no more than for astronomical computations. Hence it is just as wrong to pronounce on the existence of God as on the existence of Neptune."[7] Throughout his life, Freud remained convinced that there is a "undeniable conflict" between religion and science.[8]

This conviction had a direct impact on his family's every-day life. Freud took part in a religious ceremony only once: his own Jewish wedding in Hamburg. Freud gave in to the legal regulations in Austria, which required divine blessing for an official marriage. A civil marriage alone would not have been recognized in Austria. Freud never conceded again.

He forbade his wife to light the Sabbath candles, neither of his three sons, Martin, Oliver and Ernest, was circumcised, and he deliberately violated the rules for kosher meals. So he told his fiancée, for example, about a dinner with Jewish relatives in Berlin where he had ordered cheese after the meat.[9] It is thus not surprising that in Freud's self-designed obituary phrases like "God called him home" are missing, and although an urn burial was unheard of in the Jewish community, Freud was cremated and his ashes placed in a Greek amphora.

Even the way Freud dealt with the deaths in his family had a direct relationship to his atheism. In January 1920, his daughter Sophie died at the age of 27, and just three years later so did her youngest son – Freud's grandson Heinele. In a condolence letter to his son-in-law, Freud wrote: "Here I have spent some of the blackest days of my life mourning for the child. Finally I have pulled myself together and am now able to think of him calmly and talk about him without tears."[10] In this difficult situation Freud – paradoxical as it may sound - found support in his inability to believe in a deity. He confessed to a friend: "The death, as painful as it is, does not overturn any attitude toward life ... Since I am profoundly unbelieving, I have no one to blame and I know there is no place where you can lodge a complaint."[11]

Ten years before the death of his daughter, Freud believed that he had discovered the ground of religion. On January 2, 1910, he wrote to C. G. Jung: "It has occurred to me that the ultimate basis of man's need for religion *is infantile helplessness*, which is so much greater in man than in ani-

[4] Freud 1900, pp. 196–197.
[5] Freud 1960, p. 411.
[6] Freud 1990, p. 70–71.
[7] Freud 1990, p. 111.
[8] Reik 1956, p. 118.
[9] Sigmund Freud-Martha Bernays, 23.3.1886; Sigmund Freud Papers, Library of Congress.
[10] Clark 1980, p. 496.
[11] Freud & Ferenczi 2000, p. 6.

mals. After infancy he cannot conceive a world without parents, and makes for himself a just God and a kindly nature, the two worst anthropomorphic falsifications he could have imagined."[12] It was in helplessness and in the experience of death that Freud saw the reason why man invented evil spirits, which proved to be a precursor of mythological beings and religious deities.[13]

On his deathbed Freud read a book of the Jewish writer Rachel Berdach titled "The Emperor, the Sages and Death."[14] Freud was so impressed by the book that he invited the author to his home.

The plot of the book takes place in the 13th century. The critical moment is a dialog between an Arab doctor and a bishop about the biblical story of the raising of Lazarus. The doctor considers it awful and repellent, thinking that it must be terrible to face death for a second time. After his resurrection, Lazarus is condemned to live with his knowledge of what comes after death.

Another story in Rachel Berdach's book features a parallel to Freud's own death: The Emperor Frederick II is ex-communicated for calling a truce with the Muslims after his crusade. Frederick asks a friendly bishop whether he would still give him the last sacraments. The bishop promises to do so, and keeps his promise. Freud received a similar promise from his family doctor Max Schur: If he could no longer endure the pain he suffered after more than 30 cancer operations, Schur should end his life with a lethal injection. Freud's doctor kept his promise just as the bishop kept his to Frederick II.

Freud's decision to determine the time of his own death is a result of his attitude towards faith and religion. For him, it was always essential to keep things under control. He delegated very little, and even during his many lengthy trips abroad, he requested a daily report on what was happening at home so he could issue guidance on what to do. And the idea that his life was defined and guided by somebody else – even a God - was a great humiliation for him.

* * *

Freud's rejection of religion and religious rites was strict and consistent, and he dedicated large parts of his writings to the subject.

In his paper on "Obsessive Actions and Religious Practices" (1907), he declares compulsion the pathological counterpart of the exercise of religion, that is, he conceives the neurosis as an individual religion and religion as a universal obsessive neurosis.

His major work on religion "The Future of an Illusion" (1927) - unlike Freud's previous publications - does not concentrate on the origin of religion, but on its essence and its future. He underlines that an illusion is not an error, but a faith, essentially motivated by a wish fulfilment - for example, when a poor girl surrenders to the illusion that a prince would come and marry her. Freud expected strong opposition to his theses, but knew how to comfort himself: "In former times it was different. Then utterances such as mine brought with them a sure curtailment of one's earthly existence and an effective speeding-up of the opportunity for gaining a personal experience of the after-life."[15]

As to the future of religion, Freud has two answers: an ideal answer, which is based on rationality, science and deprivation of religion. But Freud considered that the time of the "primacy of the intellect" has not yet come. A realistic perspective, however, he saw in "retaining the religious doctrinal system as the basis of education and of man's communal life. This is a practical problem, not a question of reality-value. Since, for the sake of preserving our civilization … we are obliged to impose on the growing child some doctrinal system which shall operate in him as an axiom that admits of no criticism, it seems to me that the religious system is by far the most suitable for the purpose. And it is so, of course, precisely on account of its wish-fulfilling and consolatory power …"[16]

But Freud never abandoned his hope. In the long term, he believed, reason would overcome religion. He expressed this conviction in the sentence: "The voice of the intellect is a soft one, but it does not rest till it has gained a hearing."[17]

* * *

How can Freud's rejection of religion be reconciled with the fact that he had always felt as a Jew? Freud himself has answered this question repeatedly. In a letter to the members of the Jewish B'nai B'rith Humanitätsvereins, Freud wrote: "What

[12] Freud, Jung, & McGuire 1974, p. 131.
[13] Freud 1912–13.
[14] Berdach 1937.
[15] Freud 1927, p. 36.
[16] Freud 1927, p. 52.
[17] Freud 1927, p. 53.

tied me to Jewry was – I have to admit it – not the faith, not even the national pride, for I was always an unbeliever, have been brought up without religion, but not without respect for the so-called 'ethical' demands of human civilization."[18]

And in response to a welcome letter from a Zionist organisation on the occasion of Freud's arrival in London 1938, he answered: "To my thanks for your greetings welcoming me to England I add the request that you should not treat me as a 'Leader in Israel.' I would like to be regarded only as a modest scientist and not to be distinguished in any other way. Despite being a good Jew who has never denied his Judaism, my absolutely negative attitude towards all religion including the Jewish religion, sets me apart from the majority of our comrades and makes me unsuitable for the role you would like to assign to me."[19]

The last year of his life proved Freud right. Already exiled in London, Freud published his last major work entitled "Moses and Monotheism." The historical assumptions Freud used are no longer tenable. Nevertheless, today this book is of great importance, not because of Freud's historical reconstruction, but because it sheds light on his view of the originating mechanisms of religions.

On the motive and the background of this book, Freud wrote to Arnold Zweig: "For being somewhat at a loss what to do in a period of comparative leisure I have written something myself and this, contrary to my original intention, took up so much of my time that everything else was neglected. [...] But let me explain what has happened. [...] Faced with the new persecutions, one asks oneself again how the Jews have come to be what they are and why they have attracted this undying hatred. I soon discovered the formula: Moses created the Jews. So I gave my work the title: *The Man Moses, a historical novel...*"[20]

But only the first part reads like a novel. Here, Freud develops the theory that Moses was not a Jew but an Egyptian. That the name "Moses" is an Egyptian one, had indeed been remarked by scholars before him, but out of respect for the biblical tradition no further conclusions have been drawn. Freud supported his thesis with an analysis of the "myth of the birth of the hero." He believes that Moses was an aristocratic Egyptian who was abandoned by his parents. The legend later made him a Jew.

In the second part, Freud discusses the question, "what could induce an aristocratic Egyptian – a prince, perhaps, or a priest or high official – to put himself at the head of a crowd of immigrant foreigners at a backward level of civilization and to leave his country with them."[21]

And Moses was not only the political leader of the Jews residing in Egypt, but also the founder of a new religion. Why? Freud believed that Moses was an ardent follower of the religion of Akhenaten, known also as Amenhotep IV (1364 to 1347 BC), who introduced monotheism through religious reform. After Akhenaten's death, a kind of counter-reformation started, and Moses faced the dilemma of whether to become a turncoat or to emigrate. As a born leader, he did not chose the quiet, lonely journey into exile, but elected a people that he could take with him and that would make Akhenaten's religion that he admired so much, their own one: the people of Israel. Through them, Moses became the founder of a new religion. Not God, but Moses has chosen this people. He wanted to make it equal to the Egyptians and shook off captivity. The new monotheistic religion should support this effort. After the Exodus from Egypt, some of the Israelites – according to Freud – rebelled against Moses. The story of the dance around the golden calf might be an indication for that. As a result Moses was murdered - just as the brothers of the totem clan murdered their father. As a consequence of the murder, the Israelites denied that Moses introduced monotheism and later attributed it to the patriarchs.

The third part deals with the consequences of the murder for the history of religion. The conscience-stricken guilt of the Israelites turned into hope that the murder could be undone – perhaps by a Messiah. Later on, in Christianity this Messiah turned indeed into a saviour from sin. In this way, Freud interpreted Christianity as a delayed reaction to the guilty conscience of the Israelites for killing Moses. The generalization of this act leads to "welcome means of exculpating God: they deserved no better than to be punished by him since they had not obeyed his commandments. And, driven by the need to satisfy this sense of guilt, which was insatiable and came from sources so much deeper, they must make those commandments grow ever stricter, more meticulous and even more trivial. In a fresh rapture

[18] Freud 1960, p. 366–367.

[19] Sigmund Freud-Israel Cohen, 14.6.1938, translation by Michael Molnar.

[20] Sigmund Freud-Arnold Zweig, 30.9.1934, translation by W. D. Robson-Scott.

[21] Freud 1939, p. 18.

of moral asceticism they imposed more and more new instinctual renunciations on themselves and in that way reached – in doctrine and precept, at least – ethical heights which had remained inaccessible to the other peoples of antiquity."[22]

But the Jews have not become Christians because Moses has shaped their national character so strongly that they have always retained their special position. Freud describes these characteristics in detail and particularly stresses *self-reliance, tenacity* and *passion for intellectual pursuits*. This as well as the fact that Jews think of themselves as "chosen people," made them an ideal scapegoat and led to the emergence of anti-Semitism. For nearly 2000 years, Christianity blamed the Jews for all social and natural evils.

Freud's concerns that Jews might be offended by his book were justified. Many Jews warned him not to publish it, and after its publication Freud received threats and insults. A Jew from Boston called him an idiot and ended his letter with the words: "Renegades like you, we had thousands. We are glad that we got rid of them, and we hope soon to get rid of you too. It's just a shame that the gangsters in Germany have not put you in a concentration camp. That is where you belong."[23] But only a few months before his death, Freud refused to be intimidated any longer. About his Moses book he wrote: "Neither in my private life nor in my writings have I ever made a secret of my being an out and out unbeliever. Anyone considering the book from that point of view will have to admit it is only Jewry and not Christianity which has reason to feel offended by its conclusions. For only a few incidental remarks which say nothing that hasn't been said before, allude to Christianity. At most one can quote the old adage: 'Caught together, hanged together' … Needless to say, I don't like offending my own people either. But what can I do about it? I have spent my whole life standing up for what I have considered to be the scientific truth, even when it was uncomfortable and unpleasant for my fellow men. I cannot end up with an act of disavowal."[24]

* * *

However, I want to end my paper in a conciliatory manner, and say a few words about the friendship between Freud and the Swiss protestant pastor Oskar Pfister (1873-1956). Right at the beginning of their thirty-year-long correspondence, Freud apologized to Pfister for not having thought of the valuable insights from psychoanalysis for spiritual guidance. The reason for this failure was that as an "evil pagan," he did not think in terms of pastoral care."[25]

Because of their friendly relationship, Freud even delayed the publication of his book "The Future of an Illusion." In a letter of October 16, 1927, Freud wrote to him: "In the next few weeks a pamphlet of mine will be appearing which has a great deal to do with you. I had been waiting to write it for a long time, and postponed it out of regard for you, but the impulse became too strong. The subject matter is my completely negative attitude to religion … I feared, and still fear, that such a public profession of my attitude will be painful to you. When you have read it, let me know what measure of toleration and understanding you are able to preserve for the hopeless pagan."[26]

Pfister's response is characteristic for the relationship between the two men: "As for your anti-religious pamphlet, there is nothing new to me in your rejection of religion. I look forward to it with pleasurable anticipation. A powerful-minded opponent of religion is certainly of more service to it than a thousand useless supporters."[27]

References

Berdach R (1937). *Der Kaiser, die Weisen und der Tod*. Saturn-Verlag, Wien.

Clark R (1980). *The Man and the Cause*. Random House, New York.

Freud S (1900). *The Interpretation of Dreams* S.E. 4, 5.

Freud S (1912–13) *Totem and Taboo: Some Points of Agreement Between the Mental Lives of Savages and Neurotics* S.E. 13.

Freud S (1927). *The Future of an Illusion* S.E. 13.

[22] Freud 1939, p. 134.
[23] N.N.-Sigmund Freud, 26.5.1939 [Freud Museum London].
[24] Freud 1960, p. 453.

[25] Oskar Pfister-Sigmund Freud, 09.02.1909. The International Psycho-Analytical Library, 59:15-17.
[26] Oskar Pfister-Sigmund Freud, 16.10.1927. The International Psycho-Analytical Library, 59:109-110.
[27] Oskar Pfister-Sigmund Freud, 21.10.1927. The International Psycho-Analytical Library, 59:110-112.

Freud S (1939). *Moses and Monotheism* S.E. 23.

Freud S (1960). *Letters of Sigmund Freud*. Basic Books, New York.

Freud S (1985). *The complete letters of Sigmund Freud to Wilhelm Fliess 1887–1904*. The Belknap Press of Harvard University Press, ed. by J. M. Masson. Cambridge, Massachusetts.

Freud S (1990). *The letters of Sigmund Freud to Eduard Silberstein 1871–1881*. Belknap Press of Harvard University Press, ed. W. Boehlich. Cambridge, Mass./London.

Freud S, & Ferenczi S (2000). *The Correspondence of Sigmund Freud and Sándor Ferenczi*, Volume 3, 1920–1933. Edited by Eva Brabant, Ernst Falzeder, and Patrizia Giampieri-Deutsch. Belknap Press of Harvard University Press, Cambridge, Mass.

Freud S, Jung C G, & McGuire W (1974). *The Freud/Jung letters; the correspondence between Sigmund Freud and C. G. Jung*. Edited by William McGuire. Princeton University Press, Princeton, N.J.

Reik T (1956). *The Search Within. The Inner Experiences of a Psychoanalyst*. Farrar, Straus & Cudahy, New York.

BALKAN JOURNAL OF PHILOSOPHY. Vol. 2, Issue 2, 2010

THE FORMAL STRUCTURE OF EXPERIENCE IN CARNAP'S *AUFBAU*

Ioan Biris

West University of Timisoara, V. Parvan St. 4, Timisoara, Romania
ioan.biris@gmail.com

Abstract:

The transformation of the relations between reflection and reality and between concepts and their correspondent objects into themes represents even in the present a field for most heated discussions. The joining of conceptual schemes corresponding to the intellect and reality represents a problem which is still to be solved. A solution to this problem was proposed by R. Carnap in his extremely ambitious project from Der logische Aufbau der Welt (1928). Overlooked for a long time, this work has returned to the philosophical spotlight in recent years, because philosophers finally realized that Carnap's project comprises almost all the major themes of contemporary epistemology, such as the relation between theory and reality, between concepts and experience, the major lines of a phenomenology and of a "logic of experience"or the status of language and of the concepts of science. Situated at the confluence of some philosophical traditions such as neo-kantian philosophy, the logical analysis of Russell and Wittgenstein, but also phenomenology, Carnap's program in Aufbau starts from some premises such: the need to control experience (= major stake of any scientific knowledge); the orientation towards form of the modern concept of scientific knowledge (under Kant's direct influence); the assertion of a concept of knowledge mostly relational or structural (issue resulted in the orientation towards form of modern knowledge); the need for a methodology engaged in a constructive way and modelled on mathematical thinking; our reporting to the instance of sensitivity in the formation of experience (under E. Mach's influence).

Key words: new logic; similarity relation; structural description; formal reconstruction.

One can't understand Carnap's project in *Aufbau*[1] without taking into account the context of the "new logic" of the twenties from the last century, that is of the means in which the concepts of scientific reflection are approached, on the one hand, and on the other hand if one neglects the ways in which one can control the sensitive data of experience. These are the two instances without which one cannot grasp Carnap's approach. From the analysis of the scientific concepts – writes Carnap – one has concluded that, all the concepts belonging either to the fields of natural sciences, or to psychology and social sciences according to a common classification, lead to a common basis: they are reduced to initial concepts (*Würzelbegriffe*), which are related to the <data>, or immediate contents of living[2]. Only through this "reduction" towards "data" one can aim at an uniform science (*Einheitswissenschaft).* First of all the psychic concepts can be "reduced" to their "data" because these ones refer to the psychic phenomena of the subject who possesses the knowledge. Starting from the psychic concepts one can make the concepts which are outside the psyche and from here on one reaches the scaffolding of all other social-scientific concepts. In this way a genealogy tree of concepts results (system of constitution) in which each concept of science must have its own place, in conformity with its deduction from other concepts, and in the end from the given data (*Gegebene*)[3].

It is clear that for Carnap the role of logic is essential in view of a uniform scientific achievement and of rational and coherent explanation of reality. Having this in mind while he edited his work *Aufbau*, that is before 1928, he followed the very convincing example of *Principia Mathematica* (by Russell and Whitehead). One can notice that in his project Carnap had at least two reasons[4] to follow "the new logic": 1) constituted as a discipline (after the model from *Principia Mathematica*), the logic offered an eloquent example of the achievement of a rational reconstruction of mathematics; 2) in the form of an analytical instrument, logic could be used not only for mathematics but it also allowed an ex-

[1] Rudolf Carnap, *Der logische Aufbau der Welt*, Meiner Verlag, Hamburg, 1998.

[2] Rudolf Carnap, *Vechea şi noua logică*, (editions and translation in Romanian by Alexandru Boboc), Editura Paideia, Bucureşti, 2001, p. 32.

[3] *Ibidem*, p.32.

[4] Pierre Wagner, Le contexte logique de l'Aufbau. Russell et Carnap, in Sandra Laugier (ed.), *Carnap et la construction logique du monde*, Vrin, Paris, 2001, p. 17.

tension to all scientific knowledge. In this respect Carnap was convinced in that period that it is even more clear in the present that the theory of knowledge, which isn't in its essence anything but applied logic, can't be deprived of logistics just like physics can't deprive itself of mathematics[5].

1. New Logic and Rational Reconstruction of the World

But what kind of logic are we talking about? In *Aufbau* Carnap is rather skimpy with the explicit definitions related to this subject. It is clear that he has in mind as a pattern the reference work *Principia Mathematica*, a work in which he admired especially the method of rational reconstruction of mathematics. But one must not forget other influences also, such as the *Tractatus* of Wittgenstein, the conventionalism of Poincaré of the philosophy of science or the neo-kantian philosophy. On the other hand, without having the intention of continuing the discussions in this direction, there is a certain originality of his thinking and of his approach, since he is not a mere imitator. The new logic of which Carnap is talking is that logic developed by the mathematicians in the last decades before the edition of *Aufbau*. At least this is what he remembers in the preface of the work. It sprung from the need to overcome the crisis of the foundations of mathematics, a crisis which traditional logic could no longer face. The new logic allowed not only to overcome this crisis which is a result on the negative side but it also represents a step on the positive side, a new starting point for an overall change of philosophy itself. In other words, the new logic wasn't just a mere pattern for Carnap, but at the same time an important instrument for the radical renewal of epistemology and philosophy in general.

Of course, Carnap was not the only person excited by the potential of the new logic to renew scientific and philosophical conceptual thinking. In 1914 Russell published *Our Knowledge of the External World*[6], and a long time before it in *Begriffsschrift* (1879), Frege conceived the first formal system which questioned the procedure of the traditional logic to consider judgement simply an act in which a predicate was assigned to a subject. Frege proposed the analysis of judgement from the perspective of the notions of function and of object and he introduces universal and existential quantifiers, imposes the distinction between meaning and meant the theory of propositional functions etc. Carnap will particularly appreciate Frege's contribution in the quantification field for what he calls "the new logic", and some years later (in fact two years after *Aufbau's* publication), in the study *Die alte und die neue Logik* (published in the first number of *Erkenntnis*, 1930–1931), the Vienna philosopher is excited by the symbolical method of the "new logic" which assures the calculating of sentences, of functions, which guarantees the rigor of the conclusion, that is it doesn't allow anymore the intrusion of some unnoticed premises, as it often happens in the deductions achieved in the language of the words. If symbolism is connected more to the form of the presentation, Carnap tells us that "the new logic" is also imposed through a large development of the field of logic, the new domains and most important being considered the logic of relations and the theory of propositional functions. In the traditional logic one finds a unique form of the propositions (judgements) that is the predicative form. When one says "Socrates is a man" we attach a predicate-concept (man) to a subject-concept (Socrates). In exchange, in the case of the relation propositions (for example "*a* is bigger than *b*") "a relation is added to two or more objects (if we want: to many concepts of subject)."[7].

The relation sentences –Carnap emphasizes – are necessary to mathematics, but not only to it. The "extralogic" fields need in the same degree the new logic of relations because the past limitation only to predication sentences has lead many times to fatal errors. Following Russell's demonstration, Carnap accepts the fact that for instance the notion "absolute space" is the result of a logic error, not neccesarily of a physical science error. The constraint of the traditional logic to form a statement only in a predicative form implies the definition of the space only as "the place"of an object. Leibniz – who noticed the importance and possibility of the relation sentences – succeeded to understand in a proper way the notion of space as a "relative space". That is "the place" of a body is not the thing which expresses in a proper way the notion of space but the "relation" of that body with other bodies. It is that "relation" that represents "the elementary fact" for defining space and not "the place" attributed through a predication judgement. From the moment one accepts the new logic as the most adequate instrument for rational reconstruction of the world (not only of mathematics) the next extremely important step is that of defining the objects of reality. In 1914 in *Our Knowledge of the External World*, Russell aimed to

[5] Rudolf Carnap, *Vechea şi noua logică*, p. 21.

[6] Bertrand Russell, *Our Knowledge of the External World*, Open Court, London/Chicago, 1914.

[7] Rudolf Carnap, *Vechea şi noua logică*, p. 22.

achieve the dream of the empiricist epistemologists[8], that of succeeding to explain the construction of the exterior world from the sensitive impressions, this project being resumed with all seriousness[9] by Carnap in *Aufbau*. Such a project implies, as Quine confesses, besides a very good knowledge of the new logic, a fecund philosophical imagination and also a superior understanting of psychology and of physics and of science in general. In the process of the logical reconstruction, the introduction of any new object implies the resort to definitions, precisely to the definition of the name of the object that is introduced.[10]. But the model offered by Russell and Whitehead in *Principia Mathematica* did not allow any definition procedure. A constructive approach will be needed, in which, just as analysis and synthesis correspond to each other and assume each another, construction and "reversion"[11] presuppose one another. In conclusion Carnap distinguishes three types of definitions: explicit definitions, contextual ones (explicit definitions in a large sense) and implicit definitions. Important problems are raised only by the latter ones because the implicit definition doesn't aim at a determined object (or a concept) but at a "class", an undetermined object, or an " inappropriate" concept.[12]

What is perhaps more important regarding the definitional process in *Aufbau* is represented by the structural definitions. This step is made by Carnap in order to account for the situation in the empirical sciences, a situation for which the pattern in *Principia Mathematica* no longer offered solutions. In order to characterize empirical entities in a formal way , Carnap admits that an object can be defined by describing the relations which it has with other objects. In this way, Carnap fructifies the new field of logic (= the theory of relations) for the entire field of empirical knowledge. These structural definitions resemble up to a certain point the implicit ones, used particularly by Hilbert in the domain of axiomatic geometry, but can't be reduced to the latter ones since a defined structural description characterizes but a single object, an empirical one that is an extra-logical one.[13] With this type of structural definitions Carnap's position breaks away a lot from that of Russell, the Vienna philosopher placing himself in another philosophical perspective. Russell talks about the posibility of a "direct knowledge" of the entities (for instance, the direct knowledge of colors and of taste), while Carnap aims to overcome the subjective aspects of elementary experiences with the help of structural descriptions. In his theory of rational reconstruction, Carnap considers the objects as secondary in relation to structural relations, here being closer to Wittgenstein in *Tractatus*, for whom the sentence is a description of a state of facts. (4.023). To Wittgenstein the *sentence* had been introduced in *Tractatus* through the term *Bild* (image). "This term can be deceiving because many associate to the expression *image* a photo, a reflection in the mirror, a portrait, an icon. To Wittgenstein the word stands for patterns, for *structural descriptions* of some states of things. He follows in this respect the german physicist Heinrich Hertz, who, in his book from 1894 about the principles of mechanics, characterizes the constructions of this theoretical science as images or models of the physical systems."[14]

2. Carnap and the German Scientific Tradition

Without any doubt, the problem of the sources used by Carnap in the elaboration of his project in *Aufbau*, but also of the influences which came from different directions is a complex one, and this is not the proper place for it to be investigated in detail. We have to remember though, that besided the influences from Russell and Frege, from Wittgenstein and other logicians, Carnap's conception can't be understood outside the german philosophical tradi-

[8] W. V. O. Quine, *From Stimulus to Science*, p. 10.

[9] In his Intellectual autobiography, Carnap confesses that when he read Russell's work, *Our Knowledge of the External World*, he had the profound impression that the appeal to continue this project was addressed to him in a direct, personal way and that to search in this spirit was assigned to him as his own task (See P. A. Schilpp, ed., *The Philosophy of Rudolf Carnap*, Open Court, 1963).

[10] Rudolf Carnap, *Der logische Aufbau der Welt*, § 38.

[11] In the specialized literature, especially due to Quine's reductionist interpretation (starting with *Two Dogmas of Empiricism*), the term used Carnap in German for this intervention is "*zürükführung*" – was equated to that of "reduction", which we consider to be an overstatement and it distorts Carnap's idea. A more appropriate understanding of the spirit of Carnap's approach would be the using of this term in the meaning of " reversion", which expresses a "re-direction", a "re-coming" to the starting point, to the initial state. For example, in conformity with the basis idea in *Aufbau* , according to which all the objects and all the notions can "re-come' to the basis entities (not necessarily reduced to them), an object *a* can be realized starting from objects *b* and *c*, that is the statement about *a* can be "re-directed" through the reversion operation in the statements about *b* and about *c*.

[12] Rudolf Carnap, *Der logische Aufbau der Welt*, § 15.

[13] *Ibidem*, § 15.

[14] Mircea Flonta, *Gânditorul singuratic. Critica şi practica filosofiei la Ludwig Wittgenstein*, Editura Humanitas, Bucureşti, 2008, p. 129.

tion, especially that of Kant and the neo-Kant philosophy, but also the german scientific tradition in the physics domain where the works of Helmholtz, Boltzmann, Hertz and Mach imposed themselves. Even though their influence was not necessarily direct as C. Moulines emphasizes[15] at a certain moment but through other names such as Dingler, Driesch, Jacobi, Ziehen and others, cited in *Aufbau*, it is nevertheless true that their ideas can be felt in Carnap's construction. Hertz's idea that in the end the system of mechanics is composed of "images" structured in a deductive way, that these images are patterns of our representations about the real things, was almost a "common good" among the German scientists and among the philosophers interested in the theory of science. Along the same lines was Boltzmann's advice to consider the concepts and statements of physics as "mental images", patterns which describe the facts known through experience. Regarding Hertz, who becomes the most influent physicist in this direction, in his work *Die Prinzipien der Mechanik in neuem Zusammenhange dargestellt* (1894), he starts from the methodological requirement of deriving the future from the past, brings to our attention the fact that we form the different internal images or symbols about external objects, images which can be developed by means of other past images, considered as patterns. With the important observation that for the same objects we can have different images, which means that in contemporary terms, there is a "theoretical subdetermination of experience", Hertz prefigures in this way the successive generations of positivists[16] for which, just like for Pierre Duhem, for instance, truth must not be searched in singular sentences, but in much larger or more restricted systems of representations. But what status do these "images" have, these patterns about real things? In Hertz's view they need to answer to three requests[17]: a) the logical request, that is the request that these images be allowed (*zulässig*), that they don't contradict the principles of logic; b) the empirical request, that is the condition that the images be appropriate (*richtig*), that they be in concordance with reality, with the relations in reality; c) the pragmatic request, that is the request that the images be adequate (*zweckmässig*), as simple as possible, that they don't contain, as far as it's possible, empty or not necessary relations. If the first request can be established unequivocal by the formal observation of the means in which an image breaks or not the principles of logic, and the second one can be established also unequivocally by means of experience. The third request puts us in a delicate position, because establishing the situation of adequacy of an image doesn't represent in general a process of unequivocal decision, but different opinions of researchers and only a gradual control of more images on the time axis can offer us the more adequate image.

In order to obtain a more-or-less complete picture of the context in which Carnap's projects are stated in *Aufbau* one must not also forget the Vienna philosopher's admiration for the *Gestalt* philosophy. The *Gestalt* psychology has its originality, as J. Piaget notices at a certain moment, in the fact that it questions the existence of sensations as prior psycological elements, considering that what is given initially is a configuration, a whole which is to be explained[18]. When we see a tree, for instance we don't perceive firstly its leaves or its branches, we perceive the tree as a "form", as a "structure". But the *Gestalt* notion (as brought to our notice by Wolfgang Koehler) can have two meanings in German: form, as an attribute of things, and form as an objective and independent entity. That is why Gestaltism developed in two directions: one which, under the influence of phenomenology, considers that form is constitutive of the structural activity of the perceptive conscience; and the other one which accepts the form as being exterior to the subject, that is it is composed only of environment variables (cf. Skinner's interpretation).

In that period of time, Carnap will follow the first direction which was in consonance with E Mach's influence. One must not forget the fact that Rudolf Carnap was well aware of the crisis of physics foundations of the first decades of the twentieth century. This crisis brought to the spotlight a conceptual challenge from theoretical physics and brought conceptual and methodological renewals. In this way, Carnap found himself among those who, like Mach, hoped that a physiologist or psychophysiologist epistemology will be more capable to assure more solid foundations to scientific knowledge[19]. Mach, in his *Analyse der Empfindungen* (1886), had already tried such a project of substantiation of science in general, and of physics in particular, on the more profound and solid (considered

[15] C. Ulises Moulines, La (re)construction formelle de l'expérience. Carnap et Nicod, in Sandra Laugier (ed.), *Carnap et la construction logique du monde*, p. 44.

[16] Giovanni Boniolo, Maria Luisa Dalla Chiara et al., *Filosofia della scienza*, Raffaello Cortina Editore, Milano, 2002, p. 70.

[17] Mircea Flonta, *op. cit.*, p. 159.

[18] Ioan Biriş, *Totalitate, sistem, holon*, ediţia a doua, Editura Universităţii de Vest, Timişoara, 2007, p. 40–42.

[19] See also C. Ulises Moulines, *op. cit.*, p. 45.

by some) basis of biology, that is on the analysis of sensations. Mach's project was also appreciated in that period by the father of phenomenology, Husserl, who wrote at the beginning of the twentieth century in the following way: many of the most fecund methods characteristic of the most advanced sciences can be brought to a grateful understanding only by relating them to our psychic features. Mach expresses himself admirably in this respect[20]. The basis idea of such an approach was that the starting point in the scaffolding of scientific knowledge must be represented by a pattern of the phenomenal world of sensations, a pattern from which one can build towards the macroscopic physical objects situated in space and time.

At this moment we have merely sketched the main backgrounds of the scientific-intelectual context in which Carnap shapes his project in *Aufbau.* And it isn't just about a simple project, but a "super-program", as Moulines also states, because Carnap's work contains in fact more partial programs: the program for using the "new logic" in view of its transformation into into an efficient instrument not only in the field of mathematics, but for the entire field of scientific knowledge; the program to build a new theory of knowledge based on the physiology of senses and the psychology of perception, that what is known today as under the name "naturalized epistemology" (especially under Quine's influence), but the roots of such a project can be detected in Helmholtz; the program to math experience, that is to make a mathematical pattern of sensory experience which represents a passage towards physics.

With such an ambitious "super-program" it is no wonder that, retrospectively, in his intellectual autobiography, even Carnap saw the human subject in *Aufbau* as a sort of Turing machine which builds the world by manipulating symbols, and contemporary exegesis remarks with a certain surprise that the newest themes in the philosophy of mind – like the problems of Self, of dualism and of intentionality – had already been approached in a many ways a long time ago before the Vienna philosopher's work had appeared.

3. The Mathing of Experience

Let us see how the objective formulated by Carnap at the beginning of his work can be attained in *Aufbau*, that is to build a logical system of the concepts of science, a pyramidal system which resembles a storied building. Russell's influence from the theory of types inspires Carnap to conceive the passing from the lower level to the superior ones through a transitivity operation since all the superior levels are built starting from the basis level. In Moulines' opinion it isn't about a practical application of Russell's theory, but about a construction belonging to the theory of assemblies of Bourbaki: that is one passes from store n to store $n+1$ by composing assemblies from subassemblies of entities from n, which certain relations establish. By using the acquisitions of the "new logic" Carnap thinks that one can reconstruct succesively the objects from the basis level of sensitivity, the auto-psychological objects, then the physical ones resulting from the formal-logic combination of the sensitive data, in order to move on to hetero-psychological construction from the intersubjective level and in the end to attain cultural objects.

The starting point in building the logical system of scientific concepts is represented by the elementary sensitive experiences (*Elementarerlebnisse*), which represent for the system the originary elements (*Urelemente*), but these are "elements" in a Gestaltist meaning, that is structures, forms with a holist character. Seen in this way, for Carnap the elementary experiences are not subjective feelings which cannot be communicated, they are not psychological elements felt with a unique quality, instead they are overall configurations of "entities" of "points" of the sensitive and the relations between them. One emphasizes these relations because "the points" of the lived experience stay un-analysed. This theory may seem unclear but the Gestaltist examples can illuminate it, and in the end Carnap's idea becomes more intelligible. So, at a certain point, Carnap states that a certain sensitive quality, for example a tone or a color may appear as an "element" or a "quasi-element" [21], but what interests us in such an experience is the relation, the logical function of qualities. But what relation are we talking about? According to the Gestaltist theory, for example, in a color we can find two or more shades of the same color but a certain shade can be isolated in that configuration only by comparing it to the other shades.

Gestalt only exists through the relations between its components, the relations being in priority in relation with the elements. If the configuration and relations change, the object also changes although the elements remain the same. When one changes the relations between the keynotes of a melody, the melody itself changes; if one modifies the relations between the elements of a painting, the

[20] Edmund Husserl, *Cecetări logice*, vol. I (Romanian translation by Bogdan Olaru), Editura Humanitas, București, 2007, p. 279–280.

[21] Rudolf Carnap, *op. cit.*, § 76.

image also changes etc. Coming up to the above example, that of distinguishing between the shade of a color and other shades from the same color, it is clear that this operation of distinction can take place only through the following two steps: 1) the existence of a relation between shades since only if we base ourselves on them can we relate one shade to the other; 2) discrimination of one shade can be done only by comparing the background with the other shades, background similarity being given by the color to which the shades belong.

That is why Carnap situates the similarity relation at the basis level of the system, at the level "zero" of the construction. It is a diadem relation, interprets Moulines, that is " on the temporal level of the phenomenal reality we are having an asymmetrical relation, the *memory (the* subject or Turing machine can remember in moment *t* that they had a partially similar experience in moment *t – s*); structurally speaking, we are having a symmetrical relation, a *partially similar* one which part of the sensitive qualities. The diadem relation of the *remembering as similar* (*Ähnlichkeitserinnerung*) is in this way the cornerstone of the entire construction. By taking over this relation, Quine will describe it in the following terms: "Carnap's basic relation between elementary esperiences was *remembering as similar.* I shall call it R. One elementary experience, x, bears R to another, y, if x includes a memory of y as partially resembling x"[22]. Quine agrees in his 1995 work, together with other contemporary American philosophers, that Carnap's project in *Aufbau* is again of actuality and that the evolution of science and of epistemology imposes the reconsideration of the Vienna philosopher's stakes. Carnap didn't need, emphasizes Quine in his last work at the end of his life, supplementary predicates to denote elementary experiences, since an elementary experience can be found in the *R* relation with any other experience, in the end with everything else. In Quine's terminology, the science stimuli are exactly the elementary experiences in Carnap's language.

In the constructionist theory of *Aufbau*, the objects (or concepts) from each level are "constructed" from the objects (concepts) of the previous level. In this process one has to take into consideration at least the following main aspects: a) the characteristics of the basis level, on which all the upper levels are based; b) the determination of the ascent forms through which we rise from one level to the other; c) investigating the way in which one can build different types of objects can be built through repeated applications of the ascending forms; d) the form of a system resulting from such a construction. While the basis level from which one starts, represents extralogic entities, the other levels are logical constructs, respectively *forms* of ascent in the construction, forms of the object and forms of the system. A system built in this way is not one of a classifying type, but a derivational, genealogic one. Starting from the similarity relation, the derivational process emphasizes the partial similarities, the similarity circles the quality classes, the partial identities, the similarities between the qualities, the sensory classes etc. As Quine emphasizes[23], the partial similarities are defining elements for the similarity relation itself, and the similarity circles are larger classes of elementary experiences in which each experience is partially similar with the other. And a class of qualities is the class in which all the elementary experiences which compose it, express the given quality being more restricted than the similarity circle. In this way, a class of qualities will always be the common part of the similarity circles, which it covers in a reasonably diverse experience, each of the five senses of the knowing subject offering the largest classes of qualities.

If one takes into consideration the complex program of Carnap only under the aspect of physics, of course it isn't a singular case but it belongs to a series in which it is preceded by Brüke's, Du Bois-Reymond's and Helmholtz's initiative and E. Mach's conception to create *Berliner Physilalische Gesellschaft* (in 1845). The originality of Carnap's construction is given in a great degree by the means of mapping experience, like it appears in *Aufbau.* As we have already suggested, besides the new logic used on a large scale in the construction of the system, under a mathematical aspect Carnap uses massively the theory assemblies, but he doesn't limit himself to it but he will appeal to topological patterns. The phenomenal world is built by Carnap starting from remembering as similar, on which partial similarity is built, then the similarity circles, the sensitive qualities and the quality classes. In Moulines' opinion, the special form of mapping experience done by Carnap is "absolutely original" and it "represents probably the most creative and admirable part of *Aufbau*"[24]. It is about the idea to build a mathematical pattern of the phenomenal world which has to be independent, logically independent from physics and physiology. In a primitive manner, Moulines considers that the idea can be found in Mach also, but also in Poincaré and Russell. The closest to Carnap's ideas seems to be the French

[22] W. V. O. Quine, *From Stimulus to Science*, p. 10.

[23] *Ibidem*, p. 11.

[24] C. Ulises Moulines, *op. cit.*, p. 49.

Nicod (though it seems that Carnap didn't read his works).

For Poincaré, on a more specifically Kantian philosophical plan, the topology is especially interesting since it has a synthetic character a priori, being a systematization for the form of spacial intuition. The mathematician and thinker Poincaré is interested in a space of sensitive experiences, a "phenomenal" space (independent from the physical space) in which our sensations are placed. Moulines sees in Poincaré's project the first intention to assign a mathematical structure to the conglomerate of sensations without supposing physical knowledge. Russell's attempt, from this perpective doesn't rise to the level proposed by Poincaré, since the British philosopher isn't interested in the mapping of experience but instead in the construction of a "bridge" between sensitive elementary experiences and physical objects. In contrast, Jean Nicod, a Russell's disciple, in his work *Géométrie du monde sensible* (1923), though he seems to pose himself the same problem of the "bridge" between the sensitive and the physical, he will go in the direction which will preoccupy Carnap, that of the formal reconstruction of the space of sensitive experiences. Interested in a geometry adequate to the phenomenal world of sensitive experiences, Nicod will consider that for such a program the indivisible points of Euclidian geometry are not adequate as "prime elements", but it is necessary to have entities of a special kind, which should not be "infinitely small"; they should be a sort of sensory constellation, volumes (Moulines calls them "sensory bubbles"), of topological spaces. In a way very close to Carnap's ideas, Nicod establishes three types of relations in the phenomenal space, the purely temporal relations, qualitative similarity and local similarity. By Nicod's project and by the complex construction of Carnap, the sensitive phenomenal world starts to be thought as a geometric universe, the similarity relations allowing to built topological spaces or spacial fields, fields of qualities.

4. Carnap, Quine and Goodman

It is interesting that such a project remains attractive only for Central European philosophical thinking, thing which explains in a great degree the trajectory of Carnap's work, and why it passed through a certain degree of oblivion until the post-war period. The return in the post war actuality of the work *Aufbau* is caused in a great degree by Quine and Nelson Goodman. For Quine, Carnap's work is a very actual one and with viable multiple facets[25]. For Goodman, Carnap's project must be criticized in a severe way and in the end removed. The approaches that are intended to exploit the Carnapian project and those which intend to reveal its lack of basis, start from the basis element of Carnap's construction; that is the similarity relation is situated at the basis of the system. On an exaggerated nominalist position, Goodman[26] tries to demonstrate that starting from the similarity relation of colours, for instance, in conformity with the construction rules proposed by Carnap, one will arrive at the impossibility of making classes on the abstract principle. Let us suppose that we have in front of us many objects which have one or more colors from an assembly of 3 colors: red (*r*), yellow (*y*) and blue (*b*). Without knowing anything else besides the kinship relation or of "color similarity" that is the possession of a common color (according to Carnap's Gestaltist principle), let us try to to establish the classes of colors, i.e. what it is common (abstract in Goodman's opinion) from the similarity circles. Let us say that in the case of our objects the red color (*r*) appears only in the objects which also have the yellow color (*y*), in the following combinations: 1. y*r*; 2. *y*; 3. y*b*; 4. *b*; 5. yb*r*. Can color classes "red", "yellow" be encompassed separately, wonders Goodman? The answer is negative since the class for the "red" is included in the class for the "yellow" color, it is part of it (the class for the "red" color" {1 5} is just a part of the class for the "yellow" color {1 2 3 5}). Consequently, if no color is common to all the objects in an assembly, we won't make up a class on the basis of the abstraction principle. In the end, Goodman thinks, consistently with his nominalist position, the abstraction problem itself is disputable, since it is difficult to establish under which conditions a quality is or not common to the assembly of people[27].

Without any doubt, Carnap's conception is not easy to be deciphered, and maybe it also has some sterile and excessive elements as Michael Dummett appreciates[28], but Goodman's analysis itself seems to induce some confusions. Can the "abstract" no-

[25] True, there have been other attempts to exploit the work of Carnap, some particularly interesting as that of C. Ulises Moulines, *La estructura del mundo sensibile*, Ariel, Barcelona, 1973. Since it was published in Spanish, it remains virtually unknown.

[26] Nelson Goodman, *The Structure of Appearence*, Harvard University Press, 1951.

[27] *Ibidem*, p. 125.

[28] Michael Dummett, *Originile filosofiei analitice* (Romanian translation by Ioan Biriş), Editura Dacia, Cluj-Napoca, 2004, p. 236 (interview with Fabrice Pataut).

tions equate what is "common"? No, in a strict sense the "abstract" predicate only aims at overlooking or not taking into consideration some features considered irrevelant from a certain perspective. Then one cannot attribute "the abstraction problem" to Carnap[29]. Then, Goodman's critics is tributary to his "projective" type conception. While Carnap is preoccupied with offering a system of empirical "reconstruction" of the world of knowledge, so a "genesis" (which can be causal starting from the sensitive stimuli, which interprets Quine in the end), Goodman presents as an alternative to Carnap's project a "projective" type which, of course, can be served by the principle of abstraction. In this way, Goodman is talking of course about the induction problem, not about Carnap's construction type. By refering to this problem, Quine emphasizes that at a certain point the attempts to confirm an induction were aggravated by Hempel's paradox: that is the observation of a black raven should constitute a confirmation for the hypothesis that "all ravens are black", on the other hand the observation of an object which is not a raven and consequently it is not black either should be a confirmation for the hypothesis that "an object which is not a raven is not black". The two hypothesis are logically equivalent, from which the paradoxal conclusion that any object which is white, for instance should be a confirmation for the hypotheis that "all ravens are black".

This paradox signaled by Hempel was exaggerated by Goodman through the difficulty of the example of green emeralds[30]. If one supposes for example that one has noticed until a certain moment that all emeralds are of the green color, it is expected that the next emerald which we will see will be of the green color. Let us imagine that after a certain moment call it *t*, the noticed emeralds are of the blue color. We can built then a new adjective "greeblue" by the contraction of "green" and "blue" that is the "greeblue" emeralds can contain the green ones which can be noticed until moment *t*, and the blue ones (which exist) which were examined after moment *t*. In this way one can expect that the next emerald which will be noticed will be green as well or again greeblue. We find ourselves again in face of a paradox: the next emerald can be green and simulaneously blue since all the emeralds examined have been greeblue. Of course, on the basis of common experience we are not tempted at all to think that the next emerald will be blue. It is alarmingly difficult, though, to tell why this inference is not legitimate while the conclusion to the green one is. What we can see clearly is that to say that we expect that the future cases be like the past ones means to say nothing[31].

How can we explain the fact that we expect that the next emerald be green or greeblue? Goodman's answer is that while the "green" adjective is *projectable* -- it can be prolonged in the future – "the greeblue" adjective is not. Otherwise said, the conclusion projects in the future especially those features which are projectable, better rooted, by neglecting the others. Which is the base on which this operation takes place? Quine wonders if it is not on the basis of resemblance, letting us understand that Carnap is right and not Goodman, who stresses himself to break the idea of resemblance from the basis of Carnap's construction? Because Quine emphasizes the question of what features are projected can be then very well formulated in this way: what thing counts as resemblance?[32] The fact that emeralds have the green color in common counts as resemblance, while the color greeblu isn't in the same situation. It is a fact , Quine thinks, about our natural abilities to recognise certain characteristics, since our eye for projectability is our eye for resemblance[33]. The induction procedure, used on a large

[29] Unfortunately the different meanings assigned to abstraction often let us confused. Even in some philosophical dictionaries there are such mistakes. In our opinion it is right to talk about the abstraction process, case in which we have in mind the selection, retention of a characteristic, of a note and the neglect of the others, and abstraction is the result of this process. That is why abstraction should be made distinct from what is common, that is from "general", which is obtained following the generalization operation, of extrapolation of a characteristic of a class of objects. One can also express in this way the unity of that class. Also the abstraction must be differentiated from the "universal", which expresses rather what is repeatable, as a result of the repeat operation. If we understand in a proper way Carnap, to him abstraction has the meaning of selecting something from a multitude, of a color or a color palette, for instance, as Cassirer also thought. His accepting of transitivity, in certain situations is allowed not in the virtue of abstraction as a common element, in general but in the virtue of universality, that is of the repeatability of elements or of making relations at the different levels of the logic construction of the world.

[30] W. V. O. Quine, *Ontological Relativity and Other Essays*, Columbia University Press, NewYork, 1969, p. 114.

[31] W. V. O. Quine, J. S. Ullian, *Țesătura opiniilor* (Romanian translation by Mircea Dumitru), Editura Paralela 45, Pitești, 2007, p. 101–102.

[32] *Ibidem*, p. 103.

[33] *Ibidem*.

scale in scientific and common knowledge is not something "extremely intelectual", because other animals do it too. Our eye for resemblance ad projectability, Quine continues, is, in its roughest part a part of our animal legacy[34]. We have a born tendency, Quine concludes, to notice features such as green, and on its basis to make successful predictions. The answer to the above questions must be formulated in terms of the natural selection. That is why the teleological explanations towards which Goodman's attempt moves must be subordinated to the causal ones.

We can say in a way that in his works Quine develops naturalized epistemology starting from one of the basis programs in *Aufbau*, that is from Carnap's attempt to overcome the crisis regarding a foundation in physics by appealing to the phenomenal world of psycho-physiology and biology. Maybe this appeal is not accidental, since, as Graham Priest states[35], in the German and English language tradition there are many philosophers who have a solid education in the natural sciences. This is the case for Carnap or Quine. As we've already mentioned, in the period in which he wrote *Aufbau*, Carnap hoped that the solid foundation of scientific knowledge could be looked in the biologic strata of reality, but not in conformity with Mach's model, instead in conformity to the primary relation of *Gestalt* type resemblance. Carnap's choice is of a naturalist order, by making a criterion from the information offered by the empirical sciences for the concepts which will be used in the reconstruction process of those sciences, though the approach seems to take place in a circular way. The construction system proposed by Carnap can be considered one of a physicalist type, but let's not forget that the psychic objects can be reversed to the category of physical objects, and the physical ones in their place can also be reversed to the category of psychic objects, the problem being one of choice[36] and of coding in a language. In conformity with the gestaltist principle, for Carnap there is no isolated sensitive data, since everything that experience offers us is constituted in relational structures (*Beziehungsgefüge*), so the objects must not be thought as a sort of "points", but as "nodes" of complex and structured relations.

In the case when the primary sensitive experiences don't offer us objects, don't present us "things" (because they are *undinglich*), what then is the reality layer from which to start? Carnap tells us that it is the resemblance relation which we should put at the basis of the constructionist system of scientific knowledge. And Quine, when he analyses the natural genres will appreciate that we don't need to search beneath our similarity sense, because there is nothing more fundamental for thought and language ("there is nothing more basic to thought and language than our sense of similarity"[37]). This thing can be easily noticed when we learn how to use a new word. Such a process –Quine emphasizes – depends on a double similarity: on one hand, the similarity on a temporal line between the present circumstances and the past ones in which that word was used, and on the other hand, the phonetic resemblance between the present uttering and the past utterings of this word. With the indication that the "similarity" and the "genre" notions are variants or adaptations of the same notions, because the similarity can be defined through the term genre, we are saying that two things resemble one another if they are of the same genre. We find the same situation in Carnap, in *Aufbau*, when he presents the genre identity as a membership of two states of things (*Zustände*) to the same thing, that is to their genre. In this way, identity doesn't target simply the object to which language refers, but the genre, species which this object represents[38].

It is educational in this sense the example given by Carnap in order to make an "object", by emphasizing that the different states of things (as temporal parts, that is "parts" on the temporal line of similarity) precede the operation of building an object as an entity persistent in time. In this case the genre identity reveals to us as a report between the temporal parts of the object and these ones can be the referents of some demonstrative pronouns. From Carnap's example with a tram which is travelling in different temporal moments, one can remember two expressions: a) this rail car is the same which travels until now on line 10 (*die ist derselbe Wagen, der bisher auf der Linie 10 fuhr*); b) this tram has in A the same rail cars as in B (*dies Straβenbahn in A hat dieselben Wagen wie die in B*). In these two expressions the term "same" (*derselbe, dieselben*) expresses the identity judgement and the demonstrative "this" (*die, dies*) expresses the reference of two different states of things. If the particular object expressed by the demonstrative "this" corresponds to the temporal part, in exchange the persistent object corresponds to the genre. In other words, the genre

[34] *Ibidem*, p. 104.

[35] Graham Priest, "Quine: Naturalism Unravelled," communication presented at the *Simpozion omagial – 100 de ani de la naşterea lui Quine*, The Faculty of Philosophy, University of Bucharest, 8–9 november 2008.

[36] Rudolf Carnap, *Aufbau*, § 57 şi 64.

[37] W. V. O. Quine, *Ontological Relativity...*, p. 115.

[38] Rudolf Carnap, *Aufbau*, § 159.

identity can only be applied to physical objects, persistent in time, which means that in the auto-psychic world (*die eigenpsychische Welt*), where the objects can't be made yet, we can't talk about the genre identity and consequently, about classes. It is not accidental that this genre identity is introduced by Carnap in *Aufbau* only when paragraph 128 talks about the making of visible things (*Sehdinge*)[39]. The visible things are "classes" of entities, of "points" of the visual field.

This philosophical direction imposed by Carnap regarding the structural description of knowledge experience on the basis of the similarity relation and of quality spaces seems to be confirmed by more and more in the late years also by the researches from the cognitive perspective. Recent studies from this area of investigation bring to our attention the fact that different concepts of knowledge are structured and we can speak about *conceptual spaces* (concepts organized on the colors' space and the sounds' space etc.). Any conceptual space contains a certain number of qualitative dimensions, and these dimensions can have a sensory and non-sensory character, since they form the context in which we can design the objects of knowledge and the relations between them.

References

Biriş I. (2007). *Totalitate, sistem, holon*. Ediţia a doua, Editura Universităţii de Vest, Timişoara

Boniolo G, Dalla Chiara M. L, et all., (2002). *Filosofia della scienza*. Raffaello Cortina Editore, Milano

Carnap R. (1998) *Der logische Aufbau der Welt*. Meiner Verlag, Hamburg

Carnap R. (2001). *Vechea şi noua logică*. (Editions and translation in Romanian by Alexandru Boboc), Editura Paideia, Bucuresti

Dummett M. (2004). *Originile filosofiei analitice*. (Romanian translation by Ioan Biriş), Editura Dacia, Cluj–Napoca

Flonta M. (2008). *Gânditorul singuratic. Critica şi practica filosofiei la Ludwig Wittgenstein*. Editura Humanitas, Bucuresti

Goodman N. (1951). *The Structure of Appearence*. Harvard University Press

Husserl E. (2007). *Cercetări logice*. vol. I (Romanian translation by Bogdan Olaru), Editura Humanitas, Bucureşti

Kistler M. (2001). Le concept de génidentité chez Carnap et Russell. In: Laugier S (Ed.) *Carnap et la construction logique du monde*, Vrin, Paris, 2001

Laugier S. (Ed.) (2001). *Carnap et la construction logique du monde*. Vrin, Paris

Moulines C. U. (2001) La (re)construction formelle de l'expérience. Carnap et Nicod. In: Laugier S (Ed.) *Carnap et la construction logique du monde*, Vrin, Paris, 2001

Moulines C. U. (1973). *La estructura del mundo sensibile*. Ariel, Barcelona

Priest G. (2008). Quine: Naturalism Unravelled. Paper presented at the *Simpozion omagial – 100 de ani de la nasterea lui Quine*, The Faculty of Philosophy, University of Bucharest, 8–9 november 2008

Quine W. V. O. (1995). *From Stimulus to Science*. Harvard University Press, Cambridge/Massachusetts,London/England

Quine W. V. O. (1969). *Ontological Relativity and Other Essays*. Columbia University Press, NewYork

Quine W. V. O. Ullian J. S, (2007). *Ţesătura opiniilor*. (Romanian translation by Mircea Dumitru), Editura Paralela 45, Piteşti

Russell B, (1914). *Our Knowledge of the External World*. Open Court, London/Chicago

Wagner P, (2001). Le contexte logique de l'Aufbau. Russell et Carnap. In: Laugier S (Ed.), *Carnap et la construction logique du monde*, Vrin, Paris, 2001

[39] Max Kistler, Le concept de génidentité chez Carnap et Russell, in Sandra Laugier (ed.), *Carnap et la construction logique du monde*, p. 166.

BALKAN JOURNAL OF PHILOSOPHY. Vol. 2, Issue 2, 2010

AUSTIN ON TRUTH

Aneta Karageorgieva

Department "Philosophy", Faculty of Philosophy at Sofia University "St. Kl. Ohridski", Bulgaria
anetakarageor@yahoo.com

Abstract:
J. L. Austin's article 'Truth' is a remarkable example of his method of ordinary language analysis by which he investigates the notoriously recalcitrant philosophical concept of truth. This paper attempts to specify the character of Austin's truth conception, defending the view that despite his opinion of it being a semantic one, it is actually of a mixed nature. Correspondence is what determines the truth or falsity of a statement, but semantics delineates the segment of the world to which our statements correspond. An explication of the notions of demonstrative and descriptive conventions is given in order to add flash to the outline given by Austin.

Key words: truth, semantics, correspondence, ordinary language analysis.

In his 1950 piece, 'Truth', John Austin points out that instead of trying to say whether truth is a substance, or a quality, or a relation, "philosophers should take something more nearly their own size to strain in. What needs discussing rather is the use, or certain uses, of the word 'true'." (Austin, 1970:117). This is clearly a tribute to Austin's linguistic turn, but nevertheless it contributes substantially to the debate about the belonging of truth to the domain of semantics, as D. Davidson treats it in his 1967 "Truth and Meaning" (Davidson 2001:17-43), or whether truth should be considered an epistemological topic, as the tradition has it. In what follows I will track Austin's conception of truth-bearer (part I), then of truth proper (part II), discuss the latter's semantic nature (part III) and try to show that nevertheless it is a basically correspondence conception that he adopts (part IV).

I. Austin begins with the question: what is it that we say is true or false? In the apparently multifarious answers he tries to find some consolidation around what he calls a primary use. His suggestion is the following:

(1) It is true (to say) that the cat is on the mat.
That statement (of his &c.) is true.
The statement that the cat is on the mat is true

are the primary uses of 'is true' (Austin, 1970:118). (From now on I will be pointing only to the relevant page of the article 'Truth'.)

Austin, first, rejects the rival candidates for thruth-bearers by pointing out that ascribing truth to beliefs is not common in everyday language, but rather in the specialized language of philosophers and theologists. Moreover, beliefs are about something, or in today's jargon they have content and that content is what is captured by suggestion (1). And third, a belief is of the nature of a picture which cannot be described as true, or at least it is true *to* life, but not *of* it. The second rival candidate to be rejected by Austin on different grounds is descriptions, propositions, and true accounts. They are varieties of true statements that differ from the latter by degree of generalization; and generalizations "cannot be a direct report on current observation" (Ibid.). Of course, this is mainly the everyday usage of 'proposition' and Austin is well aware of it. Therefore afterwards he explores the philosophical usage and finds propositions incapable of bearing the predicate 'is true' because 'proposition' captures the meaning of the sentence and it is impossible to say that the meaning of the sentence is true. The last candidate for a truth-bearer – words and sentences - must be cast aside firstly as a traditional subject to linguistic inquiry as well as a concern of printers and critics. In this area words and sentences can be mis-spelled, elliptic or ungrammatical, but not true or false. However, there is a sense in which truth applies to words and sentences and here Austin makes his great discovery, with which his work in the philosophy of language is connected. It is closely related to the notion of truth, too. He draws philosophers' attention to the fact that in expressions like 'His closing words were very true' demonstratives consistently accompany the use of 'is true' so that such expressions are true or false '*as used by a certain person on a certain occasion'* (p. 119, italics by Austin). This means that such sentences refer to statements and, moreover, their utterance is a historic event, liable to localizations and easy to be dated. As Austin concludes, 'A sentence is made *up of* words, a statement is made *in* words' (p. 120, italics by Austin).

This discovery shows how the same sentences can be used in making different statements: I say 'It is mine", you say 'It is mine', so the thing in question belongs to a different person in the two cases. In addition, same words can be used on two diverse occasions or by two persons to make the same statement. So it is the statement that is true, not the sentence, although Austin accepts also the word 'assertion', noting that it is a bit wider. This is a very serious move towards a semantic conception of truth, because it is reasonable to expect that by specifying the statement's truth conditions we are to find its meaning, too, and in a footnote on page 120 Austin says so explicitly. Moreover, by emphasizing the demonstrative form of some sentences which are used to make different statements by different users Austin gives us additional ground for such an interpretation. It is only by their meaning that the different statements can be recognized as different, which simultaneously brings to the fore their truth conditions. 'It is mine' is true in my mouth only if it is mine, and is true in your mouth only if it is yours. So it is implied that the truth conditions *are* the meaning, so that we can call Austin's conception at this moment 'semantic'.

II. The next move is to specify when a statement is true. Austin resists the temptation to answer 'When it corresponds to the facts' because it can be misleading. In order to clarify this point, he sketches out a notion of linguistic communication. It needs symbols, a communicator, a communicatee, and something that the words are used to communicate about, namely the 'world' (p. 121). These components are not enough, however, because the relationships between them must be governed by some rules, e.g. the same words must have the same meaning on different occasions (not *always*, however) so that understanding can occur regularly. This is the famous principle of meaning normativity or the Humpty-Dumpty principle. Austin identifies two conventions that secure meaning normativity, namely:

Descriptive conventions that correlate the words, or sentences, with the types of situation, thing, event, etc. to be found in the world.

Demonstrative conventions that correlate the words with the historic situations to be found in the world (pp. 121–122).

The descriptive conventions are those which pick up the same referent for each use or tag the same object, or whatever your theory of meaning requires. The demonstrative conventions are those which connect each use with a particular speaker, so that it is clear, when somebody says 'I' or 'that', which one of the two or more present 'I'-s or 'that'-s is meant. In other word, these conventions contain the rules for the use of demonstrative pronouns. With these two kinds of conventions in hand we are ready to learn what it means for a sentence to be true.

'A statement is said to be true when the historic state of affairs to which it is correlated by the demonstrative conventions (the one to which it 'refers') is of a type with which the sentence used in making it is correlated by the descriptive conventions.' (p. 122)

In order to paraphrase and clarify this definition we can say the following: a statement S made by a speaker X in time t by 'I am angry' is true when the correlate of 'I' in the mouth of X in t has the property which is described by 'am angry'. The first requirement is governed by a demonstrative convention concerning personal pronouns, and the second requirement is regulated by a descriptive convention about how the property 'angry' relates to people.

After this definition the author goes into a digression which actually can be regarded as unnecessary for his explication of truth. Namely, Austin focuses on the concept of 'fact' which he finds troublesome. The reason is that 'fact' is used in sentences with a 'that'-clause, which is characteristic of statements as well. According to his linguistic method Austin points to the way of using the word 'fact', but what exactly is meant by it remains unclear. Maybe that is why Austin prefers to talk about states of affairs which make a certain statement true without elucidating further their nature. He only says that they are distinct from the true statements about them, although a somewhat lengthy discussion is devoted to possible mistakes in using 'fact' in everyday situations.

Another usage that Austin finds troublesome is that of 'correspondence'. Essential about it is that the correlation between the words and the type of situation is thoroughly conventional or, we may say, arbitrary. 'We are absolutely free to appoint any symbol to describe any type of situation, so far as merely being true goes.' (p. 124) Austin opposes strongly the idea that the words used in making a true statement mirror any trait of the worldly situations, or that there is any kind of isomorphism or structural resemblance between language and reality. To assume this, warns Austin, is to populate and overpopulate the world with linguistic entities. Here he develops further his conception of meaning, too. He contests the idea that iconic entities like photographs, replicas, pictures or copies can be said to have meaning in a semantic sense, because they are mere reproductions of the world, but not signs of it. Therefore they are true *to* the original, but not true *of* it. Actually, here Austin means that linguistic signs are intentional. This understanding is also to be seen in his next argument against 'natural meaning', if we put it in the words of his colleague H. P. Grice (see his 1957 'Meaning'). A 'natural sign' can only be reliable or unreliable, but never right or wrong, as linguistic signs can be. This argument is based on our common everyday usage of the words 'reliable', 'unreliable', 'right' and 'wrong'. Austin relies heavily on our intuitions about the relevant usage. They are indeed an indicator of appropriateness and the argument is quite plausible and cogent, although the backup is not to be accepted fully, because intuitions vary.

III. By this moment the general distinction that Austin makes is between a true account and a faithful picture. Intermediate cases exist, however, and their study can help to make the contrast salient. Austin's example is with maps, which he finds to be 'highly conventionalized pictures' (p. 126). Unfortunately he does not go further into this intriguing matter. His preferences are to the critique of redundancy conceptions of truth, to which he devotes a separate section (4). Why is the statement that 'to say that an assertion is true is not to make any further assertion at all' wrong? Because a true statement refers to the world or a part of it that does not contain this very statement (or rather, the sentence in which the statement is made). Further, a true statement about which we say it is true refers to a part of the world that includes the very sentence in which the statement is made. It is to be seen easily that the two statements refer to different parts of the world, or, as Austin puts it, the one refers to something that the other cannot refer to (p. 127). This is a brilliant move. It both solves the looming semantic paradox and establishes the falsity of the redundancy account. It shows that the two sentences have a different meaning and therefore are not logically equivalent. To make a final bow to his ordinary language analysis, Austin adds: 'Mistakes in philosophy notoriously arise through thinking that what holds of 'ordinary' words like 'red' or 'growls' must also hold of extraordinary words like 'real' and 'exists'. But that 'true' is just such another extraordinary word is obvious.' (p. 127-128). It is not quite sure that the appeal to obviousness is a good argument in philosophy (cf. Descartes' *Discourse on the Method*), but it is out of doubt that Austin has brought us to a position or point of view from which the word 'true' seems extraordinary indeed. Thus we are able to share with him 'the obvious'. Now this is a good argument, although not one to be formalized.

The next point to be made about the semantic nature of Austin's account is his analysis of the statement that the sentence S (an arbitrary sentence we may choose from the examples given in the text, for instance 'The cat is on the mat') is true. This statement seems to describe a fact, but Austin casts a reasonable doubt on it. It is hardly a fact what the sentence describes because the relation between the statement itself and the world which the extended statement about S being true asserts to obtain 'is a *purely conventional* relation (one which 'thinking makes so')' (p. 128, italics by Austin). This relation can be altered according to the will of the assertor, and facts must be unalterable at will. This means that language is what picks up the part of the world to which a statement corresponds and this is a semantic view *par excellence*.

Another usage that gives support to this analysis is connected with negation which must not be confused with falsity. It is not the same to assert that 'He is not at home' and 'It is false that he is at home'. The second implies that somebody says that he is at home, so there is at least a 'communicative' difference. There is a logical difference, too, according to Austin, but he does not explore it here. He only mentions that assertion of falsity is logically different from negation and goes back to his linguistic argument that 'no language can exist which does not contain conventions for both (assertion and negation, A. K.) and that both refer to the world equally directly...' (p. 129). This declaration is to be proved empirically, however, because actual languages tend to surprise us when compared to normative considerations about language *in abstracto*.

Austin proceeds with another linguistic argument in favor of 'is true' not being redundant, which we may say is an argument by analogy. It concerns the class of extraordinary words to which 'true' and 'false' belong; the distinguishing characteristic of this class is that it relates words with the world. As Tarski (Tarski, 1944) has shown, this is what makes a relation semantic. Therefore Austin's approach is also a semantic one. Moreover, it outlines a particular semantics. As he puts it, 'There are various degrees and dimensions of success in making statements: the statements fit the facts always more or less loosely, in different ways on different occasions for different intents and purposes" (p. 130). Austin appreciates this by saying in a footnote that coherence and pragmatic theories of truth make sense in this point. He does not say, however, what is evident in his account: namely, that his semantic is also pragmatic (relying on the notion of success) and intentionalist (introducing intent as a factor in establishing the meaning of a word).

The next analogy is with the word 'freedom' (Austin seems to prefer the somewhat loose usage 'word' to stricter ones like 'term' or 'concept' in order to remain within the everyday language). He makes the ingenious point that when we discuss whether an action is free or not, we are stuck with the word 'free' and cannot move far away from mere repetition of our stance. When we turn, however, to related words like 'unwillingly', 'accidentally' or 'inadvertently', we can easily decide on the nature of the action without it being necessary to judge the action additionally by 'free" or 'not free'. 'Like freedom, truth is a bare minimum or an illusory ideal…' (Ibid.).

IV. Austin thinks it jejune to assume that a statement's function is only to be true. Moreover, not each statement aims at being true. He calls this propensity of statement analyzers 'the descriptive fallacy' (p. 131). Many utterances are not descriptive, nor capable of being true or false. Austin enumerates cases in which a statement is not a genuine one: when it is a calculus, a performatory utterance, a value-judgement, a definition, part of a work of fiction, etc. 'It is not the business of such utterances to 'correspond to the facts' (and even genuine statements have other businesses besides that of so corresponding)' (Ibid.). A footnote on page 129 says that to be false means not to correspond to a non-fact, but to miscorrespond to a fact.

Here we can easily see that correspondence plays the main role in characterizing true statements (maybe we should add – true statements of the nature negatively specified above). It implies that Austin is still engaged with the correspondence conception, modifying it in such a way that it becomes clear what semantic relations are included in it – the relations of 'meaning' and 'describing', the first one being purely semantic, the second not so pure (as it is capable of being judged for truth or falsity). This difference can be seen in the example given by Austin: that we make a difference in using 'You said you promised' and 'You stated you promised', the first one pertaining to meaning, the second one – to describing *and* performing a statement. Austin's own analysis puts the meaning of the first utterance in present tense – 'You promise' and that of the second – in past tense: 'You promised'. The second one is capable of being characterized as true or false, while the first one is not susceptible to these characterizations. The decisive example, however, is about geometrical axioms: when people thought that they describe the world, they were called true; but when their correct nature was recognized (see for example, Plato's image of the divided line in his *Republic* 509d–513e), there are no more called true or false (they are part of a coherent system which establishes its own ontology). Here Austin admits that there are descriptive words and that the propositions using them for the task of describing the world must correspond to the world. Here again he endorses the correspondent conception of truth.

The next case which has to accommodate in full our intuitions about truth is that of non-observable states of affairs – the past and the future. According to Austin, this predicament is to be solved by expressions of possibility – 'The cat may be on the mat'. This purported statement is quite compatible with its negation that the cat may not be on the mat. The compatibility shows that it is not a statement at all, because true and false statements are incompatible with each other (see p. 132).

The last section elucidates the reasons why Austin does not accept Strawson's conception of truth. The latter declines the semantic conception (here Austin goes back to his self description of a follower of the semantic conception) as it does not use the phrase 'is true' to talk about sentences. But Strawson denies that 'is true' is a proper way to talk about anything. 'Is true' to Strawson is logically superfluous: by saying that S is true we are not saying anything different from S. Here we can trace the beginnings of the deflationary theory – a modern version of the redundancy theory popular through the works of Frege. On the other hand, Strawson does not fully accept the logical superfluity of 'is true' in that in saying that a statement is true we are *doing* something more than to state S – it is to confirm the assertion that S. Austin is unsure whether to accept this second part of Strawson's position on the ground that to say 'is true' may mean that we describe the world - not only the assertion (p. 133). This confirms the correspondence nature of Austin's conception, but it is in the nature of the semantic relation too: to describe the relation between word and world means to describe *a thing* in the world which itself can be further described so that a threat of infinite regress arises. It can be said that the correspondence relation is Austin's attempt to save his semantic understanding from this threat.

Conclusion. It must be pointed out that Austin does not need the somewhat unclear category 'fact' for his explication of truth. The notions of descriptive convention and demonstrative conventions seem to suffice, although he does not elaborate on their content. As far as one can see, a demonstrative convention regulates the use of indexicals, e.g. by 'I' in English the speaker describes himself. A descriptive convention controls the use of properties as applying to certain property-bearers. However, there are sentences that do not contain indexicals, on the one hand, and on the other hand, the everyday language correlates of logical connectors are neither indexical nor descriptive. Austin's conception, therefore, must be limited to simple empirical sentences. His achievement is to be seen in the looming view of semantic relation as a type of correlation which strongly resembles correspondence – what will be called a few years later 'a descriptive use' by K. Donnellan.

References

Austin, J. L. (1950). *Truth.* In: *Philosophical Papers*, Oxford: Clarendon Press, 1970

Davidson, D. (1967). *Truth and* Meaning. In: *Inquiries into Truth and Interpretation*, 2nd ed. Oxford: Oxford University Press, 2001[2].

Donnellan, K. (1966). *Reference and Definite Descriptions*. In: Martinich A. P., *The Philosophy of Language*, Oxford University Press, 1996[3]

Grice, H. P. (1957). *Meaning*. In: Martinich A. P., *The Philosophy of Language*, Oxford University Press, 1996[3]

Tarski, A. (1944). *The Semantic Conception of Truth and the Foundations of Semantics*. Philosophy and Phenomenological Research 4. At: http://www.ditext.com/tarski/tarski.html, retrieved on May 22, 2010.

BALKAN JOURNAL OF PHILOSOPHY. Vol. 2, Issue 2, 2010

THE PROFILE OF ROMA IN MAJORITY MEDIA*
– A Scheme for a Theoretical Research Framework –

Dragan Todorović, Dragoljub B. Đorđević

University of Niš, Serbia
todor.d@EUnet.yu ; brkab@junis.ni.ac.rs

Abstract:

Having recently acquired national minority status, Roma in Serbia were and still are the most frequent target of social distance practiced by the majority people as well as other minorities. Apart from refusing to socialize with them, the majority has created a series of negative stereotypes and prejudices, with few examples of positive attitude. The Serbian media have rarely stood up against such tirades, usually fuelling or, more often than not, instigating them.

Having ascertained the conditions through the analysis of several empirical reports, the authors in this work suggest a list of concrete acts in the media, whose realization would contribute to the genuine appreciation and overcoming of the unenviable position of the Romani national minority in our country, as well as the countries in the region.

Key words: Roma, Serbia, Social Position of Roma, Majority Media, Profile of Roma, Stereotypes and Prejudice, Social Distance, Positive Discrimination.

The Roma in Serbia

According to the census in 2002, there are 108,193 Roma in Serbia. However, demographers estimate that there is a considerably bigger number of the Roma in Serbia, about 450,000. The difference between the established and the estimated number of the Roma is the consequence of "ethnic mimicry", i.e. of the decision of the Roma themselves, who, during the census, declared to be members of some other nationality instead of Romani (Jovanović 2003). The majority of them are in the South-east Serbia, and in some municipalities (Surdulica, Bujanovac, Bojnik, Vladičin Han) the participation of the Roma in total population goes up to 1/3 of the population (*Nacionalne manjine u SR Jugoslaviji /National Minorities in FR Yugoslavia/* 2002:78). The census data reveal that only 21.9% of the Roma finish primary school, 7,8% finishes secondary and only 0.3 % have college or academic education. Almost 80% of children who attend schools for children with special needs are of Romani nationality. Nowadays, there are dozens of Romani social and political associations. Since 1989, there is the Committee for Romani life and culture studies under the Serbian Academy of Science and Art, and "The New Testament" and "The Five Books" have been translated to the Romani language. There are broadcasts in the Romani language on the National TV and daily radio broadcasts. In addition, the first independent Romani media have been founded.

The social position of the Roma in our country has been determined for decades by the following characteristic indicators (Митровић 1996:815):

1) *residential (spatial) segregation* (life in *ghetto* settlements/districts);

2) massive *unemployment* or working on *underpaid, unappreciated and stigmatizing jobs* (street cleaners, undertakers and helpers in factories);

3) unavailability of *educational possibilities* (high percentage of the illiterate people and of people with unfinished primary education and a negligible percentage of individuals with college and academic education);

4) low levels of *social and health care*;

5) the lack of *political participation* as a resource for mobilization and protection of ethnic and cultural identity.

The Federal Republic of Yugoslavia granted the Roma the status of a national minority in 2002 by adopting the *Law for the protection of rights and liberties of national minorities.* Thus were fulfilled the necessary institutional preconditions for intensifying the process of overall social and economic emancipation of the Roma. Soon afterwards, the first *National council of the Roma from Serbia and Montenegro* was formed, and the Serbian Government

* The text was prepared as a part of the project The Culture of Peace, Identity and Inter-ethnic Relations in Serbia and the Balkans in the Process of European Integrations (149014D) which is conducted by the Center for sociological research at the Faculty of Philosophy in Nis and financed by the Serbian Ministry of Science.

founded the *Council for national minorities of the Republic of Serbia*. The Romani representatives and the representatives of other national minorities in Serbia participate equally in its work.

The World Bank and The Institute for Open Society organized, in Budapest, in June 2003, the conference "The Roma in Broadened Europe – The Challenges of the Future". On the Conference the high representatives of the countries participating – Serbia and Montenegro, Hungary, Romania, FYR Macedonia, The Czech Republic, Slovakia, Croatia and Bulgaria – agreed to start the programme called *The Decade of the Roma Inclusion 2005-2015* and set the fields of education, health, housing and employment as their priorities. The goals of the programme were the reduction of unacceptable differences between the Roma and the rest of the society as well as putting an end to discrimination through development and implementation of national action plans in the areas of priority along with the active participation of national Romani communities in the realization of the Decade. The formal announcement of the beginning of the programme implementation on the international level was held on February 2, 2005 in Sofia when the Prime Ministers of the countries participating signed the *Decade Declaration*. After that, The Government of Serbia, The Ministry for Human and Minority Rights and The National Council of the Romani Minority marked the beginning of the "Decade of the Romani Inclusion" in Serbia on April 7, 2005 in Belgrade.

Before the authorized state institutions lies, already well started, a decade of patient work on creating realistic political, social, economic and cultural preconditions for overall social integration of the Roma. By continuous introduction of the particularities of the Romani ethnos to the surrounding majority, the reasons of social animosity towards them will be gradually reduced and their status as equal citizens affirmed. The political measures taken will raise the levels of educational achievements, living conditions and economic power. Also, the Romani voice will be heard when the guarantees for their full presence on all levels of decision-making and their active role in the public sphere are provided. The full support to creating the necessary preconditions for affirmation of Romani identity must, without fail, be given by the Serbian media. However, this should be done in a more complex and inclusive way than has been done in the past.

The image of minorities in the media

Based on our own experience, we can say that we get information about the problems of national minorities primarily from the media. "This is why it is important to discover what kind of messages about the minorities is being created by the media, how the media report their activities, which minorities are favored and which are neglected" (Tatalović 2003). Our perception of the minority cultures as values depends in great part on these reports. The media are not merely "the reflection of a society", but also "the window into a society".

There have been regular and detailed reports on the role of the mass media from the former SFRY in creating the public opinion about the minority issue (Радојковић 2003; Živković 2004; Filipović 2004; Vilović 2003; Kanižaj 2006; Zeković 2006). In their analyses, the authors agree that the media very often ignore or speak negatively about minorities. The members of minorities have been depicted as different and dangerous; they are described with prejudices, generalizations and stereotypes; they are given the main roles in articles and radio and TV broadcasts about incidents and conflicts; they are connected to the main causes of quarrels, misunderstandings and unsolved problems or, in the political sense, they are perceived as generators of ethnic conflicts; the underestimation and intolerance towards them is propagated; they are the victims of the media manipulation and abuse. In the sea of sensationalistic statements that appall the audience, hardly ever does a "pearl" of tolerance, inclusion in the community life, understanding and co-existence appear.

Along with the Albanians and the Muslims (the Bosniacs), the Roma are the national minority which has always been received by the Serbian majority with a high level of xenophobia, great ethnic distance and a number of prejudices and ethnic stereotypes which represent a significant psychological barrier to integrative processes. The data from empirical research of the Center for sociological research of the Faculty of Philosophy in Nis (Тодоровић 2007а) prior to the latest one are very disturbing. According to this research, a standard high percentage of people who would reject marital relations with the Roma is followed by the fact that almost one third of the participants would avoid being friends and neighbors with the Roma.[1]Apart from using impurity as a "synonym" for the Roma,

[1] More about the social distance from the Roma in Serbia in Тодоровић 2007б.

more than a half of the Serbian people describe the Roma as dishonest and rude. The range of positive stereotypes is centered on their being sensitive, peaceable, hospitable and generous as well as being friendly to other peoples (Todorović and Milošević 2004). The situation is no better in their environment (Marko 2004; Babić 2006).

The editors and the journalists in the majority media amply deepen the afore mentioned prejudices, even contribute to creating the new ones, and forget that the Roma are, just like everybody else, only people with the same short-comings, virtues and hardships. Here and there, albeit rarely, they are sentimentally represented as merry people who sing and dance all the time. More often, the national attribute "Romani", without mentioning names and surnames or followed by initials only, is put in a negative context and connected to questionable or bad intentions; it is often used to evoke the need for precaution, to evoke hostility and even violence. The background should be completely different. Instead of focusing on negative news about the Roma, the focus should be put on finding affirmative stories and giving the opportunity to the Roma for their voice to be heard.

Media reports on the Roma: examples from Serbia (and Bulgaria)

Lela Milošević and Gordana Stojić-Atanasov (2003) have conducted a qualitative and quantitative analysis of the articles in two Serbian daily newspapers with high circulation – *Политика* and *Блиц* – in 2002. The manner of reporting about the Roma in the press creates about them the image of a social group in need of help (they point to the necessity of improving their social and economic positions and objectively state the incidents which included the Roma as victims), but some texts are characterized by qualitatively negative generalizations. The published writings are about politics and culture (the elections, the activities of the international organizations and the local government, conflicts), and the most common themes were: measures for improving the positions and rights of the Roma, Romani education, skinheads' attacks and verbal and physical violence against the Roma. The positive articles describe the lifestyle and customs of the Roma. Two thirds of the texts are concerned with the collective identity of the Roma, and only one third considers Romani men and women as individuals. Other people talk about their activities: 10% of the participants in social events are quoted, 5% are paraphrased, and the rest (85%) are merely mentioned.[2] Only every fourth news story from the Romani life is accompanied by visual representations,, usually group photos.

While writing the report on human rights in Serbia in 2006, the associates of the Belgrade center for human rights, followed the work of daily newspapers *Политика, Вечерње новости* and *Блиц*, weekly magazines *Време* and *Нин,* the news agencies – national news agency *ТАНЈУГ,* private agencies *БЕТА* and *ФОНЕТ,* and internet pages of Radio *Б92.* We have singled out the information referring the members of the Romani national minority (*Human rights in Serbia in 2006.*2007:195-96, 211, 212):

"The Roma, whose number varies between 108,000 (according to 2002 census) and 800,000 (according to the estimates of the activists in Romani organizations), are the most compromised ethnic group in Serbia. Only one in 100 of the Roma lives more than 60 years, and their average life span is 40. About 80% of the Roma are completely or functionally illiterate, and less than 10% of Romani children go to kindergartens. Only 15,000 out of estimated 80,000 of Romani children go to Primary schools, and only 28% of the former finish it. About 7,8% of the Roma finish high schools and only 3 out of 1,000 finish colleges and faculties (*Данас,* March 31, p. 6; *Време,* May 4, pg 38 and *Политика,* April 8, p/ 10).

The social and economic position of the Roma in Serbia is testified by the data indicating that they are the only inhabitants of more than 600 unhygienic settlements and that the percentage of the unemployed among them is 90% (*Време*, February 14, pg 38).

The Roma are the most common victims of nationality-based incidents.

In mid-February, Bogdan Vasiljević, an employee of the sport and recreation center Krsmanovača in Šabac, was sentenced to a year parole because he refused to allow the Roma to enter the swimming pool on June 8, 2000 (*Блиц,* February 14, pg 13).

[2] The authors give an example where in the press releases from the reception organized on the World Day of the Roma 2002 by the representatives of the Romani minority, there have been quotes of the President and the Federal Minister for National and Ethnic Relations, but not a word from the organizers themselves.

During the same month, the Human Rights Fund (HRF), on behalf of the Romani citizen Dragiša Ajdarević, before the Regional Court in Niš, brought charges against Oliver M. and Nataša M. who had beat up Ajdarević in April 2004 because of his Romani origin.

In March, the Regional Court in Novi Sad sentenced Dolf Pospiš to a year in prison for provoking national, racial and religious hatred by writing Nazi symbols on the walls and threatening to kill the underage DJ for his slightly darker tan in 2004 (*Вечерње новости,* March 28, pg 12).

Towards the beginning of June, G17+ excluded their local official from Pančevo, Zlatko Bekić, against whom the charges were brought for provoking national, racial and religious hatred. During a party meeting, Bekić declared, "we should not worry about the Roma and the Jews since there are already gas chambers ready for them" (*Политика,* June 9, pg 7). During the same month, the unknown persons insulted and beat up three people of Romani nationality in a Belgrade neighborhood Borča. The police have identified the offenders and brought charges against them, but no information about the further development of the case were available at the time when this report was finished (*БЕТА*, June 20). A few days later, the inhabitants of the Romani settlement in Jagodina complained that unknown persons had stoned their neighborhood for a few nights (*Вечерње новости,* June 26, pg 11).

In July, in Ripanj, seven Roma were beaten up by a group of young people who they claimed were skinheads. The police qualified this incident as a common fight (*Блиц*, July 10, pg 14). Some ten days later, in Valjevo, a group of Romani high school kids has been attacked, and the local police again claimed it was nothing but a fight (*БЕТА,* July 21). Towards the end of the month, in the police station in Novi Sad, the police allegedly beaten up four Romany young people who suffered serious body injuries. The police stated that the young people were interrogated for theft, that they were in possession of stolen goods and said the investigation was to be started. Prior to completion of this report, there was no further information about this case (*БЕТА,* July 28, *ТАНЈУГ,* July 29).

In November, in Belgrade, two skinheads were arrested for physical assault on one of the Roma and charges were brought against them (*Блиц,* November 18, p. B5). During the same month, a group of hooligans in Jaša Tomić, allegedly beat up a Romani woman. Romani activists claimed that the assault was nationality based, while the inhabitants claimed that it was a common fight (*Вечерње новости,* November 8, p. 16).

In the beginning of February, two Romani minors, allegedly, were beaten up by the police in Subotica. According to the claims of the minors and their parents, they were beaten up in order to confess to theft. The police confirmed that they were taken into custody, but denied torturing them (*БЕТА,* February 23).

In July, five police officers from Novi Sad allegedly beat up four Romani men and insulted them on nationality basis. The police officers acquired a confession to theft and a statement that they were not beaten up by them by way of torture (*Вечерње новости*, August 1 and 2, pg 12).

Dragana Mašović (2000) conducted quantitative and qualitative analyses of the texts about the Roma in daily newspapers by monitoring the issues of seven daily newspapers during a seven-day period. It is important to mention that three of the newspapers were of pro-Government orientation, three were considered to be pro-oppositional, while the seventh was a local newspaper published in Niš. In 49 newspapers, there were only eight articles mentioning the Roma. Apart from the fact that there is not enough space dedicated to the Roma in the daily issues (no texts or short and poor texts), general attitude towards them contains yet another problem. All the texts are situated on less attractive positions in the newspapers, and, usually, the articles lack critical comments that would have given a more detailed local representation and better explanations. Then, the motives for writing about the Roma are not the difficulties they are facing, but what they have done, what has been done for them or they are merely used as a means of political propaganda. The themes about the Roma which are considered worth writing about, i.e. worthy of public attention, are either trivial or tragic, and often lacking their personal opinion. In addition, the lack of ethnically mixed articles could have long-term consequences. The texts about Roma are usually related to the Roma exclusively. The exception is made when people are mentioned because of their being connected to the topic by duty.

It is concluded in the work that it is not realistic to expect the media to shape our lives by creating a false image of inter-ethnic co-operation and giving too much attention to infrequent positive occurrences. The media cannot create the relations between ethnic groups if they do not exist, notices the author, but what should be expected of them is to recognize this lack of communication as the key problem and to devote their significant attention to it.

Nataša Simeunović (2007), a post-graduate at the Faculty of Political Science in Belgrade, performed an analysis of the texts from the tabloid daily newspaper *Курир* between 2003 and 2007. Aware that earlier research regarded tabloids as frivolous newspapers, less sensitive towards ethnic minorities in comparison to "serious" daily newspapers and given over to simplified and sensationalistic writing, nevertheless the author justifies her choice of the given press medium by the fact that, in domestic conditions, it influences the forming of values of a vast circle of readers who are of various educational levels and age.

In the absence of reports about so-called positive stereotypes (The Roma are merry, hospitable, love freedom, music and dance), the author turned to systematization and explaining of the string of negative stereotypes which, as their consequence, could (and really do) provoke further intolerance towards the Romani people. The published texts supported and additionally strengthened the following attitudes of the majority: 1) the Roma ask for charity[3] , 2) the Roma are dirty[4], 3) the Roma are lazy[5], 4) the Roma are criminals[6] and 5) the Roma are like animals[7]. Discriminatively intoned contents of the news have unjustly labeled the Romani minority in general and lead to the conclusion that it represented a social hazard.

The summary of the image of the Roma in the media in our country will be complemented by an overview of available research on public opinion about the Roma in Bulgaria, the country in the Balkans in which the social position of the Romani minority is the worst and in which they are perceived as "the social hazard". Their way of life is depicted as a deviation from the norm and, therefore, a danger to normality (Milivojević 2003). There is an insistence on their incidental and criminal actions. They are openly blamed for their own poverty and social exclusion because of their unwillingness to finish schools, their asking for charity, aggressiveness etc. (Mantarova 2006). Elisaveta Ignatova (2006) studied the relations towards the Roma created by the electronic media and, beside examples of the foundations of positive values, she also recorded those with the racist background. They were about the comments on inter-ethnic relations in a daily ten-minute show under the aggressive title *Атака* (The Attack), which was on during the prime time on a private cable TV channel.[8] Up until the moment when the

---

[3] "The adult Roma make their children ask for charity and, while the father of the family spends his whole day in a pub, his children "earn" up to 1,000-1,500 dinars" (*Šalju decu u prosjačenje*, Kurir, August 26, 2003).
"At the main bus station, for days, there is a great throng! Bags and pouches all over, hugs and kisses on the platforms, strict watchmen on the ramps and queuing at the windows which sell tickets to seaside... Gypsies of all ages beg the passers-by for some money"(*Krkljnac na štajgi*, Kurir, July 17/18, 2004).

[4] "Dissatisfied natives of the Block 72 also fear that the coming of the Roma will affect public order and the hygiene of their common space and bring the prices of their apartments down to a minimum (*Stoj, Cigani!*, Kurir, July 12, 2005). "As we speak, in the scenery made of cardboard huts, mud, garbage and immense flocks of seagulls, nothing less than "Hamlet" starring the Roma is being filmed!" (*Hamlet na deponiji*, Kurir, January 16, 2006).

[5] "A new brand of "Milka" chocolates is being sold these days in the subway at Zeleni Venac by the price of unbelievable 100 dinars per 3 bars! The sweet bars are in the authentic wrapping that differs from the original only in the last letter of the name. Namely, instead of "Milka", they are called "Milke" and their expiry date is valid. The sellers of the cheap chocolate are the Roma who get their dessert from here and there and do not actually know who and where from brings it" (*Tri Milke za sto dindži*, Kurir, February 18, 2006).

[6] "They mainly live by social aid and theft. They quarrel and fight, and, recently, there has been a fight of two Romani children which resulted in one of them attacking the other with a knife... The local residents claim that the Roma steal dead poultry from the garbage bins of the local farm and sell it afterwards" (*Tortura*, Kurir, December 4, 2006).
"The word is going that, three years ago, she fell in love with a Romani man, Slobodan, who, after a month, made it possible for five men to rape her. After that, she started to sell her body out of despair. She and L.L. claim that prostitution is impossible without patrons" (*Robinje*, Kurir, June 29, 2006.).

[7] The joke of the day in a weekend edition: "A man is walking down the street. He sees a guy carrying a Gypsy over his shoulder. / Hey you! What have you done? / Well, don't you know?! It's an open season on Gypsies. You go to a hunting society, you take a gun and.../ The man goes, takes the gun, sees two Gypsies near a dumpster, shoots and kills them./ A policeman approaches him, takes out his hand-cuffs and arrests him./ But why? Isn't it allowed to hunt the Gypsies? /Well, yes, but not on their breeding grounds!" (*Kurir*, December 27/28, 2003).

[8] During the electoral campaign, the leader of the "Ataka", Volen Siderov, held a series of seven lectures with the topic "Gypsy terror" on a private television "Skat". This

host-journalist of the show became politically involved and attained unexpected success on the elections on all levels in the country in 2005, media analysts, the elite and the politicians paid no attention to his xenophobic and insulting statements aimed at the members of ethnic minorities, and the Roma especially. In addition, the Muslim Roma in Bulgaria are perceived as *the reservoir of Islamic fundamentalism.* In serious academic debates (Колев 2005; Тодорова 2005) the assumptions of spreading untraditional, radical Islam among the socially deprived, isolated, poor and hopeless people were openly considered. It was cited that the Romani followers of Islam, those of Christian origin not being excluded, fulfill all given criteria.

Topalova and Novakov (2006) monitored the articles of different political orientations and attitude towards the Roma in three daily newspapers with large circulations – "Сега" and "Новинар" with the moderate political position and "Ataka", the paper of the extremely nationalistic political party between 2000 and 2006. "Сега" published considerably more articles with positive connotation than with negative and placed two times less reports on criminal activities of the members of the Romani population than "Новинар". Over the period of six years only one fifth of the articles in "Сега" and "Новинар" contained the ethnonym "Gypsy", while the rest (over 80%) contained the ethnonym "Roma" which is politically more acceptable in public debates. "Сега" and "Новинар" used the ethnonym "Gypsy" in the negative context in only 5.8% of their titles, i.e. in 4.8 % of the cases. In "Ataka" however, after only six months of the foundation of the party with the same name, there were 98 articles the titles of which contained the words "Gypsy" and "Roma" ("Сега" published a total of 335 in six years and "Новинар" of 215 texts, among which, as we have already underlined, there were more of those with positive background), and 79% of them were of negative quality. In this manner, the Roma are labeled by the media in an unacceptable way when it comes to practicing politically correct public speech, and the majority of them act as catalysts and call for stigmatization and discrimination of the weakest ethnic group in the country. The titles with the negative connotation comprise 70.4% of the total number of published articles, and, among these, in 88.7% of the cases the articles are dedicated to the description of the criminal action of the "sooty guys". The prevailing image of the Roma is completely transparent: the lowest positions on all ladders are guaranteed; they are "the bad guys" who should be not only avoided, but also publicly branded and severely published. It is obvious that there await us many further efforts to annul the consequences of these manipulations of public opinion and to promote the language of tolerance towards ethnic minorities on both national and regional levels.

radical right wing party won 8.25% of the votes on the parliamentary elections in 2005 using slogans "Out with the Turks and the Roma!", "Gypsies to work camps" and "No to Gypsinization!".

From "The Roma next to us" to "The Roma with us"

The Roma are the most socially neglected minority in our surrounding. There are various causes for such a state. They are without strong regional concentration, they live in unurbanized areas, at the outskirts of cities first of all. They have the highest rate of illiteracy and the lowest rate of employment (Đorđević, D. B., Živković, J. and D. Todorović 2002a, 2002b; Jakšić and Bašić 2002, 2005). They do not have social or political power, and their representation in governmental institutions is insignificant (Bašić 2004).[9] The Roma belong to a group towards which the majority peoples express deep prejudices and stereotypes which prevents their development of the higher levels of adaptation and integration on the level of primary and secondary social relations (Đorđević 1998; Jakšić 2001/2002). They are accused of being responsible for their own fate, unable to change reality and beyond any help.

The third anniversary of Serbian inclusion in the *Decade of the Roma inclusion 2005-2015* is near. The interest for the Roma in the world, the challenges of European integration and the pressure from within by the "third" sector have directed the authorities in our country towards taking concrete measures to provide basic conditions for quality integrations of the Romani into the wider community. We are at liberty to say that there are no stigmatization of the Romani by the highest authorities in Serbia, which is confirmed by the statements of the leading Roma. Their further expectations are directed towards local officials and they range from reproaches for slow application of the law to accusations of exclusion from planning and participating in

[9] The Roma managed to provide, thanks to their political parties, two members of the Parliament before the last Parliamentary election in Serbia, but they did not manage to reach the limit in the election 2008.

local events. And for slow reaction and not punishing the individual cases of discrimination, of course. On the other hand, individual examples of breaking the law by the Roma themselves, on which the media abundantly report, provoke the negative odium of the majority citizens and encourage stigmatization "from below". Requests for realizing the rights of the Romani and other minorities are perceived as downright threatening to the majority rights. We are in for an exhausting effort to change the conscience of the citizens when it comes to the Roma.

Connoisseurs of the majority-minority relations point out that "the contemporary (general) approach overpowers the phrase "the problems of the Roma" towards "the problems of non-Roma" when it comes to understanding and accepting Romani uniqueness for the purpose of integrations and tradition preservation" (Hrvatić 2004:368). The roots of expectedly poor results of Romani inclusion in social space in Serbia, and, we dare say without exaggeration, in the majority of the Balkans countries, lie in the following: 1. acquired stereotypes (we most often meet Romani beggars and smugglers); 2. the existence of a significant social distance; 3. the lack of respect and understanding for the Romani language, history, tradition, art and customs (insignificant coverage in the media and educational programmes); and 4. rare close encounters of Romani and non-Romani citizens. In a nutshell, *in the insufficient knowledge among non-Roma about Romani culture*.

The important role in overcoming the discrimination and reaching the equal treatment of the Romani community in Serbian society in the future will be played by their media representation as a collective with the specific ethnic and cultural identities and tradition. After decades of creating stereotypes about the Roma, now it is time for the media to take on the main role in diminishing and removing the stereotypes connected to the Roma. The story about positive discrimination in the media space, apart from providing equal presence to the Roma, means additional sensitivity in editorial approach and reports on existential, social and cultural problems of everyday life which the Romani community is dealing with. Using the Roma as an example, the media will have the chance to promote collective comprehension of difference as not the bare putting up with others but as active respect and tolerance for them. The *Decade of the Roma*, which has already started, is an additional imperative.

Conclusion

The improvement of the media promotion of the Romani minority can be promoted by:

1. the development of higher sensibility by the majority media in the sense that they report more precisely on the differences and that editors and journalists be educated in the field of quality and adequate presentation of minority problems to the public;
2. introducing the public to Romani tradition and cultural creativity by appropriate programme contents on national and local public services in primetime;
3. starting a campaign to teach illiterate people how to read and write and raise awareness about the necessity of including the Romani children in the school system;
4. the encouragement of bilingual and multilingual shows in the electronic media with the purpose of acquiring and getting used to the language of the environment,
5. providing translations of the broadcasts in the Romani language for the wide audience who use the language of the majority as common communication channel;
6. broadening the programmes in the Romani language on the national TV- RTS;
7. employing a certain number of journalists of Romani nationality on national and local news services;
8. the organization of a training in journalism and technology for media associates from the Romani community in local areas (assistants in media teams; correspondent reports about daily events; team work in big radio broadcasts and campaigns, jobs in editorial offices which produce specialized show in the Romani language etc.);
9. encouraging the media to form Romani regional editorial offices;
10. founding and strengthening independent and professional Romani media;
11. the affirmation of the ethical attitude that it is unacceptable to report about individual incidents in a way which brands the whole community;
12. pointing out and criticizing open racist outbursts and the community itself if it does not punish them;
13. discovering affirmative stories instead of focusing on negative news about the Roma;
14. investigative reports in the media about aspects of the marginalized position of the Roma.

References

Babić, D. (2006). *Romi u medijima BiH (Roma in the media of Bosnia and Herzegovina))*. Sarajevo: Analitički i monitoring centar MPI (Analytical and Monitoring center MPI).
www.mediaplan.ba/docs/analizaRomiUMedijimaBiH.pdf.

Bašić, G. (2004). Političko-pravni položaj Roma u Srbiji (Political and legal position of Roma in Serbia). In *Romi – od zaboravljene do manjine u usponu (The Roma – From a forgotten to a Rising Minority)*, edited by D. B. Đorđević, pp. 71–78. Niš: Odbor za građansku inicijativu.

Vilović, G. (2003). *Image* nacionalnih manjina u hrvatskim medijima (Image of national minorities in Croatian media). *Međunarodne studije (International Studies)* 3(3):19–26.

Đorđević, D. B. (prir.) (1998). Interkulturalnost versus getoizacija i diskriminacija: slučaj Roma (Interculturality versus ghettoization and discrimination: the case of Roma). In *Interkulturalnost versus rasizam i ksenofobija (Interculturality versus racism and ksenofobia)*, edited by B. Jakšić, pp. 335–42. Beograd, Forum za etničke odnose.

—— (2004). *Romi: od zaboravljenje do manjine u usponu (The Roma – From a forgotten to a Rising Minority)*. Niš: Odbor za građansku inicijativu.

—— (2005). О специјалистичким и магистарским ромолошким студијама (У оквиру Центра за студије балканских култура Универзитета у Нишу) (About a post-graduate diploma and MA programme in Romani studies /A part of the Centre for studies of Balkan cultures at the University of Niš/). *Теме (Themes)* 29 (4):461–78.

—— (2007). Romologija (Romology). In *Sociološki rečnik (Sociological Dictionary)*, edited by Aljoša Mimica and Marija Bogdanović, pp. 345–49. Beograd: Zavod za udžbenike.

Đorđević, D. B. i J. Živković. (prir.) (2002). *Romi na raskršću (Prava manjina i manjinskih zajednica) (The Roma at the Crossroads /Minority Rights and Rights of Minority Communities/)*. Niš: PUNTA i Društvo dobre akcije i Komrenski sociološki susreti i Bahtalo drom.

Đorđević, D. B., D. Todorović and L. Milošević. (2004). *Romas and Others – Others and Romas: Social Distance*. Sofia: Institute for Social Values and Structures "Ivan Hadjysky".

Đorđević, D. B., J. Živković i D. Todorović. (2002a). Romanies at Flea Markets: Mirror of Serbian Economy. *Теме (Themes)* 26(1):121–42.

——. (2002b). Unemployed Roma: Poll Research. *Теме (Themes)* 26(4):569–95.

Đurić, R. (2006) *Istorija Roma (History of the Roma)*. Beograd: Politika.

Živković, J. (2004). Medijski sadržaji i manjinske zajednice na jugoistoku Srbije (Medium contents and minority communities on the Southeast Serbia). In *Modeli kulturne politike u uslovima multikulturnih društava na Balkanu i evrointegracionih procesa (Models of cultural policy under the conditions of the multicultural societies at the Balkans and the Euro-integration processes)*, edited by Lj. Mitrović, D. B. Đorđević and D. Todorović, pp. 151–70. Niš: Institut za sociologiju Filozofskog fakulteta Univerziteta u Nišu.

Zeković, A. S. (2006). *Ljudska i manjinska prava u praksi (Human and Minority Rights in Practice)*. Podgorica: Centar za građansko obrazovanje i Centar za razvoj nevladinih organizacija (Center for Civic Education and the Center for Development of NGOs).

Ignatova, E. (2006). Ethnicity and Mass Media in Bulgaria. In *Geokultura razvoja i kultura mira na Balkanu (Etnička i religijska pozadina) (The geoculture of development and the culture of peace in the Balkans /Ethnic and religious background/)*, edited by Lj. Mitrović, D. B. Đorđević and D. Todorović, pp. 189–203. Niš: Institut za sociologiju Filozofskog fakulteta Univerziteta u Nišu.

Jakšić, B. (2001/2002). Romi između diskriminacije i integracije: društvene promene i položaj Roma (Roma between discrimination and integration: Social change and the status of Roma). *Filozofija i društvo* 19–20:333–55.

Jakšić, B. and G. Bašić. (2002). *Romany Settlements Living Conditions and Possibilities of Integration of the Roma in Serbia*. Belgrade: Ministry of Human and Minority Rights Serbia and Montenegro.

——. (2005). *Umetnost preživljavanja (Gde i kako žive Romi u Srbiji) (The art of survival – How and where lives Gipsy population in Serbia)*. Beograd: Institut za filozofiju i društvenu teoriju.

Jovanović, Đ. (2003). Mimikrija Roma – poslednje utočište (Roma mimicry – the last resort). In *Kultura u procesima razvoja, regionalizacije i evrointegracije Balkana (Culture in the processes of development, regionalization and Euro-integration of the Balkans)*, edited by D. Todorović, pp. 345–49. Niš: Institut za sociologiju Filozofskog fakulteta Univerziteta u Nišu i Sven.

Kanižaj, I. (2006). *Manjine – između javnosti i stvarnosti (Nacionalne manjine u novinama 2001-2005) (Minorities – between the public and reality /National minorities in newspapers 2001-2005/).* Zagreb: Sveučilišna knjižara.

Колев, А. (2005). Циганите-Мюсюлмани и ислямският тероризъм. В *Рискове за България от ислямски фундаментализъм и тероризъм*, съставители В. Проданов и Б. Тодорова, стр. 172–83. София: ИК ИФИ–БАН. (Kolev, A 2005 The Gypsies-Muslims and Islamic terrorism. In: Risks for Bulgaria of Islamic Fundamentalism and Terrorism, ed. Prodanov, V., Todorova, B. pp.172–83, S. IPhR-BAS)

(2007). *Ljudska prava u Srbiji u 2006 (Human rights in Serbia 2006).* Beograd: Beogradski fond za ljudska prava.

Мацура, М. (ур.). (2000). *Цигани/Роми у прошлости и данас (Gypsies/Roma in the past and today).* Београд: САНУ. (Belgrade: SANU)

Мантарова, А. (2006) Престъпността в България – проблеми и рискове. В *Българското общество пред присъединяване към Европейския съюз*, съставители А. Мантарова и М. Захариева, стр. 115–29. София: Фондация "Фридрих Еберт" и Институт по социология на БАН. (Mantarova, A. 2006 Criminality in Bulgaria – problems and risks. In: The Bulgarian society before affiliation towards European Union, ed. Mantarova, A., Zaharieva, M., pp.115–29, S. Foundation Fridrich Ebert and Institute for sociological studies-BAS.)

Marko, D. (2008). *Romi u medijima BiH i regiona (Roma in media and BiH region).* Sarajevo: Analitički centar Medijskih inicijativa i Media plan instituta (Analytical Center of Media and Initiative Media Plan Institute).
http://www.mediaonline.ba/ba/

Mašović, D. (2000). The Romanies in Serbian (Daily) Press. In *Роми – социолошки увид (The Romanies – sociological insight)*, edited by D. B. Djordjevic, pp. 99–107. Ниш: Комренски социолошки сусрети и Пеликан принт.

Milivojević, S. (2003). *Media Monitoring Manual.* London and Belgrade: Media Diversity Institute and Samizdat B92. www.media-diversity.org.

Milošević, L. i G. Stojić-Atanasov. (2003). Predstavljanje nacionalnih manjina: analiza sadržaja tiskanih medija u Srbiji (Presentation of national minorities – Content analysis of publishing media in Serbia). *Međunarodne studije (International Studies)* 3(3):60–78.

Mirga, A. i N. Georgi. (2004). Romi u dvadeset prvom veku (The Roma in the twenty-first century). In *Romi – od zaboravljene do manjine u usponu (The Roma – From a forgotten to a Rising Minority)*, edited by D. B. Đorđević, pp. 11–40. Niš: Odbor za građansku inicijativu.

Митровић, А. (1996). Положај Рома у друштву (Position of Romanies in the Society). In *Положај мањина у Савезној Републици Југославији (Status of minorities in the Federal Republic of Yugoslavia)*, edited by M. Мацура и B. Становчић, pp. 807–18. Београд: САНУ.

(2002). *Nacionalne manjine u SR Jugoslaviji (National minorities in the Federal Yugoslavia).* Beograd: Savezno ministarstvo nacionalnih i etničkih zajednica.

(2003). *Попис становништва – Вероисповест, матерњи језик и национална или етничка припадност према старости и полу (Population census – Religion, language and national or ethnicity by age and sex).* Београд: Републички завод за статистику.

Радојковић, М. (1996). Медији националних мањина између позитивне и негативне дискриминације (Media of national minorities between negative and positive discrimination). In *Положај мањина у Савезној Републици Југославији (Status of minorities in the Federal Republic of Yugoslavia)*, edited by M. Мацура и B. Становчић, pp. 407–19. Београд: САНУ.

——. (2003). Media and Cultural Diversity in Serbia. In *Demokratija i multikulturalnost u jugoistočnoj Evropi (Democracy and multiculturalism in South East Europe)*, edited by G. Bašić, pp. 435–448. Beograd: Centar za istraživanje etniciteta.

——. (2007). Радиодифузни медији и идентитет Рома у Србији (Broadcasting Media and Roma identity in Serbia). In *Друштвене науке о Ромима у Србији (Social sciences about Roma in Serbia)*, edited by Љ. Тадић и Г. Башић, pp. 67–79. Београд: САНУ.

(2006). *Romi u medijima (Roma in the media).* Sarajevo: Analitički centar Media plan instituta (Analytical and Monitoring center MPI). http://obrazovanje.skolanovinarstva.ba/2006-07/tekstovi.asp?modulID=2&ID=3

Симеуновић, Н. (2008). *Негативни стереотипи:* Роми су криминалци *(Анализа текстова из таблоидног листа* Курир*) (Negative Stereotipes: The Roma are criminals /Analysis of*

Texts from the tabloid The Courier*). Теме (Themes)* 32(4):757–760.

Tatalović, S. (2003). Mediji i nacionalne manjine (The media and national minorities). *Međunarodne studije (International Studies)* 3(3):5–7.

Тодорова, Б. (2005). Циганите-мюсюлмани в България: предпоставка за етнически и религиозен конфликт. In *Kulturni i etnički identiteti u procesu globalizacije i regionalizacije Balkana (*Todorova, B. 2005 Gypsies-Muslims in Bulgaria: a precondition of ethnic and religious conflict *Cultural and ethnic identities in the process of globalization and regionalization of the Balknas)*, edited by Lj. Mitrović et al., pp. 149–55. Niš: Institut za sociologiju Filozofskog fakulteta Univerziteta u Nišu i Sven.

Тодоровић, Д. (2006). *Нишка ромолошка школа: Библиографија 1996-2005/Romological School of Niš: Bibliography 1996-2005.* Ниш: Универзитетска библиотека "Никола Тесла".

——. (2007а). Национална, религијска и конфесионална идентификација као фактори друштвене удаљености од Рома у Србији (National, religious and confessional identification as factors of social distance from Romas in Serbia). In *Друштвене науке о Ромима у Србији (Social sciences about Roma in Serbia)*, edited by Г. Башић и Љ. Тадић, pp. 95–114. Београд: САНУ.

——. (2007б). *Друштвена удаљеност од Рома (Етничко-религијски оквир)* (*Societal Detachment from the Roma People /Ethnical-religous Relations/*). Ниш и Нови Сад: Филозофски факултет у Нишу и Stylos.

Тодоровић, Д. и Д. Б. Ђорђевић. (2006). Мањине јесу "европљани": Је ли интеркултурализам предуслов? (Minorities are "Europeans": Is Interculturalism a Precondition?) *Habitus* VII(13): 101–19.

Todorović, D. i L. Milošević. (2004). Društvena udaljenost većinskog stanovništva od romske manjine (Srbija, Makedonija, Bugarska) (Social Distance of the Majority Population from the Roma Minority /Serbia, Macedonia and Bulgaria/). In *Društvene promene, kulturni i etnički odnosi i evrointegracijski procesi na Balkanu (Social Changes Cultural and Ethnic Relations and Eurointegration Processes at the Balkans)*, edited by Lj. Mitrović et al., pp. 225–45. Niš: Institut za sociologiju Filozofskog fakulteta Univerziteta u Nišu i Sven.

Topalova, V. (2005). Натурата и културата в стереотипите за циганите: междукултурното сравнение. (Topalova, V. The nature and culture in the stereotypes about gypsies: cross-cultural comparison) In *Kulturni i etnički identiteti u procesu globalizacije i regionalizacije Balkana (Cultural and ethnic identities in the process of globalization and regionalization of the Balknas)*, edited by Lj. Mitrović, D. B. Đorđević i D. Todorović, pp. 139–47. Niš: Institut za sociologiju Filozofskog fakulteta Univerziteta u Nišu.

——. (2006). Bulgarian Media Shaping the Social Representations of Gypsies. In *Geokultura razvoja i kultura mira na Balkanu (Etnička i religijska pozadina) (The geoculture of development and the culture of peace in the Balkans /Ethnic and religious background/)*, edited by Lj. Mitrović, D. B. Đorđević i D. Todorović, pp. 205–11. Niš: Institut za sociologiju Filozofskog fakulteta Univerziteta u Nišu.

Filipović, M. (2004). Etničke manjine, masovni mediji i multikulturno društvo (Ethnic minority media in a multicultural society). In *Regionalni razvoj i integracija Balkana u strukture EU – balkanska raskršća i alternative (Regional Development and Integration of the Balkans Into the EU Structures – Balkan Crossroads and Alternatives)*, edited by Lj. Mitrović, D. B. Đorđević i D. Todorović, pp. 603–13. Niš: Institut za sociologiju Filozofskog fakulteta Univerziteta u Nišu.

Henkok, J. (2002). Borba za identitet Roma (Struggle for the identity of the Roma). *Nova srpska politička misao (New Serbian Political Thought's)* 9(3):125–36.

Hrvatić, N. (2004). Romi u Hrvatskoj: od migracija do interkulturalnih odnosa (The Roma in Croatia: From migration to intercultural relations). *Migracijske i etničke teme (Migration and Ethnic Themes)* 20(4):367–85.

Book reviews

Ancient and Modern Perspectives in the Theory of Categories

Sergiu Bălan

[for *Studies in the Theory of Categories* (edited by Al. Surdu and S. Bălan), vol. I, Bucharest, Romanian Academy Publishing House, 2009]

Since the beginnings of philosophical thought, the categories were man's main instrument for understanding reality and himself. If, on the one hand, ancient philosophers believed that categories belong both to our mind, and to the structure of reality itself, on the other hand, modern thinkers proclaimed there is a fundamental gap between thinking and reality, and therefore there is also one between categories as pure mental entities and as structures of existence. That new perspective determined a profound crisis in philosophical understanding of categories, which finally imposed new ways for using those most general concepts as instruments of investigating different fields of reality, since an all-encompassing idea of reality seems no longer possible today. As a result, contemporary philosophers no longer believe in the possibility of a unique set of fundamental categories, and instead we have today different kinds of categories, corresponding to different fields of investigating existence: logical, mathematical, scientific, aesthetic, sociological, psychological, ethical categories. Nevertheless, the study of these basic instruments of human thought represents a constant concern of philosophical inquiry, and therefore it constitutes the main theme of an extended research project of the Institute for Philosophy and Psychology of the Romanian Academy, "The Classical and the Modern in the Theory of Categories". The Institute also initiated the publication of a series of volumes of studies, intended to include not only the results of the researchers from the Institute, but also contributions of other Romanian and foreign scholars in the field of theory of categories. The first volume of this series is Studies in the Theory of Categories (edited by Al. Surdu and S. Bălan), vol. I, Bucharest, Romanian Academy Publishing House, 2009 (Studii de teoria categoriilor, vol. I, Bucureşti, Editura Academiei Române, 2009).

The volume opens with Al. Surdu's extended paper on Categories, Genera of Categories and Super-categories, concerned mainly with an interpretation of Aristotle's Categoriae from both classical and modern logical perspectives, but also with the possibility of today using genera of categories and super-categories in the field of systematic philosophy, in the tradition of classical German idealism. This issue offers the author the possibility to refer to an important aspect of Aristotle's philosophy, i.e. the prejudicative forms, which represent the main framework for the setting up of the Greek philosopher's list of categories. These prejudicative forms refer to special relations between objects, concepts and words that are supposed to determine the nature, type and use of categories.

The study of Al. Boboc, Category and Categories. Notes on the Possibility of a General Theory of Categories raises the problem of the possibility of identifying new structures of categories in the context of contemporary thinking. On the one hand, we no longer believe today that thinking already has all the conceptual structures needed for understanding reality, for in that case nothing new is possible; on the other, since the concept of reality is always changing today, there is a need to improve our conceptual schemes every day.

In his paper, Two Concepts of Knowledge, Mircea Flonta analyses two ways of understanding the concept of knowledge, Kantian philosophy and modern empiricist theory of knowledge, which are usually seen as opposite points of view. The author argues that this is a wrong perception. The article examines a distinction between two concepts of knowledge: a form-directed concept of knowledge, sketched out by Kant himself, and a content-directed concept, promoted by traditional metaphysics. When they are examined in this light, both Kantian and empiricist, scientifically inspired concept of knowledge may be seen as form-oriented. This quite surprising kind of convergence between Kantianism and modern empiricism proves once more the perennial value of the Kantian approach to knowledge.

The same concern with Kantian philosophy is present in the paper Kantian Categories. The A Transcendental Deduction by Ioan Buş. The author argues that the deduction of categories cost Kant a major part of the time needed for the development of the Critique of Pure Reason. The difficulty of the problem is counterbalanced by its importance in the entirety of the Kantian work. This paper is a short commentary on the Kantian arguments, in the order implied by their development throughout the chapter. This is a chance to address some specific theoretical issues: the strategy and meaning of transcendental deduction by analogy with the exposition of concepts in the Transcendental Aesthetics and in relation to the Metaphysical Deduction, the type of argumentation employed by Kant and the role of some key-concepts such as "unity" and "transcendental object".

The ontological function of categories comes under scrutiny in Ilie Pârvu's paper, How is it Possible to Have an Ontology Nowadays? Modalities of Justification of the Categorial Schemes. As a fundamental inquiry into the meaning of Being and Becoming, ontology (or metaphysics, for analytic philosophers) is at present in search of a new way of justifying its cognitive claims. In this study we find a justification of a categorical scheme in terms of a sui generis philosophical argument, by appeal to naturalistic or inductive strategies. The author argues that philosophical discourse on being can be legitimated is by an attempt, analogous to the one made in contemporary cosmology, to "construct" the object of research by the extension of an ontological model, worked out through the "metaphysical conversion" of the structural-generative mode of thinking in contemporary science, to all the levels and horizons of existence. In this way "existence as a whole" can be constructively defined.

Philosophical categories have also an important methodological function, as the article Categorization and Problematicity by Constantin Sălăvăstru tries to prove. This study aims to analyze the main categories that trace the defining framework of a significant direction of philosophical meditation: problematology. The author's interest focuses on identifying the pragmatic (applicative) dimension of the problematologic model of thought (founded by Michel Meyer). Concepts such as problematological situation, questioning, problematological difference, radical interrogativity are defined and operated with so as to explain topics of knowledge from fields such as metaphysics, argumentation, rhetoric or theatre. Elements of methodological originality associated with perennial aspects of philosophical reflection or other types of reflection bring out for careful consideration a way of conceptualizing that seems to oppose certain genuine traditions in the field.

The specific function of aesthetic categories is presented in Vasile Macoviciuc's article Logical-Ontological Categories in Georg Lukács' Aesthetics. The paper refers to Georg Lukács' adventure towards elaborating an aesthetics under a Marxist paradigm, which employs some aspects, postulates and issues that are specific to the ontology of the human. The problem of aesthetic subjectivity, the issue of mimesis and understanding art as a form of self-consciousness rely on an entire cultural tradition, from Aristotle to Hegel, and are dealt with starting from an ontological reference point. This is why one can speak about a Lukács against Lukács, inasmuch as the hermeneutic dimension of his system of ideas holds the front row, shading or canceling hesitations, ambiguities and/or his own critical sentences with an immediate ideological – even partisan – bias, especially related to contemporary aesthetical experiences.

The wide philosophical category of culture is the object of a comprehensive analysis in the contribution of Marin Aiftincă, Culture in Contemporary Outlook. From this perspective, culture is one among the fundamental categories of contemporary philosophical thinking. The article treats the category of culture from the point of view of modern philosophical reflection and in correlation with the challenges of today's world. The concept of culture is reconsidered from an axiological point of view and the author is also interested in explaining the perennial relations between culture and nature, culture and society, culture and civilization. At the same time, the author is interested in the understanding of the category of culture in the light of the most recent developments in the philosophy of culture: culture's unity and plurality, cross-cultural communication, cultural identity and "Global Culture", the idea of a new world cultural project.

Some of the logical aspects of the problem of categories are the object of Iancu Lucica's Paper, Consistence, Inconsistence and Para-consistence of Concepts. The subject of this study is the theme of paraconsistent concepts. The author is concerned with the consistency and inconsistency of the concepts considering that the paraconsistence of the concepts is implied by both of them. From this perspective, he also treats some issues regarding existence and questions the issue involving non-monotonic reasoning. In his opinion, paraconsistent concepts involve various forms of non-monotonic inferences such as the exceptive syllogism, to take just one example.

The same logical point of view on the problem of categories is present in Marius Dobre's paper The Category of Logical Value. The article argues that the problem of truth is the most important problem of logic and it can be conceived as the most fundamental category of this domain. The logical truth is an issue presented by the author from the point of view of the classical theory (Aristotle), but also of modern theories (Russell, Carnap, Lukasiewicz, Tarski, Brouwer, Heyting).

The article The Categorial Task of Soul in Plato's Dialogue "Philebos", by Rodica Croitoru, is a historical analysis of certain aspects of the Platonic view on categories. The categorical guide marks drawn out by Plato in Philebos, the author believes, are conceived neither with the purpose of building up a system, nor with that of introducing unity of thought; they aim first at getting together the visible world with the intelligible world; and secondly, they aim to systematize the knowledge of the visible world, starting with the superior one, realized by understanding, to the inferior one realized by senses. As a result of the two purposes aimed by these ones, the cognitive value of categories will be based on an ontological ground, due to which they could assimilate the structure of prime ideas which are the ones of beauty, good, truth. In order to satisfy these purposes two instruments will be necessary, one of them having ontological-cognitive character, respectively the understanding, and the other one having multiple existential aspects as well, respectively the soul.

The problem of categories is of the utmost importance in the analytical tradition as well, and therefore some important analytical philosophers areconcerned with a new interpretation of classical points of view on this topic. The way this is achieved is the focus of Sergiu Bălan's paper Hegelian Categories from an Analytical Point of View. The author maintains that, according to the so-called analytic philosophy's "creation myth", as Russell puts it, it was explicitly conceived and designed as anti-idealistic and especially anti-Hegelian. However, this myth was only a half-truth, and more recent philosophers (beginning with Sellars) strive to recover what they considered the valuable part of German idealism, in the context of some sort of attempted reconciliation with Kant and Hegel. The paper deals with the pragmatist-inferentialist analytical interpretation of Hegelian categories, formulated by one of Sellars' followers, the American philosopher Robert Brandom, whose 1994 book Making It Explicit was described by Rorty as "an attempt to usher analytic philosophy from its Kantian to its Hegelian stage".

Of course, not only analytical philosophers were recently concerned with a reconsideration of the problem of categories, but also thinkers from the "Continental" tradition, as Claudiu Baciu tries to prove in his paper Ernst Cassirer on Symbolic Form and Synthetic Unity. Starting with the explanation of the distinction made by Cassirer between the concept of substance and the concept of function, the paper further presents the way in which the author of the Philosophy of Symbolic Forms interprets human knowledge in general as a result of a synthesis. In contrast to Kant, the synthesis is not interpreted by Cassirer as the product of a single human faculty, but is thought as being specific to every type of science – being also irreducible to other factors – and, after the development of Cassirer's philosophical project, as being specific to every major domain of culture.