

BALKAN JOURNAL OF PHILOSOPHY

Vol. 5, Issue 2, 2013

CONTENTS

Articles

Semiotic hypostases of dispositional explanation / 131

Teodor Dima

The endless truth in the P. M. formal systems

(In honor of the *Principia Mathematica* (1910-1913) of Bertrand Russell) / 139

Mihai D. Vasile

Communication and artistic expression / 147

Marin Aiftincă

Self-deception and the selectivity problem / 151

Marko Jurjako

An acquaintance constraint and a cognitive significance constraint on singular thought / 163

Mirela Fuš

Chomsky and Foucault on human nature – a perspective for reconciliation / 175

Ana Smokrović

Sceptical theism and the problem of epistemic evil:

Why sceptical theism is philosophically costly / 181

Jimmy Alfonso Licon

The role of rationality in the formulation of and compliance with the principles
of justice / 191

Peter Cholakov

The status of philosophy – sources of the Spinozan conception of the freedom
of philosophy / 199

Gabriela Tănăsescu

Discussion

Pro and con discussion regarding the tenets of the hermeneutic philosophy of science (continuation) / 209

On the divergent being of science's theoretical objects / 209

Dimitri Ginev

Book reviews

Modern World Issues from the Perspective of Anthropology / 221

Marțian Iovan

[For *Antropologia și problemele lumii moderne (Anthropology Faced with the Problems of Modern World)* by Lévi-Strauss, Claude, Preface by Maurice Ollender, Translation by Giuliano Sfichi, Iași, Polirom Publishing House, 2011]

Information for contributors

SEMIOTIC HYPOSTASES OF DISPOSITIONAL EXPLANATION

Teodor Dima

Member of Romanian Academy,
Manager of the Institute Economics and Sociale
Researches, Branch Iași of Romanian Academy,
e-mail: tudoreldima2008@yahoo.com

Abstract:

We consider that the epistemological recourse to semiotic dimensions could lead to peace: those in favour of an epistemology without a knowing subject are right (if they keep themselves within the syntactic and semantic frames of theoretical construction) and those who add a knowing subject are also right (if they want to express the concrete process of building up scientific convictions).

Key words: semiotic strategy, syntax, semantic, pragmatic.

1. Within our prior pursuit of scientific explanation¹, we approached a dispositional explanation² for the purpose of discussing the way in which some specific types of explanation are structurally adaptable to the models proposed by Hempel and Oppenheim³. Now we will approach this type of explanation through a *semiotic strategy* that will open the door to an epistemology with and without a knowing subject.

The idea of approaching a logical, philosophical or scientific category from the triple perspective of semiotics belonged to Petre Botezatu⁴ who observed

that, before the semiotic viewpoint was promoted, the elaboration of scientific theories had been a spontaneous process, depending on the author's intuitions and subjective requirements as far as science was concerned. Such theoretical constructs, although with valuable aspects, suffered from logical vices more or less concealed. For example, on the one hand Euclidean geometry contains propositions not needed for its construction, but, on the other hand, it does not include all the necessary elements for a flawless demonstration. Only in the 20th century were the shortcomings removed by means of semiotics making new forms of geometry possible.

Today, Petre Botezatu observed in 1980, one cannot elaborate a scientific theory without meeting certain metalogical or metascientific requirements: namely, to have previously checked all materials to be used and all mechanisms necessary to assemble them. Semiotics provides a *multilateral approach* to the topic under study and so the fundamental aspects of the study are covered by means of the three constitutive parts of semiotics: *syntax*, *semantics* and *pragmatics*.

Syntax examines the way in which signs associate with one another when elaborating a theory and the internal coherence of the final theoretical construction. *Semantics* studies the relation between signs and the things to which they refer; in other words, the correspondence of the constructed theory to the facts. *Pragmatics* deals with the relation between signs and the subjects using them, including their attitudes towards the theoretical construction followed in practice⁵. Referring to Immanuel Kant, Petre Botezatu observed that he had revealed the

¹ See Teodor Dima, *Explicație și înțelegere*, Vol. I, București, Editura Științifică și Enciclopedică, 1980; Vol. II, Iași, Editura Graphix, 1994.

² *Idem*, *Tipuri de explicație științifică: explicația dispozițională și explicația genetică*, in Mircea Flonta (coord.), *Epistemologia și analiza logică a limbajului științei*, București, Editura Politică, 1975, pp. 153–174.

³ C. G. Hempel, *Aspects of Scientific Explanation*, New York and London, The Free Press, Collier-Macmillan Ltd., 1965.

⁴ Petre Botezatu, *Semiotica, strategie optimă a construcției științifice*, in *Idem*, *Interpretări logico-filosofice*, Iași, Editura Junimea, 1982, pp. 42–53.

⁵ *Ibidem*, p. 46.

triple aspect of causality, which corresponds to the three forms of existence distinguished by *syntax*, *semantics* and *pragmatics*⁶. Thus, in the Kantian system there would appear *transcendent causality* (called “intelligible causality”), *immanent causality* (called “empirical causality”), and *transcendental causality*, an *a priori* category. Generally speaking, the *transcendental* is a typically Kantian creation, an almost autonomous field, to the extent to which knowledge is possible *a priori*. But if they remained abstract, the transcendental categories, causality included, would combine with one another from the logical point of view without being gnoseologically fertile. Petre Botezatu commented⁷ that they were attributed a purely logical existence, beyond space and time, and are conditions of knowledge gained through experience. Therefore, at the ideal level of transcendentalism, there exists a *concept of causality as pure form, not related to space and time*.

To Kant, the pure forms of intellect are defined by their correspondence to the universal logical functions of thinking while the category of causality corresponds to the *classical hypothetical judgement*: *if P, then Q*. From the transcendental perspective, the causal relation is the necessary relation between principle and consequence or implicant and implicate. This relation contains no reference to space, time or substance. “As the concept of cause is concerned, I would not say the pure category is anything more than something from which the existence of something else can be concluded, and not only is not possible for the cause to be distinguished from the effect, but also, because this possibility of deduction immediately needs conditions about which I do not know anything, the concept would not have any determination to indicate the way in which it applies to any thing at all.”⁸

The question we now try to answer refers to the possibility of the transcendental construction of a *syntactic construction* of a scientific explanation belonging to the logic of scientific research. In our opinion, the answer should emphasize the *foundation* on which *that which is to be explained* will rely, and articulate an *inferential approach* able to ensure the correct passage from the foundation (*explanans*) to the thesis (*explanandum*). It was this idea that

Carl Gustav Hempel emphasized in his study “The Function of General Laws in History” published in *Journal of Philosophy* in 1942. His opinion was widely discussed, even if in 1935 Karl Popper had referred to the logical structure of scientific explanation starting from a concrete example⁹, one addressing causal explanation, that is, in Kantian terms, an explanation of the *immanent* or *intelligible causality*.

To explain causally an event means to infer a statement that describes it from *universal laws* and certain singular statements of the *initial conditions* under which the event took place. For example, Popper says that we “causally explain” how we break a thread if we establish that its resistance to breaking is 1 kg and a 2 kg weight was suspended from it. This “explanation” has several elements: first, the statement “Every time a thread sustains a weight exceeding its resistance to breaking, it will break” which is a similar to a law of nature. Second, the singular statements (there are two in this example) describing the particular case: “For this thread, its resistance to breaking is 1 kg” and “The weight hanging from this thread was 2 kg.”

Therefore, we add, syntactically in *explanans*, that there are two types of propositions: *universal* ones (“For any x , if x is S , then x is P ” – $(x) [x(S) \rightarrow x(P)]$) and *enunciative singular* ones (“This x is S ” – $x_I(S)$). The conjunction of the two types of propositions results in the following enunciative singular proposition: “This x is P ” – $x_I(P)$. Karl Popper amalgamated the syntactic and semantic analyses, immediately passing from propositions to objects and facts: “the initial conditions describe what we usually mean by cause” (the fact that from a thread with 1-kg resistance to breaking a 2-kg weight was suspended was the cause for the thread to break). This semantic negligence is remediated by saying that: *The propositions that state the initial conditions express the cause* and, from the perspective of a scientific logic, they belong to the *explanans* from which the *explanandum* comes.

For the passage from the *explanans* to *explanandum* to be grounded, a *principle of causality* has to be tacitly admitted that expresses the fact that any event can be causally explained. From the standpoint of logical syntax, this principle can be expressed tautologically or analytically, with no reference to reality, as *pure form, unrelated to space and time*, as stated by Kant. Previously we showed that, for Kant,

⁶ *Idem*, *Determinismul kantian din perspectivă semiotică*, in the volume cited above, pp. 205–232.

⁷ *Ibidem*, p. 217.

⁸ Immanuel Kant, *Critica rațiunii pure*, translation by Nicolae Bagdasar and Elena Moisuc, București, Editura Științifică, 1969, p. 248.

⁹ Karl R. Popper, *Logica cercetării*, translation by Mircea Flonta, Alexandru Surdu and Erwin Tivig, București, Editura Științifică și Enciclopedică, 1981, pp. 97–99.

the pure forms of intellect are defined by their correspondence to the universal logical functions of thinking (reason) and the category of causality corresponds to the *classical hypothetical proposition*: If *P*, then *Q*, or in our case: *If explanans, then explanandum*. An antecedent is sufficient and/or necessary for a sequent to result necessarily and/or sufficiently.

In other words, if there is the logical possibility for universal propositions and enunciative singular propositions forming an *explanans* to exist, then a statement, the *explanandum* will necessarily and/or sufficiently follow.

Concerned with the precise demarcation between metaphysics and science, Popper considered he would neither support nor reject the “principle of causality”, but would accept it as a *simple methodological rule* in the field of science, that is as a rule to “never stop searching for the laws of a unitary theoretical system and never give up trying to explain causally any event that can be described.”¹⁰ Thus, Popper passes from syntax to semantics in analysing a scientific explanation.

2. In the same way, amalgamating the elements of syntax with those of semantics, Carl Gustav Hempel resumed the topic of scientific explanation¹¹. His theory became known as the “covering-law model” of explanation, an expression coined by one of his critics, William Dray¹². In our works, we used the expression *explanation by subsuming to laws*. In some of his later published works, Hempel developed, clarified and changed his initial opinions. One of his most significant clarifications is in marking the passage from the role of nomological generalizations to the role of statements that express events or states of facts so that the statement on what is explained is logically inferred.

This moment of Hempel’s analysis corresponds to the passage to the *semantic dimension* of the semiotic investigation. Through *interpretation*, a *model* is built¹³. A famous example of semantic interpretation is that in which the *explanandum* is an event and the *explanans* consists of antecedent events and states. In other words, the intervention of time and space into the explanation leads our intervention to semantics. Petre Botezatu observed that

Kant had the same kind of approach. *Transcendental causality*, the formal design of causality, reduced to the idea of a *logical relation between entities* (for Hempel, between *explanans*, made up of universal propositions and enunciative singular propositions, and *explanandum*, the proposition referring to what is explained), is semantically interpreted within the universe of phenomena in the model of *empirical causality* being elaborated here. Semantically, the transcendental entities are interpreted as phenomena and the logical necessary relation between them is seen as a causal relation that develops in time. Thus, the causal relation is generated through particularization by the general relation of laws, of necessary connection. Therefore, for the causality relation to be possible, the component of real time, of temporal sequence, is needed.

In 1942, Hempel built a semantic *model* of scientific explanation: Let *E* be an event known to have occurred in a certain circumstance and needing to be explained. Why did *E* happen? To answer this question, the circumstances and conditions in which *E* occurred have to be known. This means that research has to be conducted and so the elaboration of an explanation generates a cognitive process, more or less complex depending on how much knowledge and understanding will be needed. As the conditions in which the event occurred are understood, statements will describe the conditions which are called *antecedents* and in this way the presence of time is suggested. Let us denote by *C*₁, *C*₂, ..., *C*_{*n*} the *statements on the antecedent conditions*; also, the antecedent conditions have to be subsumed to some necessary relations expressed by *general statements* or *laws*, *L*₁, *L*₂, ..., *L*_{*r*}. These statements should be built in such a way as to be as appropriate as possible in order to subordinate the events that generated the process of explanation. These two types of statements make up the *explanans* from which *E* is inferred¹⁴.

Let us express this semantic construction by means of Hempel’s words from the study published in 1942 and included in the volume almost entirely dedicated to scientific explanation:

The example given by Hempel has an *explanandum E* which is a statement referring to an event, and an *explanans* made up of statements referring to

¹⁰ *Ibidem*, pp. 98–99.

¹¹ C.G. Hempel, “The Function of General Laws in History”, in *Journal of Philosophy*, 1942.

¹² W.H. Dray, *Laws and Explanation in History*, Oxford University Press, 1957, p. 1.

¹³ Cf. Petre Botezatu, *Op. cit.*, p. 222.

¹⁴ Carl G. Hempel, *The Function of General Laws in History*, in *Idem, Aspects of Scientific Explanation. And Other Essays in the Philosophy of Science*, New York, The Free Press, London, Collier-Macmillan Limited, 1965, Sect. 2.1.

antecedent events and states. Thus, *E* has to state why my car radiator broke during the night. It was filled up with water; the radiator cap was sealed; no drop of liquid anti-gel was added; the car was parked in the courtyard; during the night, the temperature dropped unexpectedly much below zero. Those were the antecedent events in the *explanans* rendered by means of statements, $C_1, C_2, \dots, C_n$, in combination with the laws of physics, especially the one expressing the causal link between the temperature dropping below zero and the water increasing in volume when frozen.

Several years later¹⁵, Hempel also published a logical structure of the model that he called *deductive-nomological*:

$C_1, C_2, \dots, C_k$	Statements of antecedent condition	} <i>Explanans</i>
$L_1, L_2, \dots, L_r$	General Laws	
E	Description of the Empirical phenomenon to be explained ¹⁶	} <i>Explanandum</i>

3. Georg Henvik von Wright posed the question “whether Hempel’s model applies only to the objects of explanation that are events”¹⁷. For example, sometimes it has been stated that some behaviours or manifestations of objects should be explained by means of *dispositional properties*, characteristic of those objects. Here we can include especially human behaviour, explainable through people’s character, convictions, purposes, and other dispositional factors¹⁸. The term comes from psychology and was extrapolated to the physical world because the members of the Vienna Circle, who coined it, founded an empiricist epistemology, but the term is still in use today because it has been understood that dispositions are objective aspects of real phenomena. Their characteristic is their hidden nature; that is, that they only manifest themselves upon impact with appropriate circumstances. In other words, “*dispositions are potentialities* that do not fail to actualize” when appropriate conditions are

met. Simplified: *Disposition & Circumstances = Reality*¹⁹.

In 1975, referring to dispositional explanations, I showed that disposition intensity varies because the environment in which the disposition actualizes could be actualization-friendly or -unfriendly, but the important thing is that dispositions depend on the *internal structure* of objects and the *genetic structure* of beings. For example, glass is brittle because its internal structure determines this disposition that manifests itself when the glass is hit with a certain intensity. One may become a great tenor because an internal structure—of the vocal cords, or of a musical ear—actualizes through appropriate training. Therefore, internal structure is a *necessary condition* for the disposition to manifest itself in a certain way, and its impact with certain circumstances is the *sufficient condition for the disposition to actualize*²⁰.

4. Dispositional explanation takes us into the *pragmatic* side of semiotics. Dispositional explanation is pragmatic because it reveals the subject’s attitude towards the signs she uses, and it is confronted with psychological, biological, and sociological aspects of knowledge and explanation’s elaboration. “Relations such as *acceptance, performance, conviction*, etc. take place and their theorization mobilizes very different factors, from organic and psychic individuality to social and moral commands. Thus, it seems that pragmatics is suffocated by a very rich context of concrete phenomena that prevents it from reaching the abstract level of syntax and semantics”²¹.

¹⁵ C. G. Hempel, *Studies in the Logic of Explanation*, in *Philosophy of Science*, Vol. 15, 1948, pp. 135–175, included in *Idem, Op. cit.*, pp. 245–290.

¹⁶ *Ibidem*, p. 249.

¹⁷ G. H. von Wright, *Explicație și înțelegere*, translation by Mihai D. Vasile, București, Editura Humanitas, p. 34.

¹⁸ In the Romanian language, psychologists use the term “predisposition”; we prefer the term coined by the English literature on epistemology.

¹⁹ Mario Bunge, *Possibility and Probability*, in Harper & Hooker (eds.), *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Sciences*, Vol. III, Boston, D. Reidel Publishing Company, 1976, pp. 17–33, Romanian translation, in Mario Bunge, *Știință și filosofie*, București, Editura Politică, 1984, p. 143.

²⁰ In our previous study from 1975 we discussed the distinctions between necessary and sufficient conditions, as far as disposition manifestation is concerned, *Tipuri de explicație științifică: explicația dispozițională și explicația genetică*, in Mircea Flonta (coord.), *Epistemologia și analiza logică a limbajului științei*, București, Editura Politică, 1975. In 1994, in *Explicație și înțelegere*, Vol. II we showed that Mario Bunge presented a similar opinion in *Possibility and Probability*, 1976 (see the previous footnote): “... a disposition is a condition that is necessary but not sufficient” (see the Romanian translation, p. 144).

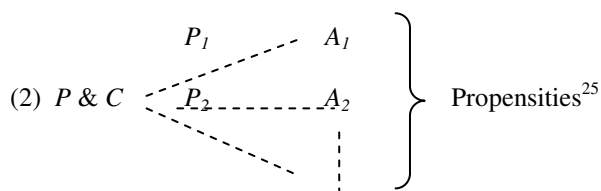
²¹ Petre Botezatu, *Semiotică și negație. Orientare critică în logica modernă*, Iași, Editura Junimea, 1974, p. 100.

Apart from dispositional properties, there are also *manifested properties*, which are always exhibited by the objects, even if they are not observable. In Popper's example, every material point has a position that can be calculated by means of classical physics. These are called stochastic properties, represented by random variables; they *tend to get exact values* while being experimentally approached. Popper called them *propensive* and elaborated *propensity*, as we showed in 1994, as a fundamental notion of his theory of probability²² that results from the measuring of the tendencies of the respective properties to manifest themselves²³.

If dispositional properties manifest themselves necessarily in one way only, when the appropriate conditions are met, for propensive properties, although inherent to objects as well, the *environment* makes the choice, it realizes their necessity. In other words, while the actualization of dispositional potentialities is determined from the inside, the actualization of propensive (tendential) potentialities is determined from the outside. *Internal determination* for disposition and *external determination* for propensity. These differentiations, which belong to us, are expressed by Mario Bunge from another perspective: "In the case of (what he called "causal") disposition things are what they are, while in the case of propensity (towards chance, as M. Bunge added) things are whatever they may become."²⁴

The main distinction between the two types of properties can be graphically rendered. So, *D* is disposition, *P* is propensity, *A_i* (*i* belongs to the set of natural numbers) is a real or manifested property, and *C* is the circumstances that, together with the properties, create the environment. The following designs are obtained:

(1) *D* & *C* *A* Manifested disposition



It follows that propensity does not reduce to disposition. For example, Johnny's mathematics skills are a causal disposition as a result of an aleatory combination of parental genes. His parents, as a

couple, had a propensity to chance determined to conceive a child with mathematical skills²⁶.

The difference between Hempel's model which is considered *standard* by most epistemologists and the structure of dispositional explanations is particularly obvious in *pragmatics*. Rudolf Carnap, the founder of logical syntax, reached pragmatics when dealing with *dispositional predicates*²⁷. He considered that any language *L* can be defined as a system of *dispositions* for the use of the expressions made up through that language. The fact that a speaker *X* is able to use a language *L* means that she has a certain system of interconnected dispositions to express herself through that language. Among these dispositions of *X*, all together called language *L*, there is the disposition to assign predicate "*q*" to any object *y* if and only if *y* has property *F* which can be directly observable or explicitly definable. Such conditions belong to pragmatics.

For example, dispositional notions such as "soluble" and "breakable" are predicates expressing the capacity of some objects to react in a certain way to given conditions. Let *D* be the disposition of glass to break (*R*) when hit with something solid (*C*):

$$(1) \quad (x)(t)\{C(x,t) \rightarrow [D(x,t) \rightarrow R(x)]\},$$

"For any *x* at time *t*, if *x* is hit with something solid, if *x* is breakable, then it will break at time *t*".

Carnap's discourse belonged to the field of pragmatics, which was clearly emphasized in *Meaning and Necessity* where he observed that the research of natural languages was of the highest importance for the understanding of people's behaviour, as well as the nature and evolution of entire cultures. Moreover, he suggested that the study of pragmatics could prove useful to a logician as well. If a logician wants to find an efficient form of a linguistic system to be used in an empirical scientific field, she could find fruitful suggestions in studying

²⁶ *Ibidem*, p. 150.

²⁷ Rudolf Carnap, "Testability and Meaning", in *Philosophy of Science*, Vol. 3 (1936) and Vol. 4 (1937), p. 440 et seqq.; *The Methodological Character of Theoretical Concepts*, in H. Feigl and M. Scriven (ed.), *Minnesota Studies in the Philosophy of Science*, Vol. I, Minneapolis, 1956, p. 62. We use the Romanian version of Carnap's study "Meaning and Synonymy in Natural Languages" published in *Philosophical Studies*, No. 7, 1955, pp. 33-47, Romanian translation by Gheorghe Enescu and Sorin Vieru, in Rudolf Carnap, *Semnificație și necesitate*, Editura Dacia, Cluj, 1972, pp. 298-313.

²² Teodor Dima, *Op. cit.*, p. 81.

²³ K.R. Popper, *Op. cit.*, p. 32.

²⁴ Mario Bunge, *Op. cit.*, Romanian translation, p. 149.

²⁵ *Ibidem*.

the natural development of scientists' language and even everyday language.

The characterization and logical expression of dispositional properties are easy to achieve and regulate at a semantic level. It is the pragmatic level that generates some difficulties because here it has to be established whether an object or a person *X* has disposition *D* at a given moment *t*. Carnap proposed two methods. The first method was called *behaviourist*: condition *C* is generated and then it has to be determined whether there is answer *R*. In other words, simplifying, a spoonful of sugar is poured into a recipient filled with water (*C*) and then it is to be noticed whether the sugar dissolves (*R*). However, there are dispositions whose determination is difficult or achieved with important losses. Moreover, this carnapien method is analytical or tautological: why does this amount of sugar dissolve into water? Because sugar is soluble in water, or sugar dissolves because it dissolves!

The second method which Carnap called the *method of structural analysis* consists in examining in detail the state in which *X* is found at moment *t* as to be possible for one to make judgements based on the description of the state of *X* and the appropriate nomological statements on the reactions of *X* under given conditions. If *X* gives answer *R* under conditions *C*, then *X* has disposition *D*.

The methods proposed by Carnap cannot be called dispositional explanations. In our opinion, they are determinations of dispositional properties that can explain past events or anticipate events likely to occur. By means of *etiologic*, which I structured in 1973 in order to express theoretically scientific discoveries of sufficient and/or necessary conditions, one can elaborate such dispositional explanations²⁸. For example, the non-tautological expression of the discovery of a dispositional property by means of the language used in *etiologic* is the following: "It is impossible that an object *X* should not be breakable and should not break under conditions *C*." That is, disposition *D* ("breakable") is a necessary condition for *X* to break under conditions *C*.

5. After these specifications, we get back to dealing with *dispositional explanation*. Wolfgang

Stegmüller²⁹ considered that the first reference to dispositional explanations was by Gilbert Ryle, a leading figure in the "Oxford" school of analytic philosophy, in his *The Concept of Mind*, published in 1949. There he argued that certain states of mind can be considered from a dispositional perspective. For example, to have an opinion does not only mean to express thoughts consciously but also to say and do things depending on circumstances. To Ryle, the difference between causal explanations and dispositional ones is the formulation of different statements. For example, if we say that a window broke because it was hit by a rock, we give a usual causal explanation; if we say that it broke because it was *breakable*, the explanation is dispositional. In other words, as we argued in 1975, while by a proper causal explanation we express *one of the possible causes* or the *sufficient condition*, by a dispositional explanation we express the necessary condition or the *sine qua non* condition.

An important characteristic of dispositional explanations was stated by Ryle: they work with individual objects, not classes of objects. This aspect is obvious at the pragmatic level, which we meet when dealing with pragmatist epistemologies, where dispositional explanations are structured according to the model of causal explanations. We are referring to the pragmatism rebuilt in the last two or three decades (1980-2010) and we will give the example of one of the papers published by Susan Haack³⁰. She starts from the characterization of convictions made by Price as "multi-form dispositions", complexes of dispositions to verbal and non-verbal behaviour: to assert to or to assert certain sentences in, as Price nearly puts it, unstudied speech, to behave in these or those ways, to feel shock or surprise if such-and-such happens ... etc.³¹

Unlike Ryle who accepted dispositional explanation only for individual objects, Haack tries to

²⁸ For further discussion of Rudolf Carnap's conception see our paper *Tipuri de explicație științifică: explicația dispozițională și explicația genetică*, in Mircea Flonta (coord.), *Epistemologia și analiza logică a limbajului științei*, București, Editura Politică, 1975, p. 157–158.

²⁹ Wolfgang Stegmüller, *Probleme und Resultate der Wissenschaftstheorie und Analytische Philosophie*, Band I *Wissenschaftliche Erklärung und Begründung*, Studienausgabe, Teil 1: *Das ABC der modern Logik und Semantik. Der Begriff der Erklärung und seine Spielarten*, Springer – Verlag Berlin. Heidelberg. New York, 1969, pp. 72–152.

³⁰ Susan Haack, *Evidence and Inquiry. A Pragmatist Reconstruction of Epistemology*, Second Expanded Edition, Prometheus Books, New York, 2009, Romanian translation by Cătălina-Daniela Răducu, Introductory study by Acad. Teodor Dima.

³¹ Price, *Belief*, London: Allen and Unwin, 1969, *apud* Susan Haack, *Op. cit.*, p. 230.

solve the ambivalence she feels when assigning convictions to animals on the one hand, and computers on the other hand. Ambivalence because in every case part of the multi-form dispositional characteristics are there, which means they resemble convictions but are not convictions in the proper sense of the word. From the pragmatic perspective, we consider that this resemblance could suggest a plausible approach of the intensity degrees of convictions by means of the intensity degrees of relevant dispositions.

We think that *pragmatically*, dispositional explanations become causal explanations, which Haack acknowledges in the following example: If semantically a dispositional explanation is given in which the *explanandum* expresses the fact that “the piece of glass broke” because “its internal structure made it fragile” (*explanans*), pragmatically “the piece of glass broke” because “he dropped it on the floor”. In other words, in the first case the reference is to a necessary, permanent condition, while in the second case the reference is to a sufficient condition, one of the possible causes. From the perspective of an epistemology with a knowing subject, Haack says something similar: “The disposition of a person to certain verbal behaviour patterns or of any other kind depends on the state of the person’s brain (in the same way fragility depends on the micro-structural configuration of glass)—in other words, something that depends on the contents of the black box.”³²

In short, a *formal model of dispositional explanation* can be elaborated in which the three semiotic dimensions are included. Let *a* be an object and let us explain why *a* behaves in manner *R*. The explanation can use singular statements regarding both situation *C* in which *a* found itself previously and dispositional characteristic *D*. It results in the general statement on necessary condition *D* which determines reaction *R* under certain conditions *C*:

A₁:	<i>a</i> was in situation <i>C</i>
A₂:	<i>a</i> has characteristic <i>D</i>
L:	Any <i>x</i> that possesses <i>D</i> behaves in manner <i>R</i> in a certain situation <i>C</i>
E:	<i>a</i> behaved in manner <i>R</i>

The schematic structure expresses the *syntactic aspect*, the reference to objects and properties gives *semantic answers*, and the content of statements **A₁**

and **A₂** generates assumptions on the existence of an observer or an agent, which means the *pragmatic dimension* is included.

We consider that the epistemological recourse to the semiotic dimensions could lead to peace: those in favour of an epistemology without a knowing subject are right (if they keep themselves within the syntactic and semantic frames of theoretical constructions) and those who add a knowing subject are also right (if they want to express the concrete process of building up scientific convictions).

References

- Petre Botezatu (1974) *Semiotică și negație. Orientare critică în logica modernă*, Iași, Editura Junimea.
- Petre Botezatu (1982) Semiotica, strategie optimă a construcției științifice, in Idem, *Interpretări logico-filosofice*, Iași, Editura Junimea, pp. 42–53.
- Mario Bunge (1976) Possibility and Probability, in Harper & Hooker (eds.), *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Sciences*, Vol. III, Boston, D. Reidel Publishing Company, pp. 17–33. Romanian translation, in Mario Bunge, *Știință și filosofie*, București, Editura Politică, 1984.
- Rudolf Carnap (1936) “Testability and Meaning”, in *Philosophy of Science*, Vol. 3 (1936) and Vol. 4 (1937).
- Rudolf Carnap (1955) Meaning and Synonymy in Natural Languages, in *Philosophical Studies*, No. 7, 1955, pp. 33–47, Romanian translation by Gheorghe Enescu and Sorin Vieru (1972), in Rudolf Carnap, *Semnificație și necesitate*, Editura Dacia, Cluj, 1972, pp. 298–313.
- Rudolf Carnap (1956) The Methodological Character of Theoretical Concepts, in H. Feigl and M. Scriven (ed.), *Minnesota Studies in the Philosophy of Science*, Vol. I, Minneapolis, 1956.
- Teodor Dima (1975) Tipuri de explicație științifică: explicația dispozițională și explicația genetică, in Mircea Flonta (coord.), *Epistemologia și analiza logică a limbajului științei*, București, Editura Politică, 1975, pp. 153–174.
- Teodor Dima (1980) *Explicație și înțelegere*, Vol. I, București, Editura Științifică și Enciclopedică, 1980; Vol. II, Iași, Editura Graphix, 1994.
- W. H. Dray (1957) *Laws and Explanation in History*, Oxford University Press.

³² Susan Haack, *Op. cit.*, p. 231.

Mircea Flonta (coord.) (1975) *Epistemologia și analiza logică a limbajului științei*, București, Editura Politică.

Susan Haack (2009) *Evidence and Inquiry. A Pragmatist Reconstruction of Epistemology*, Second Expanded Edition, Prometheus Books, New York, Romanian translation by Cătălina-Daniela Răducu, Introductory study by Acad. Teodor Dima.

C.G. Hempel (1942) The Function of General Laws in History, in *Journal of Philosophy*, 1942.

C. G. Hempel (1948), Studies in the Logic of Explanation, in *Philosophy of Science*, Vol. 15, 1948, pp. 135–175.

C. G. Hempel (1965) *Aspects of Scientific Explanation And Other Essays in the Philosophy of Science*, New York and London, The Free Press, Collier-Macmillan Ltd.

Immanuel Kant (1969) *Critica rațiunii pure*, translation by Nicolae Bagdasar and Elena Moisuc, București, Editura Științifică.

Karl R. Popper (1981), *Logica cercetării*, translation by Mircea Flonta, Alexandru Surdu and Erwin Tivig, București, Editura Științifică și Enciclopedică.

Price (1969) *Belief*, London: Allen and Unwin.

Wolfgang Stegmüller (1969) *Probleme und Resultate der Wissenschaftstheorie und Analytische Philosophie*, Band I Wissenschaftliche Erklärung and Begründung, Studienausgabe, Teil 1: Das ABC der modern Logik und Semantik. Der Begriff der Erklärung und seine Spielarten, Springer – Verlag Berlin. Heidelberg. New York, pp. 72–152.

G. H. von Wright, *Explicație și înțelegere*, translation by Mihai D. Vasile, București, Editura Humanitas.

THE ENDLESS TRUTH IN THE P. M. FORMAL SYSTEMS

In honor of the *Principia Mathematica* (1910–1913) of Bertrand Russell

Mihai D. Vasile

Romanian Academy of Sciences, Romania
e-mail: mdvasile@yahoo.com

Abstract:

Any logical analysis – and therefore philosophy itself – begins with the question: is there an indubitable knowledge? To answer this question Bertrand Russell appeals to tradition – both mathematical and philosophical – in which he recognizes himself. The process used by Russell, in order to build mathematics on a new foundation, was to build concepts logically, starting from atomic elements and known relationships. For example, a logical construction as that of the “class” is that a sentence, which includes a fixed concept, is turned into a logically equivalent statement (example analyzed by Russell is “Scott” and “the author of Waverley”) which unlike the first, entails relations that include concepts. Thus, the primitive concept can be replaced and removed. The new method was named the “theory of types” and became the intellectual program of Russell and his successors. Among the method’s performances is solving logical and mathematical paradoxes. The importance of Russell’s logical investigations exceeded the reductionist program, stimulating the formalist program of philosophy and methodology and the logical analysis of the language of science with a logical-mathematical apparatus. This program was definitively concluded, it seems, by Kurt Gödel’s results on the decidability problem in the formal axiomatic systems of P type, a problem closely linked to the question of “truth” and the two other meta-theoretical properties of P formal systems, namely “consistency” and “completeness”. Gödel’s Theorems do not affect the now established legitimacy of a meta-theoretical system concerning the formalizing processes, but prove that the hilbertian original program to reduce mathematics to logic is old and the very meaning of the whole process of realizing it must be analyzed again. Considering that in a formal P system, the decision process is demonstrability of true statements, Gödel’s research findings prove the substantial relationship of the four meta-theoretical properties Truth-Consistency-Completeness-Demonstrability.

Key words: PM system, truth, consistency, completeness, demonstrability, type theory, decidability.

The great rift between appearance and essence is exceeded by Immanuel Kant through mathematical cognition. After Kant, mathematics is an indispensable tool for the promotion of all sciences: a science that does not support the formal rigor of mathematics has not yet developed into a science. The pan-mathematization idea was reevaluated by Kant and has been theorized by scientists and philosophers in a manner not devoid of exaggerations concerning the ontological status of mathematical entities. Famous mathematicians – such as Bertrand Russell – have tried to show that modern mathematical theory developed in line with Kant, supporting and arguing for the privileged status of mathematics in the growth of knowledge.

René Thom enunciated this idea in accordance with the condition of the universality of science: for a science to be universal it must contain a translation capability such that it can be translated into all languages of the universe. This condition is fulfilled only by the system of mathematical formulas. Therefore any discipline aspiring to be science must become mathematized¹. There is something strange and frightening about math because it happens that physicists, when working on a new theory, discover that all the math they need already exists – as when Einstein found non-Euclidean geometry. This math is often developed by mathematicians who had no idea about the possibility of applying it to physics. The *phenomena* presumed or imagined by physicists are often already described through mathematics

¹ René Thom, *Towards a Revival of Natural Philosophy*, in: W. Güttinger and H. Eikemeier (eds.), *Structural Stability in Physics*, Berlin, Springer, 1979, pp. 5–11.

procedures. The *phenomena* of physics cannot be described in another language, but only in mathematical language, a language that is reliable and absolutely trustworthy in the sense that mathematical language expresses **Being** in the final analysis, *i.e.* the *noumenon*.

If we theoretically assume with Kant that mathematics always precedes philosophy in general, then the growth of scientific knowledge shows that progress in understanding the universe is due to improving mathematical tools.

The strangest, but also the most Kantian, question ever put to scientific knowledge as the “science of the ultimate goals of human reason” concerns its perfection: how far can mathematical tools be improved, if there is a limit to the perfectibility of these mathematical tools. The moral imperative fixed by Bertrand Russell for any philosopher, to use her professional competence only in the disinterested search for truth, makes the central point of any scientific knowledge on the **issue of truth**. The task of defining truth was assumed by contemporary semiotics, where the main results were obtained by Alfred Tarski who gave a formally correct and materially adequate definition of the concept of truth by (T) convention. It remains to be established how much and to what extent mathematics is able to grasp or to capture truth.

There are three trends in fundamental mathematics which have proposed this goal. After the logicist school represented by Bertrand Russell and Alfred Whitehead in England and the intuitionistic school founded by L. E. J. Brouwer in the Netherlands, the axiomatic formalist school, represented by David Hilbert and Wilhelm Friedrich Ackermann in Germany won the most spectacular and enlightening results in this respect.

After logicism, formalism emerged as a saving attempt to resolve the crisis caused by a number of paradoxes – mainly the liar paradox to which others are variations on a given theme – and by the challenge of L. E. J. Brouwer concerning the axiom about infinity. The new formalist program “was shaped by Hilbert, a breakthrough was made by his colleagues, Bernays, Ackermann, Von Neumann and others after 1920 ... he proposed a preliminary program which can be expressed as: classical mathematics must be formulated as an axiomatic theory, followed thereafter to demonstrate the consistency of the theory”².

² Stephen Cole Kleene, *A Critique of Mathematic Reasoning*, in: Stephen Cole Kleene, *Introduction to Meta-*

1. Formal systems of Principia Mathematica type. The P System

The development of mathematics towards greater precision led to the formalization of mathematics fields, in the sense that demonstrations can be given according to some mechanical rules. Some formal systems built so far are of **Principia Mathematica (PM)** type. This type of formal system³ is so broad that all methods of demonstration used in mathematics can be formalized in it, *i.e.* they can be reduced to a few axioms and rules of inference sufficient to decide mathematical problems that can be formally expressed in the given system.

Discovering formalisms power led many of the logicians to believe that the formal system cannot only eliminate the paradoxes of logicism, but also to solve the decision problem absolutely; namely formalism could determine the certainty of truth, being considered as a tool to separate absolutely true propositions in the system from false propositions. Kurt Gödel demonstrated, within a formal system, that formal systems are not an absolute and complete tool for decision. Formal systems that are imposed after efforts to resolve paradoxes have the merit that they take into account the problems and objections made by paradoxes. However, it was shown, in a procedure introduced by Gödel, that the method of formalized systems leads to a kind of paradox that cannot be avoided except with some sacrifice. Gödel directed his research towards the formal systems of **PM** type, showing that **PM** formal systems are incomplete, which means that **PM** types are limited as it concerns their power of decision. He demonstrated⁴ that in formal systems of **PM** type there are relatively simple problems of the ordinary theory of whole numbers that cannot be decided on the basis of axioms. This does not depend on the peculiar nature of the systems, but belongs to a very broad class of formal systems, among which are particularly included all those which are obtained from a **PM** system type by the addition of a finite number of axioms, assuming that no false proposition of the system can be demonstrated with the help of addi-

mathematics, Amsterdam, North-Holland Publisher Co., 1964, p. 36 sq.

³ Kurt Gödel, *On Formally, Undecidable Propositions of Principia Mathematica and Related Systems*, in: *From Frege to Gödel. A Source Book in Mathematical Logic*, ed. by Jean van Heijenoort, Cambridge, Massachusetts, Harvard University Press, 1967, p. 596.

⁴ *Ibid.*, p. 597.

tional axioms. If we build these systems as complete, they become necessarily contradictory, so that we have to choose between being paradoxical and complete or non-contradictory but incomplete.

Gödel reconstructed⁵ his **PM** system type under the following requirements:

(i) a **PM** system type contains a logical part and an arithmetical part;

(ii) in this system works a constructive process, *i.e.* the process of recursive functions, as an effective process of definition of decidable relations and their composition;

(iii) due to a special procedure, which is developed for some possibilities of the system, *i.e.* arithmetization procedure, the logic part can be represented in the arithmetical part; moreover, a class of meta-theoretical concepts, which can be formalized, can be represented in the arithmetical language. This fact is very important because the system closes again when the meta-theory can be represented in the theory language. Gödel thus rendered evident a possibility of the system, namely its being self-reporting, and he does nothing else than to update this possibility through the arithmetization method. The system thus reconstructed is called by Gödel the **P** system.

A very important issue in the **P** system is the issue of a **P** system language ranking: not only that a hierarchization is introduced in **P** by type theory⁶, but also different symbols are inserted at each level. The first and simplest level is the individual arithmetical level which belongs to simple numerical entities. Another is the level of meta-theoretical language, *i.e.* the meta-language. The arithmetization language is distinguished from the others, at least in the sense that it constitutes a distinct part in the arithmetical language due to the fact that:

(i) in this language, the digits from 0 to 9 are introduced, while in the arithmetical language there are only 0 and *successor* (hence the difference in the vocabulary);

(ii) the arithmetization rules are different from the arithmetical rules, the form originality of **P** system being precisely constituted in these arithmetization rules.

Thus, the notion of *formula*, for example, will be represented by an arithmetical formula. Then there is the logical language in which is given the aggregate of the composition rules, and finally the universal language in which we discuss everything and that is always assumed. The universal language position in relation to any inferior language is an important philosophical issue.

To present the system and Gödel's results, there are at our disposal two texts which belong to Gödel, namely *On Formally Undecidable Propositions of Principia Mathematica and Related Systems* and *On Undecidable Propositions of Formal Mathematical Systems*. The first is the classic text, rigorous, concise and self-sufficient; the second is the explanation from the first, but contains additions and further developments that we will indicate whenever appropriate. Sometimes I prefer the second text due to its clarity thorough detailed explanations. I also prefer the analysis of the original texts, although in any work about foundations theory of mathematics there are of Gödel's theorems, there are reformulations in the equivalent languages that are richer or poorer.

2. Description of P system

P is essentially the system obtained⁷ by adding Giuseppe Peano's axioms to the logic of the **PM** system (numbers as individuals, *successor* relationship as primitive concept). Note that the addition of Peano's axioms, like other changes made in the **PM** system, serve only to simplify the demonstrations, but technically are not indispensable. Specifically, the undefined⁸ terms in addition to the variables will be⁹:

I. constants: $\sim$ (not), $\vee$ (or), $\forall$ (universal quantifier), 0(zero), f (successor function), (.) (parentheses);

II. variables of type 1 (for the natural numbers):

$x_1, y_1, z_1 \dots$

⁵ Kurt Gödel, *On Undecidable Propositions of Formal Mathematical Systems*, in: *The Undecidable. Basic Papers on Undecidable Propositions, Unsolvability Problems and Computable Functions*, ed. by Martin Davis, New York, Hewlett, Ravel Press, 1965, pp. 61–63.

⁶ *Ibid.*, p. 46.

⁷ Kurt Gödel, *On Formally, Undecidable Propositions of Principia Mathematica and Related Systems*, in: *From Frege to Gödel. A Source Book in Mathematical Logic*, ed. by Jean van Heijenoort, Cambridge, Massachusetts, Harvard University Press, 1967, p. 599.

⁸ There are differences between the two texts of Gödel on this point; the second text does not make a very careful distinction between defined terms and abbreviations. Therefore we follow the first text, for the exposure of undefined terms and we will give the abbreviations according to the second specifying the text itself.

⁹ Kurt Gödel, *op. cit.*, p. 599 sq.

variables of type 2 (for classes of individuals):

$$x_2, y_2, z_2 \dots$$

variables of type 3 (for classes of classes of individuals):

$$x_3, y_3, z_3 \dots$$

and so on.

Abbreviations are: & (and), $\rightarrow$ (implies), $\equiv$ (identical with), E (existential quantifier) and $\P$ (concretization operator: $\P x (x_1 (x))$ means, “the smallest natural number x such that $x_1 (x)$ if $(Ex) x_1 (x)$, otherwise 0”).

By a sign of type 1, we will understand a combination of signs which has the form:

$$\mathbf{a}, \mathbf{fa}, \mathbf{ffa}, \mathbf{fffa}, \mathbf{a.s.o.},$$

where $\mathbf{a}$ is either 0 or a variable of type 1. A combination of signs which has the form:

$$\mathbf{a}(b)$$

where b is a n -type sign and $\mathbf{a}$ is a $n+1$ type sign, which will be called an elementary **formula**. **Class formulas** will be the smallest class that contains all the elementary formulas and formulas composed by elementary formulas and operators: $\sim$, $\vee$, $\otimes$. A formula with all tied variables will be a propositional formula. The **P** system contains some more axioms¹⁰ groups, namely:

I. arithmetic axioms group (Giuseppe Peano’s axioms, out of which Kurt Gödel holds three);

II. axiom schemes of propositional logic;

III. axiom schemes group of substitution;

IV. axiom group of comprehension;

V. axiom groups of extension, to which the following statement is available: it will be the axioms of all formulas obtained from raising the variables of the extension axiom with the same type.

The complete definition of the considered formal system calls for the list of the procedural rules to be given. Every rule of procedure should be interpreted as a statement of the conditions under which an **N** formula will be an immediate consequence of the meaningful formula **M** ($\mathbf{M}_1, \mathbf{M}_2, \dots$). Two of the rules of procedure in the **P** system are the detaching rule and the universalization rule. Note that the substitution rule does not appear in the rules of inference because it was transferred to the groups of axioms.

An **N** formula will be called immediate¹¹ consequence of $\mathbf{M}_1, \dots, \mathbf{M}_n$, if **N** is the immediate consequence of $\mathbf{M}_1, \dots, \mathbf{M}_n$ by one of the rules of inference. The **P** system must satisfy the requirement of constructiveness, that is, for each inference rule there will be a finite procedure to determine whether a given **N** formula is an immediate consequence (by that rule) of the given $\mathbf{M}_1, \dots, \mathbf{M}_n$ formulas, and it will be a finite procedure for determining whether a **N** formula is a meaningful formula or an axiom.

In a **P** system, Gödel demonstrates two meta-theorems: (i) there are undecidable sentences, *i.e.* there is (at least) one true proposition expressed in the system, but we cannot decide on it with the rules of the system; (ii) the assertion of non-contradiction of a **P** system can be formalized in the system, but it cannot be decided, despite the fact that the system is non-contradictory. In the very moment when we would try to conceive the system to be complete, we will get the contradiction.

3. Semantically undecidable propositions. Incompleteness of P

As we have seen, the **P** system presents an interesting complementarity between meta-theoretical properties of consistency and completeness in the sense that if **P** is assumed to be consistent, it is essentially incomplete, *i.e.* it contains at least one true sentence, but it is impossible to decide (checked, demonstrated) its truth by means of the **P** system.

¹⁰ As for specifying classes of axioms and rules of inference, the two texts differ essentially because, while the second introduces substitution rules in the rules of inference, we first specify them in a particular class of axioms, *i.e.* axioms of class substitution, so that, while the first text contains only two rules of inference, the second text contains six rules, four of which are substitution rules. I prefer the first text for its classicism.

¹¹ Kurt Gödel, *On Undecidable Propositions in: The Undecidable. Basic Papers on Undecidable Propositions, Unsolvability Problems and Computable Functions*, ed. by Martin Davis, New York, Hewlett, Ravel Press, 1965, p. 41.

Gödel shows that there are semantically undecidable sentences corresponding to undecidable propositions of the restricted functional calculus, in any system of **P** type, subject to the following conditions:

- (i) it includes a system of arithmetical sentences;
- (ii) any recursive predicate of the system is numerical;
- (iii) it establishes a correspondence between the **P** system and part of himself, namely restricted functional calculation (**r.f.c.**), so that if $x \ulcorner f(x) \urcorner$ belongs to the system, and if **b** is an expression belonging to the **r.f.c.**, and if $x \ulcorner f(x) \urcorner$ is derivable in the system, then **b** is realizable; and reciprocally.

A sentence is called semantically-undecidable if it is not possible to demonstrate either that it is universally valid, or that it is non-realizable.

In the formalism of **P**¹², a relationship (class) will be called arithmetical if it can be defined only by the help of concepts “+”, “•” (addition and multiplication of natural numbers) and logical concepts $\vee, \sim, \Pi, =$, where Π and $=$ refer to natural numbers. The concept of arithmetical sentences is properly defined.

Gödel established the fact that any recursive relationship is arithmetical (*Th VII*)¹³ and that for any sentence of the form $x \ulcorner f(x) \urcorner$ (where **f** is recursive), there is an equivalent arithmetical sentence, and if this equivalence is provable in **P**, it demonstrates that there are undecidable arithmetical sentences in each of the formal systems of **P** type (*Th VIII*)¹⁴, and that in all systems of **P** type there are undecidable sentences of **r.f.c.** (*Th IX*)¹⁵ (i.e. formulas of **r.f.c.** for which there is no universal validity or the existence of a counterexample is not provable). These sentences are called semantically-undecidable. Their existence is based on the fact that any sentence of the form $x \ulcorner f(x) \urcorner$ (**f** is recursive) can be reduced to the realizability problem of a formula belonging to **r.f.c.** (*Th X*)¹⁶ (i.e. for each recursive **f**, a **r.f.c.** formula can be produced whose realizability is equivalent to the truth of $x \ulcorner f(x) \urcorner$).

There are formulas built from primitive symbols that are considered as formulas of **r.f.c.**¹⁷: $\sim, \vee, \Pi, =$; individual variables $x, y, z, \dots$; variables for proper-

ties and relations $f(x), g(x, y), h(x, y, z), \dots$, where Π and $=$ refer only to the individuals. Added to these symbols is a third type of variable: $\ni(x), \ni(x, y), \ni(x, y, z), \dots$, which represents functions of objects (that is to say denotes functions whose arguments and values are individuals).

An **r.f.c.** formula which also contains a third type of variable will be called an extended sense formula (**e.s.f.**). The concepts of “realizable” and “universally valid” carry on **e.s.f.**, in the sense that for any **A** formula of **e.s.f.** can give a common **B** formula of **r.f.c.**, so that the realizability of **A** is equivalent to the realizability of **B**. As demonstrated in **P** such an equivalence, i.e. the equivalence between the realizability of an **e.s.f.** formula and the corresponding realizability out of **r.f.c.**, means that the undecidability of the first involves the undecidability of the second; from whence derives the incompleteness of a **P** system for the **r.f.c.** model.

From this result Gödel was able to obtain a very general theorem of incompleteness¹⁸. By this I mean that if we consider any formal **P** system based on a higher order language, whose rules of demonstration are quite simple, to be acceptable, then **P** is incomplete for any **C** class of models, whenever **C** contains at least one model whose field is infinite. In particular, a system cannot be complete for logically-valid sentences of a higher order language. The reason for the incompleteness of all formalisms, such as **P** type, lies in the fact that the formation of higher and higher types can be continued in transfinite¹⁹.

4. Mathematical formalism and philosophy

Gödel’s results on incompleteness raised two kinds of issues²⁰. The first kind refers to the fact that after it was discovered that certain formal-deductive systems are complete and others are not, it became necessary to seek some general criterion to identify the complete systems, or, where such a criterion does not work, to enter into individual examinations that would determine if a formal-deductive system is complete or not. Thus, instead of starting with a model and then searching for a complete formal system, we can start from a formal-deductive system of first order and ask if there is any model for which it is complete. The most interesting case of this kind,

¹² Kurt Gödel, *On Formally, Undecidable Propositions*, ..., p. 610.

¹³ *Ibid.*, p. 611.

¹⁴ *Ibid.*, p. 612.

¹⁵ *Ibid.*

¹⁶ *Ibid.*

¹⁷ *Ibid.*, pp. 612–613.

¹⁸ Leon Henkin, *Completeness*, in: *Philosophy of Science Today*, ed. by Sidney Morgenbesser, New York, Basic Books, 1976, p. 32.

¹⁹ Kurt Gödel, *op. cit.*, p. 610, nota 48.

²⁰ Leon Henkin, *op. cit.*, p. 32 sq.

found by Alfred Tarski, concerns the **R** system of real numbers. Tarski found a deductive theory **Td**, expressed in the language of first-order, adequate to **R** model, *i.e.* complete in connection with **R** model.

The second kind of problem arises from the fact that having a particular formal **J** system (of **P**-type), which is known to be complete for a class **C** of models, some sentences that are valid in **C** are derivable in **J**, while others are not. Therefore, general criteria are sought to distinguish these two types of sentences. Although Gödel's results on incompleteness raised doubts on this point, they have also served as a stimulus for further work in the mentioned directions.

New research into the general criteria that identify complete systems led to several new meta-mathematical concepts. Also looking into instruments for characterizing a class of derivable sentences in a complete formal-deductive system, this new research has discovered new mathematical structures and new ways of interpreting the language of mathematics. Thus, the general concept of the **algorithm**²¹ was analyzed in the years from 1930 to 1940. The impetus for this analysis came from the importance given to the problem of finding an algorithm for the decision procedure. In functional mathematical terminology, the problem was to find out which functions have algorithms. The analysis of the algorithm general concept has been undertaken by identifying the class of all functions which have algorithms, with a certain class of functions described mathematically. In fact, three different mathematical descriptions of a class of functions appeared at that time. These were the so-called λ -definable functions studied by Alonzo Church and S. C. Kleene in 1932 and 1933; the general recursive functions described by Gödel in 1934 which were built on a suggestion by J. Herbrand; and the computable functions, independently characterized by A. M. Turing and E. L. Post in 1936. These three classes, described in different ways, have been found to contain exactly the same functions.

The plausibility of this discovery, called the Church-Turing thesis, appears clearly within Turing's computable functions. A function of **p** and **q** is Turing-computable if there is a Turing machine which calculates its value, **r**, for each pair of values of **p** and **q**. Examples of such functions, given by Church, A. M. Turing and S. C. Kleene, opened a

new area of research because it has found a class of questions for which there is no algorithm. Thus, considering some formal-deductive **N** system, which is adequate to the elementary arithmetic of numbers and using Turing-computable functions, it is discovered that the decision problem for **N** system cannot be resolved in the affirmative. Likewise, similar results are obtained for the other theorems of Gödel.

Since the work of S. C. Kleene, Andrey Mostowski and E. L. Post, a whole hierarchy of degrees of uncomputability have been discovered. Alfred Tarski, Andrey Mostowski and others²² proved the undecidability of a variety of formal deductive theories based on first-order predicate logic, with different sets of mathematical axioms, such as those for the algebraic systems, reducing the decision problem of the considered system to the decision problem of a system for which it is known that its solution is negative.

"Philosophers tend to worry about whether formalization faithfully reflects the everyday discourse and the scientific one. Since mathematical logic lived a life of its own far too long and not quite loaded with events, professional logicians tend to overlook this problem, which, anyway, since it concerns in the first the mathematical discourse, is not very serious"²³. Inventing an ideal language would give a full and final resolution of these impediments on account of its wealth, clarity and accuracy, so that it would be self-enough as power of expression and would prevent the non-senses. The similarities between the ideal language and the automatic decision procedure of the answers to all logical and mathematical problems are that although both are impossible, their impossibility can be demonstrated provided they are given precise formulation. At the moment, the fact that a formal-deductive theory is inconsistent or not, or that it is not appropriate to formalize a demonstration of its own consistency, has led to a difference of opinion among the contemporary formalists and deepened their already existing differences. "Some believe that the consistency of mathematics cannot be determined solely on the speculative basis and mathematics has to be justified otherwise. Others argue that there are some forms of non-factual reasonings and, in a broader sense, that in their terms, Hilbert's program can be realized. Another group considers Gödel's proof as

²¹ S. C. Kleene, *Computability*, in: *Philosophy of Science Today*, ed. by Sidney Morgenbesser, New York, Basic Books, 1976, p. 37 sq.

²² Andrey Mostowski, *Thirty Years of Foundational Studies*, Oxford, Basil Blackwell, 1966, p.116.

²³ Hao Wang, *A Survey in Mathematical Logic*, Peking, Science Press, 1963, p. 60.

evidence that the whole formalist position, at least in its extreme form, cannot be sustained”²⁴. The recognition of the importance of the undecidability demonstration has, however, sparked philosophical controversies around mathematics, highlighting its really creative character²⁵.

Gödel’s results do not mean the failure of the formal method or its sterility, because of the fact that it circumscribes only the effectiveness field as well as the fact that research on logical rigor imposes narrow limits beyond which it loses touch with reality. On the contrary, his results make obvious the insufficiencies of natural reasoning, and in this respect, they remain the indispensable instrument for access to the truth. The self-restraint formalisms²⁶ are no barrier to close their creator in a speech vanity, but emerge from the view that mathematical existence becomes accessible to its researcher and gives a more accurate form to the rational field. We could say that Gödel’s results and those related to it exile the irrational in an unreal and vacuum field (and from a mathematical perspective, outside any system), in fact, the area of non-existence.

The two criteria for establishing the rational, non-contradiction and demonstrability (in disjunction, because their conjunction has been shown to be impossible for complex systems), exhaust the field of real and rational existence. The reasoning for this is simple: if a meaning system is non-contradictory (and therefore incomplete and undecidable) although incomplete and undecidable, it is rational because it is non-contradictory. If a system is complete and decidable (and therefore contradictory) although contradictory, it is rational because it is decidable. Since a system can only be non-contradictory or decidable in any case and ... and, in any case other than or ... or, and since any system, either non-contradictory, or decidable, is rational, it follows that any system is rational. So, the formal system becomes the subject and the formal mathematical thinking tool, the result of its continual genesis, opened to experience, indefinitely full, capturing the endless truth.

²⁴ Haskell Curry, *Foundation of Mathematical Logic*, New York, McGraw-Hill, 1963, p.11.

²⁵ Nachum Deshowitz and Yuri Gurevich, *A Natural Automatization of Computability and Proof of Church’s Thesis*, in: “The Bulletin of Symbolic Logic”, Volume 14, Number 3, Sept., 2008, pp. 338–340.

²⁶ Jean Ladrière, *Les limitations internes des formalismes*, Louvain-Paris, E. Nauwelaerts-Gauthier-Villars, 1957, pp. 403–405.

The existence of undecidable propositions shows that it is not possible to find, in the deductive field, the exact representation that belongs to the intuitive truth; there will always be more intuitive truth than may be expressed in the demonstrations. To deplete representation we are required to build increasingly larger systems, and as the process which allows the construction of undecidable propositions can always be applied to new systems, system enlargement may be continued indefinitely without ever being able to fully integrate the decidability²⁷. This fact is emphasized by meta-theoretical research using the models and methods in which the distance between the system and its model(s) is exactly equivalent to the necessary connection between formal and intuitive reasoning, and at the same time, brings a source of relativity. It is of course about the relativity of the system to its models.

Mathematical experience merely makes manifest the characteristics belonging to human experience in its condition, without access to pure being. Being is always hidden in phenomena, so that what is totally relevant is also dissembled. If the total system is non-realizable, it is because mathematical experience takes place amid an inexhaustible horizon.

“Bertrand Russell coined the phrase ‘philosophical logic’ to describe a program in philosophy: that of tackling philosophical problems by formalizing problematic sentences in what appeared to Russell to be the language of logic: the formal language of *Principia Mathematica*”²⁸. “Post-Fregean mathematical logic began with a concern for foundational issues in mathematics. However, by the 1930s philosophers had not only contributed to the building and refinement of various systems, but they had also begun an exploitation of them for primarily philosophical ends. While many schools of philosophy today eschew any kind of technical, logical work, an ability to use (or at list a familiarity with) the tools provided by formal logic systems is still taken as essential by most of those who consider themselves analytic philosophers. Moreover, recent years have witnessed a growing interest in formal logic among philosophers who stand on friendly terms with computer theory, cognitive psychology, game theory, linguistics, law, and so on. At the same time, techniques developed in formal logic continue to shed

²⁷ *Ibid.*, p. 411.

²⁸ Mark Sainsbury, *Logical Forms. An Introduction to Philosophical Logic*, Second revised edition, Oxford, Blackwell, 2001, pp. 1–2.

light on both traditional and contemporary issues in epistemology, metaphysics, philosophy of mind, philosophy of science, philosophy of language, and so forth”²⁹. Elucidating the nature of mathematics does not return to mathematics, but to philosophy; and it is possible only within the frame of some major conceptions that do not break mathematics from the rest of sciences, but on the contrary, take into account its origin from the natural sciences, its application and its links with other sciences, and finally, its history³⁰.

References

- Haskell Curry, *Foundation of Mathematical Logic*, New York, McGraw-Hill, 1963.
- Nachum Deshowitz and Yuri Gurevich, *A Natural Automatization of Computability and Proof of Church's Thesis*, in: “The Bulletin of Symbolic Logic”, Volume 14, Number 3, Sept., 2008.
- George Englebretsen and Charles Sayward, *Philosophical Logic. An Introduction to Advanced Topics*, New York, Continuum, 2011.
- Kurt Gödel, *On Undecidable Propositions of Formal Mathematical Systems*, in: *The Undecidable. Basic Papers on Undecidable Propositions, Unsolv-able Problems and Computable Functions*, ed. by Martin Davis, New York, Hewlett, Ravel Press, 1965.
- Kurt Gödel, *On Formally, Undecidable Propo-sitions of Principia Mathematica and Related Sys-tems*, in: *From Frege to Gödel. A Source Book in Mathematical Logic*, ed. by Jean van Heijenoort, Cambridge, Massachusetts, Harvard University Press, 1967.
- Leon Henkin, *Completeness*, in: *Philosophy of Science Today*, ed. by Sidney Morgenbesser, New York, Basic Books, 1976.
- Stephen Cole Kleene, *Introduction to Meta-mathematics*, Amsterdam, North-Holland Publisher Co., 1964.
- S. C. Kleene, *Computability*, in: *Philosophy of Science Today*, ed. by Sidney Morgenbesser, New York, Basic Books, 1976.
- Jean Ladrière, *Les limitations internes des formalismes*, Louvain-Paris, E. Nauwelaerts-Gauthier-Villars, 1957.
- Andrey Mostowski, *Thirty Years of Founda-tional Studies*, Oxford, Basil Blackwell, 1966.
- Mark Sainsbury, *Logical Forms. An Introduc-tion to Philosophical Logic*, Second revised edition, Oxford, Blackwell, 2001.
- René Thom, *Towards a Revival of Natural Phi-losophy*, in: W. Güttinger and H. Eikemeier (eds.), *Structural Stability in Physics*, Berlin, Springer, 1979.
- Hao Wang, *A Survey in Mathematical Logic*, Peking, Science Press, 1963.

²⁹ George Englebretsen and Charles Sayward, *Philosophi-cal Logic. An Introduction to Advanced Topics*, New York, Continuum, 2011, pp. 1–2.

³⁰ Andrey Mostowski, *ibid.*

COMMUNICATION AND ARTISTIC EXPRESSION

Marin Aiftincă

Romanian Academy, Bucharest
e-mail: aiftinca@acad.ro

Abstract:

On the assumption that, among forms of human communication, artistic communication is one of the most rich and complex, this paper mainly analyses two major aspects. The first refers to the phenomenon of contemporary art characterized by its reflective, but also commercial, nature, by its questioning of the essence, the forms of expression and the purpose of art, which seeks its identity with respect to the requirements of our time. The second brings into focus another major issue: the current strained relations between contemporary artistic creation and the public, each with its own development and justifications that are not easy to reconcile. Finally, this study advocates the idea that, beyond any indulgence to "anti-aesthetic", post-historical artistic creation, the public claims that art should offer aesthetic satisfaction, for she seeks beauty in artistic creation, to the delight of contemplation and spiritual perfection.

Key words: communication, art work, public, aesthetic value, symbol, language, artistic expression.

It has become a commonplace idea that art is a way of perceiving and rethinking the world, but also, we must add, a way of expressing it. We always relate to the outside world of the reality of things and phenomena, and also to the inner world of feelings and thoughts that inhabit our soul. The perception of these realities requires expression, otherwise they may sink into nothingness. A wide range of possibilities, always extended by means of our inventive spirit, is at reach for the embodiment of one's perceptions and experiences. In the case of art, we express them by words, gestures, sounds, lines, colors, etc., that constitute the configuration of the objects of artistic creation – objects to which we ascribe meanings and which we propose to others, thus establishing communication relations. It is generally accepted that, among forms of human communication, artistic communication is one of the most rich and complex. This feature comes from the

very nature of the objects of artistic creation, which in their sensible appearances, comprise a plurality of meanings, limited only by the hermeneutic ability of the contemplator.

Fitting here is a general remark of C. Noica, relating two different notions: communication and communion. If the first term involves information, signals, significances and meanings, the second refers to the subtlety of meanings. Being intimately linked to our life and culture, it can be said that "we progress through an increase of communication, but we are not really progressing if we do not get more communion" (1). Paraphrasing the Romanian philosopher's subtle statement, in connection with this issue, the following reformulation seems adequate: artistic communication has progressed more with respect to techniques and forms of expression, but genuine progress implies an increase of communion. In other words, the evolution of art also requires, whether we accept it or not, the development of receptor's interpretative capacity of grasping the subtle nuances of the art work, beyond its appearance. Of course, this idea refers to the widely argued issue of the relationship between art and the public and, more specifically, to the issue of the art work's accessibility, a theme to which we will return later. For now, we note just one key aspect to our approach: artistic communication, like any other type of communication, occurs, as it is well known, through a specific language, introducing a content of feeling and thinking within the artistic field. Naturally, we talk only in a synthetic manner about an artistic language in general. In fact, each particular field of artistic creation is defined by its own language that enables its existence and assures its identity (2).

Without going into the details of this issue, I must add that an essential component of this type of

language is the artistic expression. Understood as the synthetic unity of thought and feelings (3), the artistic expression is specific to each art field. Therefore, we can talk about literary, or artistic, or musical, or theatrical, etc. expression. Besides this artistic expression, it is impossible for us to understand *The Divine Comedy*, *The Brandenburg Concertos*, *The Romanian Rhapsody*, *The Swan Lake*, *Gioconda* or any other work that ennobles the kingdom of arts.

As it is well known, in any sphere of art, only by appropriate means, namely by its specific language, the artistic expression acquires form and becomes an objectified creation with a plurality of potential meanings. This is just the art work itself, presented as a unitary structured whole, from which nothing can be removed or changed without damaging its value. The artistic expression aspires to reflect in its specific manner the essence of reality; its cognitive elements are not shaped in the logical-scientific reflection of reality, but in a spiritual sensible way.

The structure of the artistic expression, in agreement with the specificity of the different fields of creation, has as essential qualities meaning, emotional character, originality and accessibility (4). Each of these qualities entails separate theoretical developments, which exceed the intentions pursued here. We note, however, that in general the complexity of artistic expression enables the art work to transcend the mere sign, rising up to the level of symbol – that is, a synthesis of sensible and idea (5), of concrete and abstract.

The objectification of the artistic expression in a work of symbolic value constitutes, in fact, a “natural” reaction against imitative art, enslaved to the world of appearances, and also a reorientation, active since the beginning of the last century, towards abstraction and symbol. That is, the intentional cultivation of a type of creation that refuses to be a mere image duplication of external reality. As one theorist noted, “an artist does not want to create perceived images, but represented images; he tends to institute symbols and to create configurations” (6). According to this aesthetic theory, if imitative art presents only a fragment cut out of this world, which it sends subjectively to the viewer, symbolic art provides a spiritual whole, “where objects do not move in spaces imitating” those offered in experience, “but in systems of spiritual, imaginary dimensions, where the data provided by our senses are no longer valid” (7).

Hence, the idea that art is neither an imitation nor a limitation of nature, but her ideal interpretation. In this case, regarding the art work as a symbol, there appears the difficult problem of the relationship between signifier and signified, between art object and its contemplator, which constitutes the source of the plurality of different approaches and aesthetic controversies. We note, however, the argument that “significance is (...) a creation of sign, but one that allows us to go back to reality, to extract its deep and unlimited meanings. The sensible reality does not lack ideality, nor the significance” (8) provided by the artist’s generalizations.

Unlike conceptual symbols, such as scientific, religious, or historical ones, artistic symbols are sensible and, as such, versatile, and their significance is, as T. Vianu argued, *illimitable*. It is appropriate that spirit finds meaning within the symbol, not outside it; and, in order to recognize the artistic character of the symbol, this meaning must have a way of being open, illimitable.

The multiplicity of meanings sought by contemporary art, whether symbolic or not, poses serious communication difficulties to a public accustomed to a different artistic tradition. With extension to the so-called anti-art, symbolic art is versatile as content and complex as expression. In other words, it allows and even calls for correlative representations instead of merely expressing the effects of perception. The access to its inner structure is not an easy exercise, because of the symbol ambiguity and its significant connotations. If one thinks of Brancusi's works in Tg. Jiu (*The Table of Silence*, *The Gate of Kiss* and *The Endless Column*), one can see that, taken together or separately, these art works “incite to almost endless reveries and interpretations” (9). This is an obvious consequence of the versatility of the aesthetic symbol created by Brancusi.

According to a relevant comment, symbolic art “has the character of the picture-writing, which must be read and interpreted” (10). Therefore, in the context of artistic creation becoming more reflective, especially with postmodern or “post-historical” art, free of any external finality, art opens to different interpretations and significances – which raises a rather difficult task for the public to accomplish. It is, among other things, a matter that led to the opinion, also supported by current artistic experience, that, in order to facilitate communication with the contemplator, art work requires an accompanying comment made by the author himself

or by a specialist in the field. This requirement covers the entire field of contemporary art, mainly the one sprung out of anti-aesthetic impulses.

If we put the question this way, at least two important interrogations are occurring: 1. Does the interpretation offer to the public a full “translation” of the artistic work?; 2. Does the comment enable or obscure the public access to the work of art? The answers to these questions lead aesthetics to complex approaches that go beyond this paper’s aims. Therefore, we limit ourselves to say, on the first question, that the object as unique artistic expression has a number of specific qualities, among which is a prevailing of the aesthetic value that gives its identity, thus transcending the mere thing character. For, besides what an ordinary thing can tell us, the artistic object communicates in more profound and exciting ways. The unalterable unity of the coexistence of material support (word, color, sound, marble etc.) and spiritual content, in the case of the art work, makes it a symbol that radiates a cluster of meanings and significances, catching the public’s attention (10). Of course, a comment or interpretation may lead the public to note or understand the emergence of the spiritual content of an artistic creation, which is undoubtedly a good thing, especially when it comes to any rebel form of art. But they are not able to fully “translate” the profound significances of an art work, because those cannot be entirely disclosed. We have noticed both the acknowledgment of these difficulties and the need for transcending them in an experiment unique in the world, initiated by the museum of the Benedictine monastery of Admont, in a small Austrian settlement. In 2002, several contemporary artists started a collection of works especially designed for being valued not with the eyes, but with the “mind’s eyes”. Gathered after a decade in an exhibition entitled “Beyond what you see” these works, paintings, installations, multimedia works, with particular forms, offer to the public new possibilities of complex evaluations from entirely different perspectives. The exhibition’s guiding idea is that art can be valued in various ways, so visitors can get an “impression” of how the world looks if it is not seen (11). Therefore, it proposes that art should not be judged solely depending on its aesthetic value.

Returning to our issue, the truth is that the universe of every art work contains a mystery, which refers to a content only suggested by its creator’s intuition, and never shaped as an image. Hence, art

work can never be an object of absolute disclosure. It remains always free, open and inviting in its relations with the public (12). Precisely its intrinsic mystery explains the sense of inner joy and the fascination that every human being experiences in contemplating an authentic art work, feeling it close, but never possessing it.

Regarding the second interrogation formulated above, we note that real communication takes place through artistic expression objectified as art work. Undoubtedly, it requires from the public adequate aesthetic culture, and flexibility. Such a condition appears to be insufficient when we face an art work which tells us nothing or confuses us. In this case, a comment or an interpretation of the art work facilitates decoding the author’s intentionality and may also transgress the immediacy of feeling and of concrete reality. Based on the aesthetic object, interpretation itself is a new spiritual creation and, therefore, may distort the act of personal reception, trying to say something else than the art work itself does. In this situation, the artistic object is eclipsed by the interpretation, which gives priority to the ideas that the author intended to convey to the public. We should remember, however, that what is important is not the referent, what an art work symbolizes, but the art work itself as a whole. The art work alone constitutes the object of aesthetic evaluation and, therefore, can be considered with respect to aesthetic values.

As shown above, the theme of communication through artistic expression objectified in art works, approached particularly from a contemporary perspective, is accompanied by a number of challenging and even corrosive difficulties aiming at both artist and contemplator. Such a difficulty consists in the very phenomenon of art. Marked both by its reflective and commercial nature, by what is called minimalist art, practical art, functional art, conceptual art, etc., contemporary art is beset with various questions and also pulled into conflicting interpretational thickets. Subject to questioning aiming at its essence, its forms of expression and its goal, contemporary art proves that it seeks its identity with respect to the requirements of our time.

Another major issue is that the strained relations between artistic creation and the public, each with its own development and justification, are not simple, nor easy to reconcile. On the one hand, the tendency to release artistic creation from its exclusiveness and to expand its commitment over extra-artistic areas (13), on the other hand, the public – although

strongly urged to change its attitude towards art work and to initiate a dialogue without prejudice – beyond any indulgence to “anti-aesthetic”, post-historical creation, claims that art should offer, however, aesthetic satisfaction, for he seeks beauty in the structure of art works, to the delight of contemplation and to spiritual perfection.

References and notes

Constantin Noica (1978), *Cuvânt împreună despre rostirea românească (Together Word about the Romanian Utterance)*, Bucharest, Publishing House Eminescu, p. 188.

Marin Aiftincă (2011), *Perspective filosofice (Philosophical Perspectives)*, Bucharest, Romanian Academy Publishing House, pp. 74–75.

See: *Estetica* (1983) (*The Aesthetics*), Bucharest, Romanian Academy Publishing House, p. 2013.

Dicționar de estetică generală (1972), (*General Aesthetic Dictionary*), Bucharest, Political Publishing House, p. 120.

Tudor Vianu (1966), *Simbolul artistic (The Artistic Symbol)* in vol. “Postume” (Posthumous), Bucharest, Publishing House for World Literature, p. 149.

Werner Hofmann (1977), *Fundamentele artei moderne (Grundlagen der Modernen Kunst)*, Bucharest, Publishing House Meridiane, pp. 100–101.

Ibid., p. 101.

Tudor Vianu, *op.cit.*, p.149.

Werner Hofmann, *op.cit.*, p.106.

See: Marin Aiftincă (2007), *Misterul artei și experiența estetică (Mystery of Art and Aesthetic Experience)*, Bucharest, Romanian Academy Publishing House, pp. 10-11.

See: Die Sammlung *JENSEITS DES SEHENS, Kunst Verbindet Blinde und Sehende, Sammlungskatalog anlässlich ihres 10. Jahrgigen Bestehens und der Ausstellung .25. März bis 26. August 2012 im Benediktinerstift Admont-Bibliothek Museum, A-8911 Admont*.

Marin Aiftincă, *op.cit.*, pp. 12–17: for more details on this subject.

See: Werner Hofmann, *op.cit.*, p. 49.

SELF-DECEPTION AND THE SELECTIVITY PROBLEM

Marko Jurjako

Faculty of arts and sciences at the University of Rijeka
Marohnićeva 4, 51000, Rijeka
Mob. 385 98 1866089
mjurjako@ffri.hr

Abstract:

In this article I discuss and evaluate the selectivity problem as a problem put forward by Bermúdez (1997, 2000) against anti-intentionalist accounts of self-deception. I argue that the selectivity problem can be raised even against intentionalist accounts, which reveals the too demanding constraint that the problem puts on the adequacy of a psychological explanation of action. Finally I try to accommodate the intuitions that support the cogency of the selectivity problem using the resources from the framework provided by an anti-intentionalist account of self-deception.

Key words: action explanation, folk-psychology, dispositions, epistemic virtue, Mele's deflationary account, self-deception, the selectivity problem.

1. Introduction

The phenomenon of self-deception has proven to be a very controversial philosophical subject whose paradigmatic cases are hard to characterize in a general manner. However, one can find a fairly neutral description of the phenomenon of self-deception that encompasses varieties of views on the Stanford Encyclopedia of Philosophy: "self-deception is the acquisition and maintenance of a belief (or, at least, the avowal of that belief) in the face of strong evidence to the contrary motivated by desires or emotions favoring the acquisition and retention of that belief" (Deweese-Boyd 2006).

Standard examples of self-deception that can be found in philosophical literature usually involve a person that is under the grip of emotion forming a self-deceptive belief about some state of affairs. For example, a mother is present at the court where her son is being tried for murder. She listens to all the evidence that the prosecutor is presenting and despite the persuasiveness of the evidence, which would be more than enough for an unbiased ob-

server to convict her son, she stays convinced that her son is innocent. Intuitively, in this case we say that the mother has a self-deceived belief that her son is innocent because she is not responsive to the available evidence to the contrary and because the formation of her (self-deceptive) belief is caused in an irrational manner; it was caused by her attachment to the son and not by the available evidence.

1.1 Intentionalist and anti-intentionalist accounts of self-deception

There are two general groups of accounts that try to explain the phenomenon of self-deception. On the one hand, *intentionalists* think that self-deception essentially involves an intention on the part of the self-deceiver who actively (intentionally) deceives herself (e.g. Bermúdez 2000, Davidson 2004). On the other hand, *anti-intentionalists* think that self-deception can be explained without invoking the concept of intention in forming the self-deceived belief (see e.g. Johnston 1988; Lazar 1997, 1999; Mele 1998, 2001).

To illustrate differences between the two accounts let us consider Davidson's (2004) example:

Carlos has a good reason to believe he will not pass the test for a driving license. He has failed the test twice before and his instructor has said discouraging things to him. On the other hand, he knows the examiner personally, and he has faith in his own charm. He is aware that the totality of the evidence points to failure. [...] But the thought of failing the test once again is painful to Carlos (in fact the thought of failing at anything is particularly galling to Carlos). So he has a perfectly natural motive for believing he will not fail the test, that is, he has a motive for making it the case that he is a person who believes he will (probably) pass the test. (p. 209)

According to Davidson we should interpret Carlos' case as a case in which Carlos goes through a

process of practical reasoning that ends up with his forming an intention to form a belief that he will pass the test. Because of the uneasiness that Carlos would feel if he failed his test he forms a desire to avoid feeling this uneasiness. Together with the belief that this desire can be satisfied by forming a belief that he will not fail the test, Carlos intentionally leads himself to believing that he will not fail the test. This pattern of practical reasoning exhibited by Carlos is parallel to the reasoning that a person undergoes when intentionally doing something. For example, a construction worker wants to make concrete: she knows that in order to make concrete she must mix water, cement, and sand; therefore, she forms an *intention* to mix water, cement, and sand, and consequently she executes her intention into an (intentional) action.

However, intentionalist accounts traditionally face at least two problems, the so-called static and dynamic paradoxes (see Bermúdez 2000; Mele 2001).¹ These *prima facie* problems for intentionalist accounts gave a boost to anti-intentionalist proposals for explaining self-deception (cf. Bermúdez 2000, 309; Mele 2001; see also Bermúdez 1997). According to anti-intentionalists we do not have to postulate the presence of an intention in the mind of the self-deceiver in order to explain the phenomenon of self-deception. Anti-intentionalists claim that putative unintentional mechanisms of the mind are enough to explain garden-variety cases of self-deception. In this respect, the main strategy of the anti-intentionalists is to explain self-deception as a case of motivationally biased belief.

For example, anti-intentionalists would claim that Carlos does not intend to deceive himself; rather, they would claim that he has a strong desire not to fail the test, since that would cause him dire pain. Consequently, the desire biases his belief-forming process and causes him to believe that he is not going to fail the test. In the next subsection I will present what is currently the most influential anti-intentionalist account of self-deception and then in the following section I will present and discuss the objection to this account proposed by Bermúdez (1997, 2000)

1.2 Mele's anti-intentionalist account

According to the most influential contemporary anti-intentionalist, Alfred Mele (1997, 2001), there are four jointly sufficient, though not necessary, conditions that explain the acquirement of a belief through a process of self-deception:

1. The belief that *p* which *S* acquires is false.
2. *S* treats data relevant, or at least seemingly relevant, to the truth-value of *p* in a motivationally biased way.
3. This biased treatment is a non-deviant cause of *S*'s acquiring the belief that *p*.
4. The body of data possessed by *S* at the time provides greater warrant for $\sim p$ than for *p*. (50-51)

The first condition is intuitive; Mele claims that "A person is, by definition, deceived in believing that *p* only if *p* is false" (ibid., 51), and that the same holds for self-deception.²

The second condition concerns the mechanisms through which people acquire self-deceived beliefs. Mele (ibid., 26-27) names four possible mechanisms. Negative misinterpretation is a phenomenon that happens when a person's desire that *p* causes her to misinterpret that some evidence does not count against *p* when the same evidence would be considered to count against *p*, if it were not for the desire that *p*. In our first example, it could be the case that the mother believes that the murder weapon with her son's fingerprints does not count as evidence for her son's being guilty, even though she would believe that it counts as evidence were it not for the desire that her son be innocent.

Positive misinterpretation is manifested when a person misinterprets some evidence as favoring her desired proposition when that evidence, in absence of the biasing desire, would easily be recognized as counting against the desired proposition. Mele (ibid., 27) gives an instructive example: Sid is in love with Roz, but Roz does not love Sid. However, Sid misinterprets her refusal of his affection as evidence that she really loves him (and pretends that she is 'hard to get'), and that she wants him to prove his real love for her.

Another psychological process is the selective focusing on evidence. The process in question is such that the mother's desire that her son be innocent can lead her to focus her attention on evidence that counts in favor of her son being innocent and that can lead her to fail to recognize the available

¹ To indicate briefly, the static paradox concerns the fact that intentionalists are seemingly committed to a claim that a self-deceiver has two contradictory beliefs (belief that *p* causes the belief that not *p* [see Davidson 2004, 208]). And the dynamic paradox reflects the seeming problem according to which the intentionalist needs to ascribe a self-defeating intention to a self-deceiver (i.e. the intention to deceive herself).

² However, not all authors would agree with this claim. For example Bermúdez (2000, 310–311, 312) argues that even in interpersonal deception the deceiver does not have to hold that the deceived belief has to be false, and by analogy the same should work for self-deception.

evidence that goes against the truth of her desired proposition.

The fourth mechanism that Mele presents is selective evidence-gathering. To illustrate this mechanism we can suppose that the mother is trying to guide her own investigation about her son's case. However, while investigating, the mother's desire that her son be innocent biases her into overlooking the evidence that goes against her desired proposition and leads her to find some less obvious and less convincing evidence that her son is innocent.

The third condition is added by Mele to secure the exclusion of cases of distorted belief-forming processes that are not cases of self-deception. For example, a mother may want to believe that her son is innocent and that desire can bias her search for evidence in favor of her desired belief. However, while trying to investigate the crime scene (or some scene more available to her) in order to find some evidence in her son's favor, she slips on the wet floor and falls on her head. After she wakes up, because of the trauma, she starts to believe that her son is innocent. In this case we can say that her desire to believe that her son is innocent caused her to believe that he actually is innocent. However, the causal chain from the desire that *p* to the belief that *p* is causally deviant in a way which prevents us from saying that this is a case of self-deception.

The fourth condition is meant to capture the often present characteristic of self-deception; that is, the failure of epistemic rationality. An epistemically rational person is someone who forms her beliefs in accordance with the best available evidence. However, a self-deceiver is a person who, despite the availability of good evidence to believe $\sim p$, starts to believe that *p* because she has a biasing desire that *p* be true. So, a self-deceiver is epistemically irrational because she forms her beliefs on the basis of her desires and not on the basis of the available evidence. However, unlike Davidson (2004), Mele does not take this condition to be necessary for entering into the process of self-deception.³

³ In a recent article Mele (2012) introduces two amendments to his jointly sufficient conditions for acquiring self-deceptive belief. These are: condition 5: "S consciously believes at the time that there is a significant chance that $\sim p$ "; and condition 6: "S's acquiring the belief that *p* is a product of 'reflective, critical reasoning,' and S is wrong in regarding that reasoning as properly directed" (12). Condition 5 is meant to accommodate the intuition that there is some cognitive tension (conflict) present in cases of self-deception. While condition 6 captures the intuition that self-deception requires mistaken higher-order thought about the causes of one's own cognitive activities (see Mele 2012, 10–12). I am mentioning these

In order to make his account more empirically adequate, Mele (1997, 2001) relies on an empirical model of everyday hypothesis testing, developed by the psychologists Friedrich (1993), and Trope & Liberman (1996). The idea behind the FTL model⁴ is succinctly summarized by Bermúdez (see also Mele 2012):

People have different acceptance/rejection thresholds for hypotheses depending upon the expected subjective cost to the individual of false acceptance or false rejection relative to the resources required for acquiring and processing information. The higher the expected subjective cost of false acceptance the higher the threshold for acceptance; similarly for rejection. Hypotheses which have a high acceptance threshold will be more rigorously tested and evaluated than those which have a low acceptance threshold (Bermúdez 2000, 316).

In garden-variety cases of self-deception Mele proposes that the structure of everyday hypothesis testing will enable one to enter into a process of self-deception. This idea can be illustrated with an example. A mother who desires that her son is innocent would experience more painfully the thought that her son is guilty than that he is innocent. These emotions, and the desire that her son is innocent, structure her expected subjective cost of believing a hypothesis in a way that makes it easier for her to falsely believe that her son is innocent, than to falsely believe that her son is not innocent. Because of this subjective cost, her acceptance threshold for believing that her son is innocent will be much lower than the acceptance threshold for believing that her son is guilty. According to Mele, the desire that the son be innocent and the low threshold associated with the desired state of affairs is what explains the mother's acquiring the self-deceived belief that her son is innocent.

2. Anti-intentionalism and the selectivity problem

Bermúdez (1997, 2000) has put forward a problem that threatens to undermine anti-intentionalist accounts in favor of intentionalist models. In this section I will present his objection. Here is how Bermúdez formulates the problem:

Any explanation of a given instance of self-deception will need to explain why motivational bias

two amendments in a footnote because they do not play any significant role in what follows in the present article.

⁴ FTL stands for Friedrich, Trope and Liberman, psychologists who developed the model of everyday hypothesis testing.

occurred in *that* particular situation. But the desire that *p* should be the case is insufficient to motivate cognitive bias in favour of the belief that *p*. There are all sorts of situations in which, however strongly we desire it to be the case that *p*, we are not in any way biased in favour of the belief that *p*. How are we to distinguish these from situations in which our desires result in motivational bias? I will call this the selectivity problem (Bermúdez 2000, 317).

According to Bermúdez, the conditions enumerated by Mele for having a case of self-deception are insufficient because they cannot explain why the desire that *p* be the case in one situation will cause us to believe that *p* is the case, while in another situation it will not. Bermúdez (1997, 108) claims that Mele illegitimately draws “analogy between familiar examples of unintentional cold bias and motivationally hot bias.” The difference between instances of cold bias and self-deception is that cold bias is *nonselective*, which means that regardless of its subject matter it gets activated, while the process of self-deceptive belief formation is subject dependent (ibid.).

Mele (2001, 63-65) tries to answer the selectivity problem by using the FTL model for hypothesis testing in order to explain what makes the difference between two situations in which a desire with the same content has different effects on the formation of the belief. He gives an example of Gordon who is a CIA agent that has been accused of treason. Gordon’s parents and all of his colleagues want him to be innocent and they all have access to the same information regarding Gordon’s case. However, based on the available evidence, Gordon’s parents come to the conclusion that he is innocent, while his colleagues, on the basis of the same evidence, conclude that he is guilty.

According to Mele, what explains that the parents and Gordon’s colleagues come to believe different things, despite having the same desire and access to the same evidence, is the fact that their acceptance/rejection thresholds for the hypothesis that Gordon is innocent are different. For Gordon’s parents it is less costly to falsely believe that he is innocent, than to falsely believe that he is guilty. While for his colleagues it is the other way around; since their lives depend on the verdict, it is much costlier for them to falsely believe that Gordon is innocent than to falsely believe that he is guilty. Accordingly, the difference in associated costs causes parents to have a low acceptance threshold for the hypothesis that their son is innocent, while his colleagues have high thresholds for the acceptance of the same hypothesis.

However, Bermúdez (2000, 317–318) asserts that the selectivity problem reappears again even if we introduce the cost/benefit analysis as Mele does. Bermúdez’s claim is that “there are many hypotheses for which my motivational set dictates a low acceptance and high rejection threshold and for which the evidence available to me is marginal enough to make self-deception possible. But I self-deceptively come to believe only a small proportion of them. Why those and not others?” (2000, 318).

Bermúdez’s answer is that what is missing and what solves the selectivity problem is the intention “on the part of the self-deceiver to bring it about that he acquires the belief that *p*” (ibid.). This is why, according to Bermúdez, intentionalist accounts (in some form) must be right, since they have a ready answer to the selectivity problem that anti-intentionalists supposedly do not have.

In the next section I will try to show that, if genuine, the selectivity problem can also be raised against intentionalist accounts. After that I will examine one possible solution to the selectivity problem proposed in (Pedrini 2010).

3. The selectivity problem for intentionalists

If the selectivity problem is genuine, then introducing intentions into the account will not solve the problem. As Mele (2001, 65-66) already noticed, even if intentions were necessary for accounting for an episode of self-deception, intentionalists will have a selectivity problem of their own. Let us suppose that what explains the difference between situations in which agent *A* with a strong desire that *p* self-deceptively believes that *p* and situations in which she does not, is the fact that in one situation she had an intention to bring it about that she believes *p* and in another she had no such intention. Nevertheless, it is clear that there are many cases in which we intend to do something and fail to accomplish it. I can intend to become a basketball player, and fail to become one; also, I can intend to raise a glass that is on the table and fail to do that, just like I can intend to hit the bulls-eye and not manage to do that. So, having an intention to bring oneself to believe something does not seem to be sufficient for entering into a process of self-deception since nothing in the nature of intention guarantees that the intention will be effective. Hence, the question of selectivity of self-deception would arise for intentionalists as well; they seem to be faced with the question of what is the difference between situations in which agent *A* has the intention to be self-deceived and manages to self-deceive herself and

situations in which she does not succeed in executing that intention.

Moreover, a stronger point can be made against the intentionalist who thinks that the selectivity problem can be resolved by introducing the relevant intention. In his (1997, 108; 2000 317) Bermúdez put the following desideratum on the proper explanation of self-deception: “Any explanation of a given instance of self-deception will need to explain why motivational bias occurred in *that* particular situation.” If we generalize Bermúdez’s proposal, we can say that at the most general level, the selectivity problem boils down to the question: why some people deceive themselves and others do not?⁵ From this standpoint, even if, for the sake of argument, we grant that intentionalist accounts are true, we still would not have a proper explanation why in a particular case an episode of self-deception occurred. Namely, Bermúdez supposes that introducing an intention to acquire a belief will explain why in *this* particular situation an episode of self-deception occurred. However, an intention by itself cannot explain an occurrence of self-deception. To rephrase Bermúdez (2000): Any explanation of a given instance of self-deception will need to explain why *intention* was formed in that particular situation.

To see why let us consider a more general example of an instance of action explanation. In folk-psychological explanations of action we usually explain why someone did something by invoking the reasons for which that person acted. For example, we say that Jennifer went to a store because she wanted some candy, and that she believed that by going to the store she could buy some candy. Since we suppose that Jennifer’s action was intentional we ascribe to Jennifer an intention to go to the store. However, ascribing an intention to Jennifer does not explain *why* she went to the store; the reasons (i.e. desires and beliefs) explain why she formed that intention and performed that action. A similar point can be made in the case of self-deception. Introducing an intention to self-deceive will not explain why in this particular situation an instance of self-deception occurred. We need to know the reasons for which the intention is formed, and, in the folk-psychological sense, this will include the desire and a means-end belief.⁶

⁵ I am very grateful to Patrizia Pedrini for giving me an idea how the following case can be used against the intentionalist’s solution of the selectivity problem.

⁶ See Bermúdez (1997, 108) where he writes: “the intentionalist holds that the subject must intend to cause himself to believe that *p* by biasing his cognitive processes because (a) he desires to believe that *p* and (b)

So, in order to explain why in this particular instance self-deception occurred, we need to invoke a desire and a belief. But, now, we can ask why in this particular situation a desire that *p* be the case caused an intention to believe that *p* is the case? As Bermúdez noted, we have all kinds of desires that, nevertheless how strong, do not cause us to believe that *p* is the case; similarly we can say that we have different strong desires to believe that *p* be the case (or that we believe that *p* is the case), that nevertheless do not cause an intention to believe that *p*. So, in this way we can raise the selectivity problem against the intentionalist account. Namely, we can raise the question why in this *particular* situation desire that *p* be the case (or to believe that *p*) caused an intention to believe that *p* is the case since, according to Bermúdez, in all kinds of situations, no matter how strongly we desire that *p* be the case it does not cause us to believe that *p* is the case.

Someone could argue on behalf of Bermúdez that the intentionalist account of self-deception explains the occurrence of an episode of self-deception and solves the selectivity problem as a complex whole that includes intentions and reasons for the intention. She can say that the selectivity problem only indicates the possibility of a situation in which we have two persons (we can call them A and B) with equally strong desires that *p* be the case and equal thresholds for accepting and rejecting hypothesis, of which only one person self-deceivingly starts to believe that *p* is the case. And the intentionalist supposes that what explains the difference between A and B is the formation of the intention to believe that *p* on the part of the individual that self-deceives (e.g. A).⁷

However, this response could only work, as a solution of the selectivity problem, if we suppose two things (not necessarily as a conjunction) about the situation⁸ in which we have A and B. (i) We can suppose that intentions to self-deceive are caused by sheer chance by the reasons that precede the inten-

he believes that the best way to achieve this is to bias his cognitive processes in the ways that Mele discussed.”

⁷ One could object that the example with two persons is not adequate since Bermúdez explicitly writes that the “selectivity problem is not a problem of how two people in similar situations can acquire different beliefs” (2000, 317). However the example would work even if we suppose that there is one person with an identical psychological makeup at different moments in time.

⁸ This situation, by supposition, involves complete psychological (functional) identity between A and B in moments before A forms an intention to acquire some belief.

tion formation. Or (ii) we can suppose that intentions to self-deceive are an effect of a conscious decision.

Response (i) does not sound very plausible, since we would need to suppose that what precedes formation of an intention does not influence in any substantive way the intention to form a self-deceiving belief. This would go against our folk-psychological conception of ourselves, which include law-like generalizations that figure in action-explanations (see e.g. Churchland 1981).

If we suppose (ii) then we could explain the difference between A and B in a plausible way, since we could say that A (unlike B) deliberated about the issue at hand and decided that she will intentionally deceive herself into believing that *p* is the case. As a consequence of that decision A could then engage in performing the necessary steps to cause herself to believe that *p*. However, response (ii) is not plausible as an intentionalist account of self-deception because it could not encompass and explain garden-variety cases of self-deception. Namely, it overly intellectualizes the core cases of self-deception since it supposes that they have the form similar to Pascal's wager and this is something that even Bermúdez (2001, 318) recognizes as not being vindicated by his argument. Furthermore, such overly intellectualized cases could even be incorporated in anti-intentionalist accounts since they do not suppose that a person can never decide to acquire a belief and then to undertake the necessary steps to accomplish that goal. However, the crucial thing that divides the two accounts of self-deception is exactly the explanation of garden-variety cases of self-deception (see Mele 1997, 99).

Hence, it seems that intentionalist accounts do not present a proper explanation of garden-variety cases of self-deception. Since, if genuine, the selectivity problem can be raised even against them.

4. Value-based solution to the selectivity problem

More recently Pedrini (2010) has offered a possible solution to the selectivity problem by supposing that what makes the difference between a self-deceiver and a person who resists self-deception is the *embodiment* of epistemic values that the latter exhibits. To embody an epistemic value is to be disposed to behave in certain ways. Here is an example from ethics:

[i]f *S* embodies the value of generosity, to the extent that generosity can be attributed to *S* as one of her traits, *S* is disposed to a wide range of behaviors that we define as generous: perhaps she offers her friends a dinner to celebrate their birthdays, or she

will help others in need even if they are strangers, etc. (Pedrini 2010, 133.)

Similarly, Pedrini claims that a person who *embodies* epistemic values, such as a value of truth, epistemic accuracy, completeness and impartiality (ibid.) will be disposed to behave in such a way that will preclude that person from entering into a state of self-deception. Thus, according to this account, a person can have a desire that *p*, a desire that could cause her to lower the acceptance threshold for the hypothesis that *p*, but she will not be disposed to self-deceive if she really embodies the epistemic values protective against self-deception. For example, if a mother has a desire that her son is innocent and has a low acceptance threshold for this hypothesis, if she embodies the relevant epistemic values, which constitute her dispositions to behave in an epistemically virtuous way, she will override her biasing desire and start to evaluate the evidence in order to find out the truth, no matter what are the possible costs associated with finding new and possibly devastating evidence for her son's case.

Pedrini's account is an interesting *prima facie* solution to the selectivity problem. It introduces the evaluative element by supposing that what explains the selectivity of self-deception is the presence or absence of epistemic values to which the agent is committed. Moreover, it seems that Pedrini's account can improve and amend Mele's use of the FTL model, since now we can explain what, besides desires, determines the rejection/acceptability thresholds for hypotheses. In an epistemically virtuous agent, her epistemic values determine the relevant threshold, and protect from its being affected and modified by desires. So, using Pedrini's account we can explain the difference between agents who deceive themselves and ones who do not by appealing to their evaluative structure.

However, there is an objection that can be pressed against this latter account. Pedrini (2010, 134-135) examines one possible objection. One might say that epistemic values cannot be enough to secure one from exhibiting an episode of self-deception because one can be weak-willed and in the presence of a desire that *p* can become akratic and unable to employ their best epistemic strategies. That is, in that case, a person would, against her best epistemic judgment about how to treat the evidence, be unable to act in accordance with that judgment, and consequently the desire that *p* would cause her to treat available evidence in a motivationally biased way. So in this kind of case (epistemic *akrasia*) embodiment of epistemic virtues apparently would not secure one from forming self-deceptive beliefs, and

furthermore, would not resolve the selectivity problem.

Pedrini tries to show that the case of *akrasia* is not a counterexample to her account because her account of embodied values is *designed* to rule out cases of epistemic *akrasia*⁹ (Pedrini *ibid.*, 134). She explains her idea by introducing the distinction between “something being judged valuable” and “valuing something” (*ibid.*). For example, one may say that helping others is good, but then never exhibit the helping behavior when in a situation where *helping* behavior is appropriate. For that person we would say that she does not really *value* helping others. On the other hand, if someone values something, then if the right circumstances occur, she will be *disposed* to act in accordance with her judgment. According to Pedrini (*ibid.*), to *really value* something is to *embody* that value and if we really embody a value then it seems that we cannot fail to exhibit that value when the circumstances are appropriate.

However, the introduction of the distinction between ‘judging valuable’ and ‘valuing’ does not get to the core of the problem of the possibility of epistemic *akrasia*. In order to evaluate the plausibility of the latter counterexample (and its supposed rebuttal) we need to restrict our attention only to cases of ‘valuing’, where this attitude can be construed as being manifested in a *sincere* judgment that something is valuable.¹⁰ So what is ‘valuing’? According to Pedrini (2010) to value something is to embody that value, and to embody a value is to be disposed to act in accordance with the corresponding sincere value judgment when the circumstances are appropriate. Therefore, it seems that what plays a crucial role in explaining valuing is the notion of a *disposition*, since disposition explains what it means to *embody* a value. The notion of a disposition involves all kinds of intricacies that have not yet been philosophically resolved; however there are a few commonplaces (platitudes) concerning dispositions

that are relevant for the present article. It is widely recognized that dispositions involve some notion of conditionality that connects their *stimulus* and *manifestation* conditions.¹¹ For example, let us take into consideration the notion of *fragility*. Typically, the fragility of an object manifests itself in the object’s tendency to break under suitable conditions. However, to ascribe fragility to an object it is not necessary that its typical manifestation always occurs; a fragile object will still be fragile even if it never manifests that property (e.g. it never breaks). Dispositional properties will be manifested only if certain conditions are satisfied, that is, only if certain antecedent or *stimulus* conditions are fulfilled (cf. Mumford 1998, 6). Since veridical ascription of dispositions does not require occurrence of typical manifestation, a thing can have a dispositional property that is never manifested. For example, an object can be fragile, without ever breaking because stimulus conditions for its becoming broken never occur. Also it is possible that even if stimulus conditions are present the circumstances in which manifestation usually occurs do not have to obtain because some interfering factors are also present. For example, when emerged into the water (stimulus condition), sugar will typically dissolve (manifestation of solubility), but if the water were already saturated, then sugar’s solubility would not be manifested.

These same features of dispositions when translated into the talk about judgments and value take the following form. For example, a plausible formulation of value judgment internalism, given by Michael Smith, takes the following form: if a person judges that it is right to do A, then if rational, she will do A (Smith 1994, 62). According to this view, if someone does not behave in accordance with her sincere value judgment, it does not mean that she does not embody the value expressed by that judgment, it only means that she is exhibiting some kind of irrationality (perhaps she is depressed or weak-willed, etc. (*ibid.*)); that is, it just means that the disposition to act in a way that corresponds to one’s sincere value judgment does not get manifested.¹² Now the question can be posed: does the notion of an epistemically virtuous agent (or an agent that embodies epistemic value) play a role similar to *stimulus* conditions (that encompass the rationality

⁹ When I use the expression ‘epistemic *akrasia*’ I intend it in the sense in which it is used in theory of action with the obvious limitation, since in the epistemic case, the action will be concerned with theoretical investigation whose final product is a belief (and I do not presuppose that the belief formation itself needs to be intentional). Accordingly, epistemic *akrasia* will concern the failure of employing epistemic strategies that one judges to be the best (see Pedrini 2010, 134–135).

¹⁰ When I say sincere judgment I mean to include the idea that the agent not only believes of herself that her value judgment is sincere (where the latter belief could be mistaken), but that the value judgment really (actually) manifests what she really (actually) *values*.

¹¹ What kind of conditionality will the notion of a disposition involve and whether it will always be a counterfactual, is a controversial philosophical issue. For a discussion see e.g. (Bird 2007, section 2.2); Mumford 1998, chapters 3, 4).

¹² Still in other words, we can say that the disposition gets defeated, but not erased.

condition) or the role similar to the dispositional property itself that manifests certain properties in the right *stimulus* conditions?¹³

If we equate the embodiment of epistemic values with the disposition to exhibit epistemically virtuous action then the problem of epistemic *akrasia* lurks back in. Namely, if embodiment of epistemic values involves dispositions to act in certain ways, then the ascription of the latter disposition(s) can be veridical without manifesting its consequence(s) in any particular case. For example, if we take the latter line, it can still be the case that a person self-deceives that *p* is the case even though she still has the disposition to behave in an epistemically virtuous way. To take the standard philosophical platitude, the disposition would be defeated, but not erased, because the desire that *p* be the case is so overwhelming in this particular situation that it manages to *silence* (defeat) the otherwise epistemically stable character. Thus if we take the latter reading as something that Pedrini proposes then the case of epistemic *akrasia* becomes a problem for her account once again.

Hence, it seems that in order to avoid the former consequence we should equate epistemic value embodiment with *stimulus* conditions or, in Michael Smith's sense, with the rationality condition. If we take this line of thought then it seems that we can exclude cases of epistemic *akrasia*. Namely, just like in the practical case, where a person who does not act in accordance with her sincere practical judgment exhibits a case of practical irrationality, we can say that when a person self-deceives she exhibits behavior (or thought) that is *ipso facto* not epistemically virtuous.

However, even if plausible to a certain degree, the value-based solution of the selectivity problem is not without its problems. For example, we can ask whether the embodiment of epistemic values come in degrees or is an absolute characteristic; and connected to the latter issue, whether we should take the embodiment of values to be contextual, so that one can fully embody one and the same value in certain life circumstances and fail to embody it in another, or not.¹⁴ If we suppose that embodiment comes in

degrees then it would seem that the problem of *akrasia* could again become salient because there is no *obvious* boundary that would delimit degrees of embodiment that would exclude cases of *systematic akrasia*. *Prima facie*, it seems that we could only be certain that full (absolute) embodiment of values prevents cases of *systematic akrasia* to occur. So, *prima facie*, it seems that a person who is not fully epistemically virtuous could in certain circumstances always be overcome with a desire that *p* and consequently be induced in believing that *p*; and there does not seem to be any obvious reason to suppose that this akratic self-deception cannot be systematic.

We can avoid this potential problem by supposing that the embodiment of epistemic values should be taken in an absolute (non-degree) sense. However, in that case one could object that the solution of the selectivity problem in terms of embodiment of epistemic value loses its explanatory power. The idea is that one could admit that on the general level non-embodiment of epistemic values explains why self-deception occurs; usually people self-deceive because they are not epistemically virtuous. However, one could still argue that proper non-embodiment of epistemic values does not explain Bermúdez's original formulation of the selectivity problem, that is, why self-deception "occurred in *that* particular situation" (Bermúdez 2000, 317). Since most people do not exhibit stable epistemic dispositions (they are not epistemically virtuous [see e.g. Samuels and Stich 2004]), the fact that they do not fully embody epistemic values would not have much explanatory power to explain why self-deception occurred in some, but not in other particular cases. In this case, someone could say that the selectivity problem becomes acute again even for the account that introduces embodiment of epistemic values. At this point one might again argue that two persons can have the same desire that *p*, be at the same (non-perfect) level of epistemic value embodiment, and have the same rejection/acceptance thresholds and still differ with regard to the exhibition of self-deception (one self-deceives, while the other does not). Consequently, that person might argue again that what is needed is the intention on the part of the agent in order to explain her exhibiting self-deceiving belief. However, as I argued earlier, I do not believe that this strategy actually works as a solution of the selectivity problem.

The latter remarks about the value-based account are not meant to prove the falsity of the account; however, they do point out some of the gaps

¹³ Of course an epistemically virtuous agent will embody a whole set of epistemic values and correspondingly will be disposed to exhibit a whole range of subdispositions that comprise epistemic virtues (such as evidence gathering, evidence evaluation, sound reasoning, etc.).

¹⁴ For example, a successful scientist may embody all the relevant epistemic values in the context of her professional work, but be completely unresponsive to them when issues concerning her private life become salient. Does such a person fully embody relevant

epistemic values in one context but not in the other, or does she overall embody them to a certain degree, etc.?

or misgivings of the account. Pedrini (2010, 135) also claims that her account is not a final word on the matter but that it provides a framework in which the selectivity problem could be solved. In section 5 I will try to show why I find the selectivity problem dubious, and how the intuitions that support its genuineness can be explained away by using the framework of Pedrini's value-based account.

5. Dissolution of the *selectivity problem*

At this moment one might start to despair that there is no solution of the selectivity problem, since the same problem can be raised even against intentionalist accounts, and the other proposed account (value-based account) does not seem to be adequate. However, I believe that the discussion in section 3 enables us to see the mootness of the *selectivity problem* as a genuine problem that sophisticated accounts (such as Mele's 2001) of self-deception face. In section 3 I stated that, if genuine, the selectivity problem can be directed towards intentionalist accounts¹⁵, because we can always raise the question of why in this particular situation a desire that *p* be the case (or to believe that *p* is the case) has caused one to acquire an intention to form a belief that *p* is the case.

However, I believe that an intentionalist may properly solve the selectivity problem (as posed in section 3) simply by saying that that is just how the mechanism works, or at least how our folk-psychological explanation, in simple cases, depicts the workings of the mechanism that produces intentional action. In normal cases, we have a strong desire that *p* and a belief that if I do *q* I will accomplish *p*, which causes us to decide to form an intention to make *q* true. Similarly in the case in which we adopt an intention to self-deceive that *p* is the case. I think that from a folk-psychological perspective the selectivity problem puts too demanding a requirement on the adequacy of an action-explanation; if we take it as a genuine problem then it would be baffling how even in everyday cases an intention gets formed on the basis of the reasons that precede it.

However, if we grant this answer to the intentionalist then by parity of reasoning we have to grant it to the anti-intentionalist as well. The only difference will consist in the fact that the anti-intentionalist will claim that it is not necessary to postulate an intention to acquire a belief in order to account for garden-variety cases of self-deception.

At this point the question that arises is why we feel the intuitive cogency of the selectivity problem as proposed by Bermúdez (1997, 2000)? The intuitive pull of the problem might be invited by Mele's assertion that his conditions for self-deception are *conceptually* sufficient (see e.g. his [2012, 2]). This idea of conceptual sufficiency of Mele's conditions should be read as giving a proto-analysis (ibid.) of the *self-deception* concept, and presumably its application conditions. So, one might wonder and doubt whether it is really *conceptually sufficient* that when Mele's conditions are satisfied then somehow by conceptual necessity it follows that we have a case of self-deception.

However, it seems to me that, regardless of what is meant by *conceptual sufficiency* in this context, when it comes down to explaining actual cases of self-deception the sufficiency of the thresholds for hypotheses acceptance and rejection (modified by the desire that something be the case) must include a *ceteris paribus* clause since psychological explanation is a paradigmatic species of explanation that demands *ceteris paribus* laws (see e.g. Churchland 1981, Davidson 2001, Fodor 1991), and if I understand Mele (2001, 2012) correctly, in giving a deflationary account of self-deception he intends it to be psychologically adequate. Hence, the question that now becomes pertinent is in what sense desire's lowering the acceptance threshold of a particular hypothesis is not sufficient for self-deception?

Logical sufficiency (in the sense that the threshold for a hypothesis entails self-deceptive formation of a belief in that hypothesis) presumably cannot be adequate, since an empirical phenomenon does not presuppose logical necessity. Thus, just because we can imagine that a motivationally modified hypothesis testing mechanism (that satisfies Mele's other conditions) does not *necessarily* lead to self-deception is not an argument against its empirical sufficiency. On the other hand, if we take it that the notion of sufficiency is causal then this notion of causality cannot simply involve *nomic necessity* since self-deception depends on contingencies of human nature (its less than perfect *epistemic nature*).¹⁶ Thus we need to take into account the notion of *ceteris paribus* condition. But when we consider the selectivity problem in this light then the problem itself loses its weight. The question 'why one self-deceives in one situation and not in others' (Ber-

¹⁵ At least against those accounts that have the form displayed in (Bermúdez 1997, 108).

¹⁶ For example, when Bermúdez states "[e]ven if one desires both that *p* be true and that one come to believe that *p* it is not *inevitable* that one will form the belief that *p*" (2000, 318, my emphasis) it is possible that he has this strong reading of sufficiency in mind.

múdez 2000, 318) in a sense dissolves because the anti-intentionalist may argue that cases where self-deception does not occur even though the acceptance/rejection thresholds are appropriately motivationally modified comprise cases in which *cetera* are not *paria*,¹⁷ that is those are cases in which normalcy conditions for the occurrence of self-deception do not obtain.¹⁸

So, my claim is that from the perspective of a (folk) psychological explanation there is no problem for a sophisticated anti-intentionalist to explain why in certain situations a desire can be sufficient to induce a self-deceiving belief and in others will not be sufficient. The dissolution of the selectivity problem, according to this view, consists in taking seriously into account the *ceteris paribus* conditions that psychological explanations contain. Now, this invoking of *ceteris paribus* conditions might seem as a sort of magic bullet. In order to dispel these kinds of doubts, in the next section I will sketch a model of the conditions in which self-deception can be reasonably expected to occur.

5.1 Amending the value-based account of self-deception

Earlier, in section 4, I argued that Pedrini's (2010) value-based account could involve two notions of epistemic virtue: epistemic virtue can be fully (absolutely) embodied or it can be embodied to a certain degree. On the one hand, the problem with the notion of full embodiment was that it was not very explanatory in the context of the selectivity problem since people are usually very far from embodying epistemic virtues. On the other hand, the seeming problem with the graded notion of embodied epistemic virtues was its inability to determine when a real cognitive agent is secured from self-deception or will undergo a biased treatment of evidence caused by a desire that something be the case, since ordinary cognitive agents regularly fall short of embodying epistemic virtue (usually to a varying degree).

¹⁷ Despite the fact that Bermúdez claims the opposite. See (Bermúdez 2000, 317) where he claims that: "[...] possessing a desire that *p* be true is not sufficient to generate cognitive bias, even if all other things are equal [...]."

¹⁸ This also refutes Pedrini's contention that desires are not sufficient to cause motivationally biased treatment of evidence that leads to self-deception (see Pedrini 2010, 134). Since conditions in which desires become sufficient to cause biased treatment of evidence that lead to self-deception exclude conditions in which an agent fully embodies epistemic virtues (i.e. *cetera* are not *paria*).

However the above remark concerning the graded conception of epistemic value embodiment is only a problem if we think that desires are not sufficient to bias one's cognitive abilities, other things being equal. But if we take desires to be sufficient, other things being equal, then grades of distance of embodiment of epistemic value from the ideal can figure in the sufficiency clause as a part of the normalcy condition that characterizes the background of desires' (that cause self-deception) *ceteris paribus* clause.¹⁹

Let me try to develop the latter thought a bit. As I interpret it, according to anti-intentionalists (of Mele's deflationary kind) a desire that *p* be the case is sufficient, *ceteris paribus*, to bias one's cognitive abilities and consequently to produce self-deceptive belief. *Ceteris paribus* condition involves the background conditions that constitute the *normal* circumstances in which desire causes self-deception. In our case, normal conditions that allow self-deception to occur, among other things, involve the condition that self-deceivers embody epistemic virtues to some degree.

So, how can this vague idea of a degree of embodiment of epistemic values play any explanatory role in this account? The answer I have in mind is provided in the following rough sketch: when relevant conditions are satisfied and there is a strong desire that something be the case, then, normally people (who have that desire) self-deceive.²⁰ That people normally self-deceive means that the statistical probability of self-deception is sufficiently high given the aforementioned conditions.²¹ So where do degrees of epistemic value embodiment enter the picture? They enter at the level of background normal circumstances; the supposition is that what determines the statistical probability of forming the self-deceptive belief that *p* given the desire that *p* is the distance of the behavior of the actual epistemic agent from its corresponding ideal counterpart.

¹⁹ *Ceteris paribus* clauses (and scientific laws that use them) are notoriously hard to adequately characterize. For a taxonomy of different senses of *ceteris paribus* sentences see Schurz (2002). However, it seems that recently a *prima facie* consensus (or a promising optimism) emerged that *ceteris paribus* clauses may be characterized by using so called normality accounts (see e.g. Reutlinger, Schurz and Hüttemann 2011).

²⁰ Or more specifically people exhibit confirmation bias that leads them to self-deceive, but not to complicate things I will just write as if desire in normal situations lead directly to self-deception.

²¹ This is called the statistical probability consequence thesis. The thesis is somewhat controversial, but see Schurz (2001) for a forceful defense.

To make this idea more precise let us represent an actual cognitive agent and her corresponding ideally epistemically rational counterpart with functions Ψ and Φ , respectively. Cognitive agents can be described as mappings between input world states (or states as perceived by the agent) and output mental (belief) states, e.g. mapping Ψ takes evidence (world states) and outputs hypothesis (belief states) such that $\Psi(e_i) = h$, where i is an index of some evidence set.

We can define a distance between the real and the ideal cognitive agent in the following way: let d be a distance function that takes Ψ and Φ as its arguments. Also let ρ be a function from the set of cognitive agents to real numbers between zero and one such that $\rho_i(\Psi) = 1$ if agent Ψ behaves in a fully epistemically rational way (fully embodies epistemic virtue) and $\rho_i(\Psi) = 0$ if Ψ behaves in a completely irrational way from the epistemic perspective (does not embody epistemic virtue at all).²² In other words, the idea is that ρ gives a degree to which agents rationally treat/use/interpret evidence and form/evaluate hypotheses and it tries to capture the rationality of an agent's behavior across different situations, where index i equals some natural number and signifies the difference in contexts in which a cognitive agent may exhibit (or not) its epistemic virtues. As a tentative and provisional formalization we can take the following equation to determine the distance function between the actual cognitive agent Ψ and its ideal counterpart Φ ²³:

$$d(\Psi, \Phi) = \frac{1}{N} \sum_{i=1}^N (\rho_i(\Phi) - \rho_i(\Psi))^2$$

Since Φ fully embodies epistemic virtues ρ_i will always assign her 1, but $\rho_i(\Psi)$, given that Ψ is a real, cognitively (and otherwise) limited, agent can take any value between zero and one. Hence, if we take only the extreme possible values, then across situations, if $\rho_i(\Psi) = 1$ then $d(\Psi, \Phi)$ will equal zero. Which means that there is no difference between the actual agent and its epistemically rational counterpart in that particular situation; or we can say that Ψ fully embodies and manifests in a particular situation the epistemic virtue. If $\rho_i(\Psi) = 0$ then $d(\Psi, \Phi)$ will equal one which would mean that Ψ is at a complete distance (according to this measure) from its ideal

counterpart and accordingly does not embody epistemic value.²⁴ Thus in general, $d(\Psi, \Phi)$ will equal the average difference between the epistemic behavior of Ψ and Φ , across different situations in which cognitive agents might find themselves.

Using the above formalization we can express the idea that the distance between the ideal epistemic agent and the real agent determines the statistical probability that the agent will be prone to epistemically irrational behavior, and the greater the distance between Ψ and Φ the more likely it is that the a desire that p be the case will influence Ψ 's belief forming processes. Hence, the *appropriate* values of the distance ($d(\Psi, \Phi)$) measure constitute the background normal conditions that need to obtain in order for the desire to be able to influence (cause) one's self-deceiving belief formation. I wrote appropriate values of d since, if e.g. $0.09 \geq d(\Psi, \Phi) \geq 0$ then presumably desire that p be the case will have negligible or no influence on the Ψ 's belief forming processes. However, since we suppose that self-deception is a real phenomenon then there must be appropriate distance between Ψ and Φ (relative to some standards of rationality or virtues that an ideally rational being embodies) which constitutes the background conditions that allow the phenomenon to occur given that the desire that something be the case is present (and suitably intense or strong), that is, that the statistical probability of self-deception (given that a certain desire occurs) is sufficiently high.²⁵ So, the basic idea is that the greater the distance between the actual and ideal cognitive agent the more probable it will be that the actual cognitive agent self-deceives that p given her desire that p be the case.

From the perspective of the ideas developed in this article Bermúdez's (2000) claim that "possessing a desire that p be true is not sufficient to generate cognitive bias, even if all other things are equal (...)" (p. 317) is not adequately supported. The dissolution of the selectivity problem, in the framework of folk-psychology, invokes the idea that the desire that p be the case causes motivational bias, *ceteris paribus*; where *ceteris paribus* conditions include an appropriate distance degree between the real cognitive agent and her ideal counterpart or equivalently that there is a certain average difference between the

²² What determines the standards of epistemic rationality or epistemic virtue is a matter for epistemology and philosophy of science to discuss and settle. Here I will remain silent regarding the determination of standards of epistemic rationality.

²³ One can immediately see that the equation is an instance of the Brier Score.

²⁴ Or we can say that she embodies the epistemic value to a degree 0.

²⁵ And ideally d should be empirically measurable. One of the conditions for this depends on our being able to agree on which epistemic standards should be used as standards of evaluation of cognitive abilities and generally epistemic behavior of individual agent.

real and the ideal that constitutes the *normal conditions* that *allow* or create circumstances in which a desire can cause biased treatment of evidence and consequently self-deceptive belief formation.

6. Conclusion

In this paper my aim has been to discuss and examine the selectivity problem as proposed by Bermúdez (1997, 2000). Bermúdez presented the *selectivity problem* as a problem that counts against anti-intentionalist accounts of self-deception. Bermúdez also argued that intentionalists have a ready-made solution to the selectivity problem and consequently that this problem counts in favor of intentionalists' accounts of self-deception.

I argued that the selectivity problem can be raised even against intentionalist accounts. Furthermore, I argued that the fact that a selectivity problem can be raised against intentionalist accounts indicates the way in which the problem could be dissolved. The latter fact shows that the insistence on the cogency of the selectivity problem puts a too demanding requirement on what constitutes a sound (folk) psychological explanation. Therefore, in the final sections I set out to analyze the intuitions that gave a convincing tone to the selectivity problem and tried to accommodate them using resources that do not go beyond the commitments of a sophisticated anti-intentionalist account.

References

- Bermúdez, J., L. (1997). "Defending intentionalist accounts of self-deception." *Behavioral and Brain Sciences*, 20: pp. 107–108.
- Bermúdez, J., L. (2000). "Self-Deception, Intentions, and Contradictory Beliefs." *Analysis*, 268: pp. 309–319.
- Bird, A. (2007). *Nature's Metaphysics: Laws and Properties*, Oxford: Clarendon Press.
- Churchland, P., M. (1981). "Eliminative Materialism and the Propositional Attitudes." *The Journal of Philosophy*, 78: pp. 67–90.
- Davidson, D. (2001). *Essays on Actions and Events*. Oxford: Clarendon Press.
- Davidson, D. (2004). *Problems of Rationality*. Oxford: Clarendon Press.
- Deweese-Boyd, Ian, "Self-Deception." *The Stanford Encyclopedia of Philosophy (Spring 2012 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2012/entries/self-deception/>.
- Fodor, J., A. (1991). "You Can Fool Some of The People All of The Time, Everything Else Being Equal; Hedged Laws and Psychological Explanations." *Mind*, 397: pp. 19–34.
- Friedrich, J. (1993). "Primary Error Detection and Minimization (PEDMIN) Strategies in Social Cognition: A Reinterpretation of Confirmation Bias Phenomena." *Psychological Review*, 2: pp. 298–319.
- Johnston, M. (1988). "Self-Deception and the Nature of Mind." *Perspectives on Self-Deception*. In B. McLaughlin and A. O. Rorty (eds.). Berkley: University of California Press: pp. 63–91.
- Lazar, A. (1997). "Self-Deception and the Desire to Believe." *Behavioral and Brain Sciences*, 20: pp. 119–120.
- Lazar, A. (1999). "Deceiving Oneself or Self-Deceived? On the Formation of Beliefs 'Under the Influence'." *Mind*, 430: pp. 265–290.
- Mele, A. (1997). "Real self-deception." *Behavioral and Brain Sciences*, 20: pp. 91–102.
- Mele, A. (1998). "Motivated Belief and Agency." *Philosophical Psychology*, 3: pp. 353–369.
- Mele, A. (2001). *Self-Deception Unmasked*. Princeton: Princeton University Press.
- Mele, A. (2012). "When Are We Self-Deceived?" *Humana Mente*, 20: pp. 1–15.
- Mumford, S. (1998). *Dispositions*. Oxford: Oxford University Press.
- Pedriani, P. (2010). "Cognition and Desires: How to Solve the 'Selectivity Problem' for Self-Deception." *Pratiche della cognizione, Proceedings of the AISC2010, Italian Society for Cognitive Sciences*: pp. 131–135.
- Reutlinger, Alexander, Schurz, Gerhard and Hüttemann, Andreas, "Ceteris Paribus Laws." *The Stanford Encyclopedia of Philosophy (Spring 2011 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2011/entries/ceteris-paribus/>.
- Samuels, R. and Stich, S. (2004). "Rationality & Psychology." *The Oxford Handbook of Rationality*. In A., R. Mele and P. Rawling (eds.). New York: Oxford University Press: 279–300.
- Schurz, G. (2001). "What is 'Normal'? An Evolution-Theoretic Foundation for Normic Laws and Their Relation to Statistical Normality." *Philosophy of Science*, 4: pp. 476–497.
- Schurz, G. (2002). "Ceteris Paribus Laws: Classification and Deconstruction." In *Ceteris Paribus Laws, Erkenntnis*, 57: pp. 351–372.
- Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell Publishing.
- Trope, Y., & A. Liberman. (1996). "A Social Hypothesis Testing: Cognitive and Motivational Mechanisms." *Social Psychology: Handbook of Basic Principles*. Edited by E. Higgins and A. Kruglanski. New York: Guilford Press.

AN ACQUAINTANCE CONSTRAINT AND A COGNITIVE SIGNIFICANCE CONSTRAINT ON SINGULAR THOUGHT^{*}

Mirela Fuš

Central European University, Hungary, Budapest

e-mail: mirela.fus@gmail.com

Abstract:

Among Singularists, it has been widely accepted that one can have singular thought by acquaintance, and that acquaintance encompasses the perceptual acquiring, memorizing and communicating of singular thoughts. I defend the possibility of having a singular thought via extending acquaintance to intermediaries other than just through written and spoken words. On my account, singular thought includes two types of representations, namely *indexical-iconic representation* and *indexical-discursive representation*. Also, it is determined by two constraints: (i) *the acquaintance constraint*: singular thought includes a causal-historical relation to the object; and (ii) *the cognitive significance constraint*: we have a cognitive capacity to deliberately encode an *indexical-iconic representation* into an *indexical-discursive representation*, which enables us to have a singular thought.

Key words: singular thought, acquaintance constraint, cognitive significance constraint, indexical-iconic representation, indexical-discursive representation.

1 Introduction

Singularists find it utterly implausible that none of our thoughts are about a particular object and that all that we perceive is given to us in an entirely qualitative manner. Therein lies their strong motiva-

tion to define and defend a singular or *de re* thought. Some Singularists have argued that this is possible only in the case of our more or less immediate perception or, as Acquaintance Theorists and Cognitivists argue, due to a cognitive significance for a thinker, while Semantic Instrumentalists place less or no restrictions on one's ability to have a singular thought.

In a recent book, *New Essays on Singular Thought*, the editor Robin Jeshion (2010, 16) identifies the four main topics concerning singular reference and thought. Let me summarize and list them briefly: (i) *the structure and nature of singular thought*; (ii) *conditions on thinking singular thought*; (iii) *the nature and scope of acquaintance* for perceptual-demonstrative and communication-based reference and thought; (iv) *the relationship between singular thought and the semantics of fictional characters and mythical terms*.

I defend the possibility of extending acquaintance in order to have a singular thought. In other words, I am particularly interested in the problems concerning Jeshion's third topic as I extend the notion of acquaintance to non-propositional media of reference¹. Furthermore, I allow a possibility of hav-

^{*} I wish to extend my warmest thanks to Professor Kent Bach and Professor François Recanati and to the other outstanding researchers I had the opportunity to discuss my presentations with. I have especially benefited from the comments and discussion with Robin Jeshion and Michael Devitt. I am also grateful to Professor Luca Malatesti and Professor Nenad Miscevic for several useful suggestions as well as remarks on earlier drafts of this paper.

¹ My view, even though I defend the possibility of having a singular thought via pictures and similar intermediaries, substantively differs from the standard views in theories of depiction. For this reason, I will briefly mention the standard views, but I will not discuss them here. To put it in a nutshell, an alternative or additional view about *singular thoughts* has been discussed by authors (e.g. Zeimbekis, Lopes, Walton, Goodman) who argue *for* or *against* the view according to which it is possible to claim the same for pictures and (in some cases) other

ing a singular thought via extending acquaintance to intermediaries other than just written and spoken words. I argue that intermediaries such as photographs or video and audio recordings can, in certain cases, acquaint us with objects. My approach to singular thought is two-pronged because it includes both external and internal elements, and it is broad in scope since it is not restricted to just certain intermediaries.

In this paper, I emphasize that singular thought about material objects is deeply rooted in our perception no matter whether we talk about immediate perceptual acquiring, memorizing or communicating singular thoughts. Second, I stress that proper names and, in certain cases, even indexicals and definitive descriptions have played a significant role in communication of singular thoughts. However, extending acquaintance to intermediaries other than just written and spoken words has been problematic for many advocates of singular thought. At the same time, we have a very strong intuition that we can have singular thought about persons or objects we perceive via other intermediaries such as photographs or video recordings.

2. General vs. singular thought – *pro et contra*

2.1. Russell and Frege

Russell and Frege are considered to be the pioneers of the notion of singular thought². However, as noted by Jeshion, the sole demarcation between singular and descriptive thought does not tell us “what it is for a thinker to grasp a general or singular proposition nor the conditions under which a thinker can grasp such propositions” (Jeshion 2010, 3) since it is about the structure of thought contents only.

intermediaries. Very briefly, theories of depiction considering remote acquaintance can be divided into two camps. One camp supports the possibility of a remote acquaintance either through *demonstrative identification* (Lopes 1994) or *transparency theory* for photographs (Walton 1984). Another camp argues that pictures cannot acquaint us with particulars. Rather, as some of them will argue, they can acquaint us with certain *kinds of properties*. For example, Zeimbekis (2009, 11) claims that “the contents of picture-perceptions do not themselves provide the kind of numerical and contextual information required for singular thought.”

² Frege (1984/1892, 1984/1918); Russell (1905, 1911, 1946).

Russell’s (1912) distinction between *knowledge by acquaintance* and *knowledge by description* has been used in semantics to demarcate *de re* from *de dicto* thoughts. However, according to him we could only have acquaintance (or singular thought) “with sense-data and (probably) with ourselves” (Russell 1912, 109). Present-day Singularists, even those who defend the Acquaintance Theory of singular thought, often reject Russell’s ‘strict conception of acquaintance’. Following this line of thought there is a difference in content between singular and descriptive thoughts. The content of the former is a singular proposition (having properties and individuals for constituents), while the latter takes a general proposition as its content (consisting of concepts or properties). Both kinds of Russellian propositions are structured and have truth conditions.

On the other hand Frege’s sense/reference distinction, used both for expressions as well as for sentences, characterizes his descriptivist semantics. He applied the sense/reference distinction to singular terms, not only to proper names, but to definite descriptions, demonstratives and indexicals as well. According to Frege, the referent is the object the singular term is about whereas the sense which is taken to be its ‘mode of presentation’ is given in terms of its conceptual condition. For example, the proper name “Lewis Carroll” refers to a person, namely to Charles Lutwidge Dodgson and the sense of it can be given using a definite description “the author of the *Alice’s Adventures in Wonderland*”. This singular term is also interesting since it reflects another important notion Frege introduced, and it presents his argument from cognitive significance. Singular terms which co-refer, as “Lewis Carroll” and “Charles Lutwidge Dodgson” do, both contain the same referent³. As Jeshion puts it:

Yet, intuitively, a fully competent, rational agent might accept one while denying (or just not accepting) the other, and thus the propositions semantically expressed by the sentences cannot just contain, as the value of the proper names, their referents. To account for the differing cognitive significance of the co-referring singular terms, Frege postulated that they contribute senses, not referents, to the propositions expressed by the sentences in which they occur. Such propositions,

³ Frege’s well-known example of the co-referring terms Hesperus/Phosphorus which both refer to planet Venus.

composed of senses that are purely qualitative, are what we grasp and think in understanding sentences, and thus all our thought has as its content general, not singular, propositions. (2010, 4)

Thus, for the Fregeans, introducing senses and thinking under different modes of presentation when it comes to (co-referring) singular terms should save the rational agent from being inconsistent, even though the referent of the singular terms she uses is one and the same object or person.

In the next two sections I will very briefly summarize the main advantages and disadvantages of descriptive and singular thought.

2.2. Descriptive thought

In addition, Jeshion (2010) mentions two advantages of the accounts that recognize only descriptive thoughts. I will give a brief description of both of them. I will call the first advantage the *straightforwardness of a descriptive thought* and the second one the *non-sensitivity to remoteness of relation to the object*. The advantage of the former consists in that a thinker only has to grasp a general proposition “that has a conceptual condition that [an] object satisfies” (Jeshion 2010, 4) in order to have a descriptive thought. Moreover, thinkers are able to “employ them compositionally in thought” (*ibid.*). For example, one can think of the greatest philosopher by combining the concepts *greatest* and *philosopher* together with the definite concept *the*. The latter advantage tells us that “requirements for descriptive thought remain constant across various different ways in which a thinker may be related to the object of thought” (Jeshion 2010, 5). To illustrate, my relation to an object that I perceive when I stand in a close perceptual relation to it, for example to the kitten, is not as remote as when I stand in relation to some other spatially distant object, for example Daniel Dennett. However, according to Descriptivists, my thought content and the constraints on thinking of them are the same in both cases. Similarly, in the case of temporal remoteness one’s thought does not change no matter whether one thinks about Obama or Roosevelt. The reference-determination for descriptive thoughts seems to be unproblematic as well, since “the mechanisms of reference-determination for descriptive thought are constant despite variations in thinker’s relation to the

object of thought” (Jeshion 2010, 5). At the same time, the major dilemma of Descriptivist’s accounts is their non-plausibility. Jeshion argues that,

... our conceptualizations of objects are often not what determine which objects we think about. Our thought about object is more direct; it draws upon and uses our various complex relations to the world in securing and determining the reference of the objects we think about (*ibid.*).

2.3. Singular thought

Jeshion (see 2010, 5–13) emphasizes three philosophical “avenues” that influenced the survival of singular thought and a widening of its scope⁴. First, there have been significant developments on the semantics of singular terms. These new developments occurred during the last three decades of the twentieth century, and Fregean semantics (of proper names, natural kind terms, definite descriptions, demonstratives, and indexicals) has been heavily criticized and a new theory of direct reference has been established. In particular, several philosophers argued for direct reference in semantics, e.g. Kripke (1980) on the semantics of proper names with his famous ‘semantic argument’; Donnellan (1972, 1979) on the semantics of definite descriptions by showing that they can be used both referentially and attributively; Kaplan (1979, 1989a, 1989b) and Perry (2000, 2001) on the semantics of demonstratives and indexicals establishing that what is expressed by them is determined in a context of use⁵.

Second, in the philosophy of perception the relationship between the content of perceptual experi-

⁴ She also mentions the achievements from the other areas that have stimulated research on singular thought such as: “grappling with the problem of the contingent *a priori*; debate over conceptual and non-conceptual content of perceptual experience; the rise of dynamic semantics and especially discourse representation theory; the influence of externalism about belief content for natural, social, and artifactual kinds; exploration of the relationship between *de se* and *de re* thought; contemporary research on vision science and auditory perception; analysis of knowledge who; semantic analysis of propositional attitude reports” (Jeshion 2010, 6).

⁵ It is prudent to mention in addition to the aforementioned philosophers, two other prominent philosophers: Putnam (1975) and Marcus (1995).

ence and the content of thought has been discussed⁶. There are two Russellian-inspired positions. The first one suggests that the content of perceptual experience is singular and object-involving. On this account, one experiences the very object in the world one perceives:

The content of my perceptual experience of the grapes actually contains the grapes I see, together with their properties of being purple and ripe, and can be captured as a Russellian singular proposition (Jeshion 2010, 10).

The second position, on the contrary, takes the content of perceptual experience to be a general proposition, one whose contents are existentially quantified⁷.

However, Jeshion points out that the second position has often been criticized by Singularists for two reasons: in short, its failure “to account for the phenomenology of perceptual experience” and that “such an account of content does not allow that perceptual content itself provides one with knowledge of what one is thinking about” (Jeshion 2010, 10).

Finally, the third influence comes from the interface of philosophy of mind and epistemology, developed from the discussion about the mind’s acquaintance with real world objects. These additional questions come to mind: is there any object in the real world that we can be directly acquainted with, what that object is, and what is the scope of acquaintance? Russell’s answer was positive, yet Jeshion explains that:

[w]hile he championed the existence of an acquaintance relation, he held that we are not acquainted with ordinary objects like

grapes and roses because for him standards on acquaintance were exacting, requiring unmediated and complete awareness of objects of acquaintance, immune to errors of identification, and perceptual experience never reaches that standard (Jeshion 2010, 11).

To this argument, the arguments from illusion and hallucination, that is when one’s perceptual experience of objects, phenomenally speaking, is identical to one’s normal perceptual experiences, can be added⁸. Thus, Russell needed a common object for both cases and, since no external object can satisfy that condition, he appealed to a mental object that he thought would do the job. For that reason he introduced the sense-data which have the characteristics of being “mind-dependent, private, possibly non-physical mental images, and they possess all and only the properties that are phenomenally presented to us in perceptual experience” (Jeshion 2010, 11–12). As a result, our thought about the external objects is general⁹.

However, despite the popularity of sense-data theory¹⁰ in the first half of the twentieth century, in the second half of the twentieth century numerous problems arose. For example, one criticism claims that it is hard to identify the properties presented in perceptual experience. To illustrate, Chisholm (1942) argued that when having a perceptual experience of a speckled hen, one is not able to say how many spots it has. While this may be true, Austin (1962) contended that it is implausible to say that this sense-datum possess an indeterminate number of spots. These criticisms opened the door to direct reference, especially with help of recent developments from vision sciences which “affirm that visually attending to objects affords mental reference directly to objects” (Jeshion 2010, 13)¹¹.

⁶ “The content of perceptual experience aims to capture what perceptual experience is like: how we experience the world when we are in a state of perceiving or apparently perceiving the world. How is this related to the content of those thoughts we think on the basis of occurrent perceptual experiences? One facet of this complex topic concerns the representation of objects in perceptual experience. There is much agreement that we typically perceptually experience properties as being properties of objects. We do not simply perceptually experience purple ripeness as at a certain location in the visual field. We perceptually experience the grapes themselves as purple and ripe. Or at least we are disposed to attribute perceptually experienced properties to perceptually experienced objects.” (Jeshion 2010, 9–10)

⁷ See McGinn (1982), Davies (1997), and Tye (2000) for generalist accounts of content.

⁸ For classical presentations of the arguments from illusion and hallucination, see Price (1932), Russell (1946), Broad (1965), and Moore (1993).

⁹ “I think about the grapes by thinking, roughly: there is an object that is a grape-bunch, purple and ripe that is the cause of this sense datum, where ‘this’ makes demonstrative, acquaintance-based, reference to the sense datum I presently have in seeing the grapes” (Jeshion 2010, 12).

¹⁰ Apart from Russell (1946), the sense-data theory was defended by Price (1932), Broad (1965), Moore (1993).

¹¹ Scholl and Pylyshyn (1999), Scholl (2002), Scholl, Pylyshyn, and Feldman (2002), Pylyshyn (2003, 2007).

In the end, I would like to briefly enumerate several sets of problems concerning singular thought as identified by Jeshion (see 2010, 13–16). First, the fundamental set of problems deals with *what singular thought is*. Under this set there are included three problems: (i) If singular thought is necessarily object-dependent, the theory of singular thought needs to explain tokening of empty names and, perhaps, cases of hallucinating. (ii) How and if a singular proposition can be grasped? (iii) What is it that happens in cognition when one thinks referentially and how can one maintain this kind of intentional relation¹²? The second set of problems deals with acquaintance, especially its nature and scope. The third one is about fictional characters and mythical objects.

In the next section I wish to distinguish clearly between two main cases of extended acquaintance and to argue that the situation is not so pessimistic, that there is a way out, both for the cases of perception and memory (that appear in Bach's formulation), and at least for some cases of communication-based *de re* thoughts.

3. A two-pronged approach to singular thought

In semantics, direct reference via direct perception can be achieved through three “degrees” of acquaintance. Hence, among most Acquaintance Theorists, including prominently Bach (1982/84, 1987/94, 2010) there are three types of *de re* or *singular thoughts* about objects. The first and most basic type of *singular thought* is *perception-based de re* thought while the other two, namely *memory-based* and *communication-based de re* thoughts, are derived from the first one (see Evans 1982; Bach 1982/84, 1987/94, 2010; Recanati 1993/1997). The connection between the thinker and the object is *causal-historical*, and we are thinking via a “chain of representations originating with a perception of the object” (Bach 2010, 55). It is also important to stress that singular thinking is determined *relationally*, not *satisfactionally* (see Bach 1987/1994: ch. 1).

The following problem arises among Singularists, especially among defenders of Acquaintance Theory: How far can the acquaintance with the object be extended? In his recent work, Bach asks himself this very question:

... I'm stuck on questions like these. So, for example, does seeing a photograph or film of someone put one in a representational connection with them? Hearing someone's voice? Does reading someone's name do the trick, even outside the context of communication? For example, can you have singular thoughts about someone whose name you read on a luggage tag, in a phone book, or on a tombstone? I'm inclined to think not (Bach 2010, 58).

Conversely, once we admit the possibility of remoteness acquaintance via communication, we have to bear in mind that communication is a process which involves not only spoken or written words, but, potentially, all sorts of physical tokens – i.e. intermediaries – which can be causal-historically included in the representational connection relation. How do we discern which of them are in some way relevant for singular thought, if any?

Bach (see 1982/84; 1987/94) tried to extend the acquaintance relation via the representational connection relation even to the most remote examples (where the causal-historical chain exists), but in his (Bach 2010, 57) later work he questioned this possibility¹³. As I have tried to demonstrate elsewhere (Fuš, forthcoming) this might lead to a sorites paradox. In this paper I give my reaction to the potential threat of the paradox. I will argue that many concepts and relations deserve better treatment and further enhancement. Furthermore, I wish to distinguish clearly between two main cases of extended acquaintance and to argue that the situation is not so pessimistic, that there is a way out, both for the cases of perception and memory (that appear in Bach's formulation), and at least for some cases of communication-based *de re* thoughts.

¹² Those advocating and developing such an approach include: Evans (1982), Bach (1987), Recanati (1993), Millikan (2000), Perry (2001), Lawlor (2001), Jeshion (2002), Taylor (2003), Sainsbury (2005).

¹³ Including when he gives the following examples of extended acquaintance and of one's relation to photography, film, hearing someone's voice, reading someone's name: on a luggage tag, in a phone book, or on a tombstone.

3.1. *Seemingly two types of cases of physical tokens*

Consider first the physical tokens we perceive. One might say that there is a significant difference between cases of hearing or reading words that refer to a certain object on the one hand, and seeing photography or film of this particular object on the other. In both cases physical tokens (made out of the mental tokens) that are causally connected to the object are used. However, we have an intuition that physical tokens of acoustic or visual patterns of items heard or read resemble the object at the end of the causal-historical chain rudimentarily (or not at all) whereas photography or film resemble it more (or to a greater degree). The cases of photography or film might seem especially confusing since we can find a similarity between properties of the object-as-presented in the photograph or on the film on one side, and the actual object the photograph or film really depicts on the other. But let us not be fooled, since in both cases, at least according to Bach, what is important is a causal-historical relation to the object at the end of the communication chain, not the appearance of the physical token itself, at least not in a sense that it should resemble the object at the end of the causal-historical chain.

While this may be true, physical tokens of spoken or written words can be perceived as well, and thus one might say that the receivers might also have immediate perceptual *de re* thoughts about these physical tokens themselves. I believe that considerable confusion stems from not taking into consideration this possibility. Yet with this in mind, demarcation is crucial when it comes to perceiving other kinds of physical tokens, e.g. photographs or film, since we might also have immediate perceptual *de re* thoughts about these physical tokens themselves on the one hand, when on the other we have a strong intuition that they can connect us with the object at the end of the communicational or causal-historical chain, at least when they stand in the causal relation to the object. For example, the perceiver can perceive a visual or acoustic physical token such as the word 'Mary' written or spoken respectively, or perceive the photo or video recording of Mary. In the case of the physical token, which should stand for a (mental) proper name according to Bach, the perceiver can have a communication-based *de re* thought about Mary, while in the case of the photograph or video recording Bach is inclined to think that he cannot.

Do we really want to come to this paradoxical ending? We might agree that the causal link plays the crucial role, and not the object resemblance of the physical token. But we should try to see if there is anything else that matters apart from the pure causal-historical connection to the object at the end of the communication chain.

3.2. *The acquaintance constraint and the cognitive significance constraint*

One way to go is to amend Bach's view and to add different criteria to the causal-historical link in order to have a singular thought. A causal connection to the object is necessary, but there must be additional criteria to answer the problem of a singular thought. Let me propose that singular thought is determined by two constraints: (i) the (*external*) *acquaintance constraint*: singular thought includes a causal-historical relation to the object; and (ii) the (*internal*) *cognitive significance constraint*: the thinker has a cognitive capacity to deliberately encode an *indexical-iconic representation* into an *indexical-discursive representation*, which enables her to have a singular thought. This approach to singular thought is two-pronged because it includes both external and internal elements, yet it is broad in scope since it is not restricted to just certain intermediaries. To illustrate further, I argue that the information that the causal-historical link carries and our cognitive ability to properly¹⁴ unpack this information are two components that matter in order to have a singular thought.

I would like to analyze this a bit further. I propose to distinguish between two types of cases of extended acquaintance representations taking also into consideration problematic examples of extended acquaintance. In particular, I will briefly explain what I mean by *indexical-iconic representations* and *indexical-discursive representations*¹⁵.

¹⁴ What "properly" means is to be defined by what it takes to encode an *indexical-iconic representation* into an *indexical-discursive representation* in order to have a singular thought. Due to space considerations I do not cover this here.

¹⁵ The motivation for this distinction I owe to Fodor's (2007) distinction between *iconic* and *discursive* representations. I add to both the element of indexicality that is motivated by Peirce's (1967) distinction between indexes, icons and symbols. However, my view differs from theirs at several points, especially the ones that are relevant for transmitting singular thoughts. I cannot fully develop this

3.3. The indexical-iconic representation and the indexical-discursive representation

In my view, an *indexical-iconic representation* is constructed in our mind each time the thinker perceives something, either in the case of an Object-Mind (hereafter, O–M) relation, namely in direct perception of an object (Bach's perception-based *de re* thought), or in a Mind1–Mind2 (hereafter, M1–M2) relation, namely through perceiving the physical token which was made in communication (Bach's communication-based *de re* thought). In other words, the thinker's immediate sense perceptions – sight, hearing, smelling, tasting, touching or their combination – always produce indexical-iconic representations. Thus, most indexical-iconic representations depend on one's senses. What is available to the senses is the immediate perceptual information, gained in an O–M relation or a M1–M2 relation, which brings along a causal-informational link¹⁶. The thinker (or the receiver or the perceiver, in this case) has to dismantle this informational package received from her immediate perception. She can unpack this causal-informational link in the way that will be beneficial in order for her to have a singular thought about an object. Consequently, there are always families of intermediaries¹⁷ that are present in perception and that are responsible for producing the *indexical-iconic representations*. If there is a certain analogy between an O–M relation and a M1–M2 relation, then what Bach's communication chain presents in a M1–M2 relation can be seen as a label for just a part of causal-informational link between O and M2.

In my view, the indexical-iconic representations are non-propositional representations and have a non-conceptual content. They are produced in perception by those intermediaries included in the causal-informational link, which connect the thinker

to the object at the end of the causal-informational link¹⁸. Thus the hard question is whether the propositional content¹⁹ contained in the mental token of one thinker, i.e. the transmitter, can be transmitted to another, i.e. the receiver, via physical tokens. Is the propositional structure of a mental token lost once we use a different kind of physical token (i.e. other than one produced in language) in an attempt to transmit a singular thought? Can it be preserved only in the case of physical tokens such as spoken or written words? I doubt that. This further opens the problem of how to transmit the propositional content via any kind of physical token from one person (who already has gained the propositional content and has it as a certain mental token) to another who should gain it via communication. I believe that each time the thinker perceives some physical token and has an indexical-iconic representation she is in a position to derive a propositional content out of this token, i.e. to decode an indexical-discursive representation.

I want to argue that, apart from the acquaintance constraint, what makes it possible to have a singular thought (either immediately or transmitted to us in communication) is the possibility of encoding indexical-discursive representation, i.e. the cognitive significance constraint. Fodor (2007) characterizes discursive representations as conceptual (as opposed to iconic representations), but I would like to distinguish between fully conceptualized, i.e. discursive representations and non-fully conceptualized, i.e. indexical-discursive representations²⁰. For this rea-

distinction here, but I will indicate differences in what follows, accordingly.

¹⁶ What Bach calls a causal-historical link I would rather call a *causal-informational link* to emphasize the informational package it carries.

¹⁷ The intermediaries might be more or less remote, e.g. air, glasses, photography, film, map, microscopic or telescopes image, etc. However, I want to stress that although there are intermediaries present between our senses and the object on the end of the causal-informational chain, the perceivers are not always, and do not have to be, aware of all the intermediaries present.

¹⁸ Another related issue is the following: what determines that a certain thing is at the end of the causal-informational chain? I am inclined to think that, in fact, causal chains are often not linear and are not with 'the sharp ending'. They do not lead solely from the object-point A to the object-point B, but they often go further to the object-point C, D, etc., and (sometimes) also come back to the previous points. They might, but not necessary, go linearly backwards. We use the model of 'causal links' with sharp endings in order to help us schematize the connections, and we deliberately 'break' the causal-historical chain at a certain point in order to say that this is the object one is having singular thought about.

¹⁹ Here the assumption is that a propositional content is characteristic of singular thoughts. I argue in favor of this notion by introducing the indexical-discursive representations.

²⁰ Bach's conceptualized representations are also not like Fodor's (2007) – they are narrow-psychological properties. And adding to them an indexical element makes them not fully conceptualized. Those are characteristics of "mental" *proper names* – they are not fully conceptualized. Bach assumes, contrary to Burge

son, indexical-discursive representations are, in my view, not fully conceptualized since they are, apart from being symbols, partly indexical in the Peircean sense. Indexical-discursive representations can be encoded from the indexical-iconic representations the thinker has when he perceives (both in case of O–M1 relation, as well as M1–M2 relation). Once the indexical-discursive representation has been encoded, this then opens a mental file²¹. Plainly speaking, the thinker perceives something and, thus, he has an indexical-iconic representation. After that he can use his cognitive capacity to decode the information into indexical-discursive representation that has propositional content²². Once he does that, he is in a position to have a singular thought.

3.4. Why not cut out the middle man?

Why introduce the indexical-iconic representation when we already have indexical-discursive representation? I posit indexical-iconic representation as a middle step, one that ties together the object the singular thought is about and the indexical-discursive representation which, in fact, is a singular thought. To elaborate, the indexical-iconic representation is not a singular thought, yet it might lead to one. I agree with Acquaintance Theorists that one has to be in a certain relation for a singular thought to arise, but I also posit this middle step, the ontological entity between the object singular thought is about and the singular thought itself.

This entity, namely the indexical-iconic representation is not a singular thought (yet), although it opens the possibility that the transmitter can use other intermediaries other than just words in order to communicate her singular thought about the object she has, if any. However, her communication might

not always be successful, for what she does is not to transmit a singular thought in the strict sense of the word transmitting, but to enable the receiver to encode one. Communication of a singular thought is a complex process. Some of the crucial ingredients which it includes are the object singular thought is about, the transmitter and the receiver on the one hand, and on the other it involves the perception of sense-data, mental and physical token, indexical-iconic and indexical-discursive representation. Finally, it includes the causal connection that goes from the object to the receiver. I will expand my view on this in what follows.

It is worth noting that, according to my account, transmitting a singular thought means something else. Let me clarify the matter: In my view, singular thought is not transmitted in the way Bach and other Singularists argue for²³. Moreover, even when the external condition is satisfied and there is a causal connection from the object to the receiver, the failure of transmitting still might be at work if the internal condition is not satisfied, namely if the receiver is not cognitively capable to accept and decode an indexical-iconic representation into an indexical-discursive representation.

3.5. The nature of an indexical-iconic representation

Thus, an indexical-iconic representation is both external and internal in its nature. It is a step from external to internal. Singular thought about an object is always generated anew depending on, among the other things, the cognizer's abilities. However, although the singular proposition via indexical-discursive representation is created afresh, a difference between singular and descriptive thought still remains. For, in the case of the inception of a singular thought, there exists the causal-informational path and other elements that lead to its inception. In short, there is a certain sense-data that is responsible for an indexical-iconic representation, and in that sense we might say that, transitively, the object singular thought is about becomes closer to us²⁴.

(1977), that conceptual "content is a narrow-psychological property" and "*de re* beliefs have truth-conditions which are not determined solely by their narrow-psychological properties (...) since they are indexical..." (Bach 1982/84, 128). *De re* thoughts have also a non-conceptual element which is distinctive of *de re* beliefs. Bach uses the notion of thought/belief interchangeably. My notion of the latter differs but, for the purpose of this paper, let us suppose we are talking about thought.

²¹ Recanati (2013) writes extensively on mental files.

²² I believe that decoding can be extended to symbol systems other than ones of natural language. The thinker can do so via other semantic systems she has the cognitive capacity to encode.

²³ For this reason, I believe that the kind of transmitting they offer is an illusion since it somehow presupposes that the whole proposition is transmitted, whereas I will argue that the singular proposition is created anew.

²⁴ Again, we have to bear in mind that the object alone cannot come into one's mind, one can only grasp a certain aspect of it via its sense-data. Precisely because of that, I

By claiming that, we must bear in mind that my view is closer to Russell's sense-data theory than to those who claim that the relation to the object is more immediate and direct. Even though there is no direct relation to the object properly speaking, still one's thought is not merely descriptive. What fixes this problem is the acquaintance constraint I am committed to. It manages so for at least two reasons. First, this external constraint pulls my theory out of the possible methodological solipsism one might arrive at if she only follows the internal constraint²⁵. Second, it secures that singular thought is about the particular object that exists outside and the thinker thinks about. As a result, the indexical-iconic representations reconcile the internal and external condition for having a singular thought.

3.6. *Indexical-iconic without indexical-discursive representation*

From what I have written so far, it is also important to see if one sometimes has an indexical-iconic representation without having a correspondent indexical-discursive one. As already stated, this is mainly determined by the thinker along with her ability to decode the right discursive system²⁶. I claim that the communication-based singular thought is possible when both of the thinkers, namely the transmitter and the receiver, are able to successfully encode the right indexical-discursive representation, each on their own. What is then transmitted, if anything, is the physical token²⁷ which is eventually responsible for the indexical-iconic representation. We can say that while the thinkers are to a certain degree powerless in having

the indexical-iconic representations, they are in charge of encoding the indexical-discursive one, i.e. of having a singular thought. In other words, there is no real transmitting in the Bach's sense when it comes to communication-based singular thoughts.

In brief, in order for one to have a singular thought, apart from the causal-informational connection to the object one has singular thought about, the thinker must acquire two indexical representations: (i) indexical-iconic representation and (ii) indexical-discursive representation. As I argued above, the indexical-iconic representation is a necessary, but not a sufficient representation one has to have in order to have a singular thought. In other words, the indexical-iconic representation which does not lead to the indexical-discursive representation is not enough for one to have a singular thought. Indexical-iconic representation can lead to indexical-discursive representation when the thinker's cognitive capacity is fulfilled. It is fulfilled when the thinker is capable of decoding the indexical-iconic representation into an indexical-discursive representation²⁸.

Next, I want to look at how my proposed account including these two kinds of representations reflects on the standard view which distinguishes three standard kinds of singular thought, i.e. the perception-based, the memory-based, and the communication-based singular thoughts, with special emphasis on the third kind, namely the communication-based singular thought.

3.7. *The inception of a communication-based singular thought*

In my view, in the case of communication-based singular thought the transmitting of a singular thought is not done in the ordinary Bachian manner since singular thought is not transmitted, but generated anew. The transmitter's intention is relevant since it motivates her to produce a physical token out of her mental token that the receiver then perceives. Moreover, I believe that, on the one hand, singular thought is always perception-based since it is derived from perception, yet on the other it is not a form of perception-based singular thought in the Bachian sense. More precisely, there is an object the transmitter perceives and its sense-data produces an

am inclined to think that this is the only way one can have an object in her mind.

²⁵ As already noticed by Descartes, our representations might not resemble the object they represent.

²⁶ I am aware that I owe an account of how these (indexical) discursive systems are generated and what it means that two thinkers share the same (indexical) discursive system.

I also stipulate here that these systems can be learned and recognized by a cognizer. Of course, it might happen that one does not, in the end, recognize or encode a certain physical token into an indexical-discursive representation whereas another one does. Depending on the cognizers and the physical tokens, some of them are more suitable and they are in a position to encode them more easily. However, I do not discuss the problem of encoding here.

²⁷ It is transmitted along with the causal connection to the object singular thought is about.

²⁸ The indexical-discursive representations might include different propositional systems. Due to space considerations I do not cover this here.

indexical-iconic representation in her mind which she as a cognitive capable thinker encodes into an indexical-discursive one and memorizes it. After that the transmitter, who has acquired a singular thought about that particular object, intentionally uses a physical token out of her mental token of that particular object. However, this physical token is another sense-data the receiver has to deal with. Similarly, the receiver perceives the sense-data of a physical token which produces an indexical-iconic representation in the receiver's mind which she, again, as a cognitive capable thinker encodes into an indexical-discursive one. Apart from this, there has to be a causal-informational connection originating from the object via the transmitter to the final thinker.

A further question regarding sense-data comes to mind. How is it possible that both the transmitter and the receiver have a singular thought if the sense-data the receiver perceives might be completely different from the sense-data the transmitter had about that particular object? My very tentative solution is that the singular thought about a particular object rests on the notion of a certain indexical-discursive system singular thought is based on. I am inclined to think that a certain system can be encoded out of sense-data. Take, for example, the natural language system. Language is a generated system connected with certain sense-data. One can learn to recognize and transform this sense-data into spoken or written words. I believe that cognitively capable thinkers can also learn certain already generated discursive systems which refer to certain classes of objects or entities (general thoughts or propositions) to which then, in certain circumstances suitable for a singular thought, the element of indexicality is added (singular thoughts or propositions).

4. Conclusions concerning a two pronged approach

What is the novelty of this approach to singular thought? I offer two important roles. The first is a *restrictive role*: I believe that introducing indexical-iconic representation is a necessary element which can resolve the dilemma whether any physical token whatsoever can be taken as a means of transmitting a singular thought. And my answer is that not any physical token will do for in order to have a singular thought one has to have the right indexical-

discursive representation originating from an indexical-iconic one. The second is an *explanatory role*: this account enables the possibility of having a singular thought in cases of remote acquaintance which, despite our strong intuitions, other Singularists, in particular, and less restrictive Acquaintance Theorists, unsuccessfully strive to explain²⁹. In my opinion, they are struggling with the dilemma of how we can argue for the transmission of a singular thought via the physical token of words (either spoken or written of which one perceives the sense-data) and not of the other physical tokens.

To conclude, we should not lose sight of the fact that the two pronged approach to singular thought opens the possibility of a remote acquaintance under new terms of grasping what the acquaintance to the remote object is. Acquaintance is an external condition, given both in terms of a causal connection and the Russellian sense-data theory³⁰. However, in order for a thinker to have a singular thought about an object the internal condition in terms of indexical-iconic and indexical-discursive representation is needed. While the former condition externally fixes the *singularity* and the reference to the particular object singular thought is supposed to be about, the latter internally gives the *thought-likeness* to the singular thought.

References

- Austin, J. L. (1962) *Sense and Sensibilia*. Oxford: Oxford University Press.
- Bach, K. (1982/84) *De re* Belief and Methodological Solipsism. In *Thought and Object*, ed. A. Woodfield, 121–151. Oxford: Oxford University Press.
- Bach, K. (2010) Getting a Thing into a Thought. In *New Essays on Singular Thought*, ed. R. Jeshion, 39–63. Oxford: Oxford University Press.
- Bach, K. (1987/94) *Thought and Reference*. Oxford: Oxford University Press.

²⁹ For example, the ones who would like to allow the possibility of having a singular thought via photographs, audio and video recordings, etc.

³⁰ Even though my account relies on sense-data theory, I think that it does not exclude or it is at least compatible with the fact that a singular thought is about the object.

- Fuš, M. Forthcoming. *Bach's Constraint on Extending Acquaintance: Some Questions and a Modest Proposal*. Croatian Journal of Philosophy.
- Burge, T. (1977) *De re* Belief. *Journal of Philosophy* 74: 338–62.
- Broad, C. D. (1965) The Theory of Sensa. In *Perceiving, Sensing, Knowing: A Book of Readings in the Philosophy of Perception*, ed. R. Swartz. Garden City, NY: Doubleday.
- Chisholm, R. (1942) The Problem of the Speckled Hen. *Mind* 51: 368–373.
- Davies, M. (1997) Externalism and Experience. In *The Nature of Consciousness*, eds. N. Block, O. Flanagan, and G. Guzeldere. Cambridge, MA: MIT Press.
- Donnellan, K. (1972) Proper Names and Identifying Descriptions. In *Semantics of Natural Language*, eds. D. Davidson and G. Harman. Dordrecht: Reidel.
- Donnellan, K. (1979) The Contingent A Priori and Rigid Designators. In *Contemporary Perspectives in the Philosophy of Language*, eds. P. French, T. Uehling, and H. Wettstein. Minneapolis: University of Minnesota Press.
- Evans, G. (1982) *The Varieties of Reference*. Oxford: Oxford University Press.
- Fodor, J. (2007) The revenge of the given. In *Contemporary Debates in Philosophy of Mind*, eds. B. P. McLaughlin and J. Cohen, 105–116. Oxford: Blackwell.
- Frege, G. (1984/1892) On Sense and Reference. In *Collected Papers on Mathematics, Logic and Philosophy* McGuinness, eds. and trans. Black, et al. Oxford: Blackwell.
- Frege, G. (1984/1918) The Thought. In *Collected Papers on Mathematics, Logic and Philosophy* McGuinness, eds. and trans. Black, et al. Oxford: Blackwell.
- Jeshion, R. (2002) Acquaintances *De Re* Belief. In *Meaning and Truth: Investigations in Philosophical Semantics*, eds. Joseph Keim Campbell, Michael O'Rourke & David Shier. Seven Bridges Press.
- Jeshion, R. (2010) *New Essays on Singular Thought*. Oxford: Oxford University Press.
- Jeshion, R. (2010) Singular Thought: Acquaintance, Semantic Instrumentalism, and Cognitivism. In *New Essays on Singular Thought*, ed. R. Jeshion, 105–140. Oxford: Oxford University Press.
- Jeshion, R. (2009) The Significance of Names. *Mind and Language* 24 (4): 370–403.
- Kaplan, D. (1968) Quantifying In. *Synthese* 19: 178–214.
- Kaplan, D. (1989a) Demonstratives. In *Themes from Kaplan*, eds. J. Almog, H. Wettstein, and J. Perry. New York: Oxford University Press, 481–563.
- Kaplan, D. (1989b) Afterthoughts. In *Themes from Kaplan*, eds. J. Almog, H. Wettstein, and J. Perry. New York: Oxford University Press.
- Kaplan, D. (1979) Dthat. In *Contemporary Perspectives in the Philosophy of Language*, eds. P. French, T. Uehling, and H. Wettstein. Minneapolis: University of Minnesota Press.
- Kripke, S. 1980. *Naming and Necessity*. Oxford: Blackwell.
- Lawlor, K. (2001) *New Thoughts about Old Things: Cognitive Policies as the Ground of Singular Concepts*. New York: Garland Publishing.
- Lopes, D. (1994) *Understading Pictures*. Oxford: Oxford University Press.
- Marcus, R. (1995) *Modalities: Philosophical Essays*. Oxford: Oxford University Press.
- McGinn, C. (1982) *The Character of Mind*. Oxford: Oxford University Press.
- Millikan, R. (2000) *On Clear and Confused Ideas: An Essay about Substance Concepts*. Cambridge: Cambridge University Press.
- Moore, G. E. (1993) Sense Data. In *Selected Writings*, ed. T. Baldwin. London: Routledge.
- Peirce, C. S. (1967) *Exact Logic*. Collected papers, vol. 3 Third printing Cambridge Mass.: The Belknap Press of Harvard University Press.
- Price, H. (1932) *Perception*. London: Methuen.
- Perry, J. (2000) *The Problem of the Essential Indexical and Other Essays*. Stanford: CSLI Publications.
- Perry, J. (2001) *Reference and Reflexivity*. Stanford: CSLI Publications.
- Putnam, H. (1975) *Philosophical Papers II: Mind, Language and Reality*. Cambridge: Cambridge University Press.
- Pylyshyn, Z. (2007) *Things and Places: How the Mind Connects with the World*. Cambridge, MA: MIT Press.
- Pylyshyn, Z. (2003) *Seeing and Visualizing*. Cambridge, MA: MIT Press.
- Recanati, F. (1993/1997) *Direct reference: From Language to Thought*. Oxford: Blackwell.
- Recanati, F. (2013) *Mental Files*. Oxford: Oxford University Press.

-
- Russell, B. (1912) Reset 1946. *The Problems of Philosophy*. Oxford: Oxford University Press.
- Sainsbury, M. (2005) *Referents without Referents*. Oxford: Calderon Press.
- Scholl, B. (2002) *Objects and Attention*. Cambridge, MA: MIT Press.
- Scholl, B. and Z. Pylyshyn. (1999) Tracking Multiple Items Through Occlusion: Clues to Visual Objecthood. *Cognitive Psychology* 80: 159–77.
- Scholl, B., Pylyshyn, Z., and J. Feldman. (2002) What is a Visual Objects? Evidence from Target Merging in Multiple Object Tracking. Reprinted in *Objects and Attention*, ed. B. Scholl. Cambridge, MA: MIT Press.
- Taylor, K. (2003) *Reference and the Rational Mind*. Chicago: Chicago University Press.
- Tye, M. (2000) *Consciousness, Color, and Content*. Cambridge, MA: MIT Press.
- Walton, K. (1984) Transparent pictures: On the Nature of Photographic Realism. *Critical Inquiry* 11: 246–276.
- Zeimbekis, J. (2009) Pictures and Singular Thought. *Journal of Aesthetics and Art Criticism* 68: 11–21.

SCEPTICAL THEISM AND THE PROBLEM OF EPISTEMIC EVIL: WHY SCEPTICAL THEISM IS PHILOSOPHICALLY COSTLY

Jimmy Alfonso Licon

San Francisco State University, USA
e-mail: jimmylicon01@gmail.com

Abstract:

Sceptical theism is supposed, by a number of philosophers, to undercut the evidential basis for the evidential problem of evil. In this paper, I argue that even if sceptical theism succeeds, its success comes with a hefty epistemic price: *it threatens to undermine a good deal of what we supposedly know*. Call this the *problem of epistemic evil*. Thus, sceptical theism has a costly philosophical price of admission. In light of this, it seems that the evidential problem of evil is harder to dislodge than it might have initially seemed; i.e. with sceptical theism, we trade the evidential problem of evil for the problem of epistemic evil.

Key words: problem of evil, suffering, appearances, epistemic evil, defeater.

I. Introduction

The philosophical basis for the evidential problem of evil amounts to this: it appears that there are instances of gratuitous suffering; as such, there is good reason to suppose that no individual who deserves the honorary title 'God'¹ exists. This is because God, as a perfectly loving, omni-being would have the ability and motivation to eliminate gratuitous evil. However, these evils remain. Thus, there is probably no such individual who deserves the honorary title 'God' (Rowe 1979, Draper 1989). Apparent gratuitous suffering is supposed to be *some* evi-

dence that there *are* instances of gratuitous suffering; even though suffering *tout court* is compatible with the existence of God, *gratuitous* suffering is not. In response to this challenge, from apparent gratuitous suffering, theists have tried to find a way to undercut the evidential inference from it appearing that there are instances of gratuitous suffering to the conclusion that *probably* there are such instances. If the evidential link can be severed, the threat to theism is significantly weakened.

A prominent attempt to undermine the evidential basis for the problem of evil is known as *sceptical theism* (cf. Alston 1991, Bergmann 2001). Roughly put, sceptical theism says that the appearance of gratuitous suffering is, at best, a questionable epistemic guide to the nature of suffering. This is so merely in virtue of the wide epistemic and cognitive gap that exists between humans and the individual who deserves the honorary title 'God.' After all, just as it might seem to a dog that veterinary visits are unnecessarily cruel because of her ignorance and cognitive impoverishment, (even though she lacks the epistemic and cognitive resources to make such an inference – but I'm ignoring that for the sake of the paper), likewise, there is a similar, if not greater, epistemic and cognitive gap between humans and God in the following sense: God could have any number of reasons, moral and otherwise, beyond our ken, to allow the amounts and kinds of suffering that God does. Thus, inferences from seeming-to-be-gratuitous to the claim that there is *actual* gratuitous suffering appear epistemically weak. After all, God might have reasons to allow suffering, *which only appears gratuitous*, which are beyond our ken. Thus, it should be no surprise that some suffering *appears to be* gratuitous, even if it isn't.

¹ For instance, McBrayer (2010, 611) writes: "The metaphysical thesis is that God exists, where 'God' is an honorific title describing the most perfect being possible." The intuition at work is this: the most perfect being possible would not allow gratuitous suffering; if you think that G is the most perfect being possible, but G would allow *even a bit* of gratuitous suffering, then there is a greater being, G*, which does not even allow *that much* gratuitous suffering.

In this paper, I propose a dilemma, with serious epistemic consequences, for those who rely upon the sceptical theist strategy as a way to undercut the basis for the evidential problem of evil. In a nutshell, the dilemma says that either:

- (a) Despite the epistemic and cognitive gap which exists between humans and anyone who deserves the honorary title of 'God,' *apparent* gratuitous suffering is good evidence that there is gratuitous suffering, and thus, there is *some* reason to reject theism;
- (b) *Apparent* gratuitous suffering is *not* good evidence that there is gratuitous suffering, and as such, the epistemic and cognitive gap that exists between humans and God need not be contained solely to matters of suffering and their supposed moral justification; God could have good reasons for any number of things having little to do with suffering.

Thus, the epistemic implications of sceptical theism bleed into that which we take ourselves to know. Simply put, it might turn out, for all of our evidence, that God could have reasons, beyond our ken, to massively deceive us, e.g. perhaps, unbeknownst to us, it will aid in our moral development. However, it seems plausible to suppose that there are plenty of things we know. Thus, there is something wrong with our starting assumptions.

The problem, particularly potent under sceptical theism, amounts to this:

The Problem of Epistemic Evil

Just like it might *seem as though* God is incompatible with the presence of suffering, in light of God's perfect benevolence, although there *might be* reasons which morally justify, if not morally demand, that God allow various kinds of suffering, by a similar line of reasoning, it may be that God allows us to suffer from epistemic deficiency, if not outright deceiving us that such-and-such is so, for moral reasons beyond our ken. If suffering *tout court* is compatible with the truth of theism, then so is God deceiving us that such-and-such for similar reasons, even if they are beyond our ken. Put differently, although sceptical theism serves to undermine the inference that *there are* instances of gratuitous suffering, it lays the philosophical

groundwork for a virulent strain of theologically-based scepticism.

The layout of this paper is simple. In the next section, I explain how it is that sceptical theism undercuts the justification for the evidential problem of evil, in greater detail. In the section after that, I explain why it is that that undercutting the justification for the evidential problem of evil comes at a serious philosophical cost *by the sceptical theist's own lights*.

II

If you recall, the evidential problem of evil is often supported by appeals to suffering which *appears* to lack moral justification (Rowe 1979, 337). For instance, it seems that a small infant, who is slowly dying of a stomach cancer, does little, if any good, for anybody; that is, rather than building the moral character and resolve of the parents, it breaks them; rather than helping the child morally develop later in life, it deprives her of the opportunity. It *seems* that the infant dying from stomach cancer is, at a minimum, a plausible instance of gratuitous suffering. If we assume that appearances are a good epistemic guide to how things are, there is good reason to think that there are instances of gratuitous suffering; thus, we have a challenge to the plausibility of theism. After all, we think that appearances, as an epistemic kind, confer *defeasible* justification. Surely, this includes how things appear morally speaking, e.g. it appears morally wrong to light a puppy on fire, *merely* for giggles; and, in virtue of this seeming to be wrong, I have *some* reason to think that it *is* wrong.

For instance, Tucker writes:

It is natural to think that many of our beliefs are rational because they are based on seemings, or on the way things seem. This is especially clear in the case of perception. You believe that there is a paper document (or a computer screen) in front of you because it seems visually that way ... I believe that I ate cereal for breakfast because I seem to remember eating it for breakfast. And we believe that torturing for fun is morally wrong and that $2+2=4$ because those claims *seem intuitively obvious* (forthcoming).

Swinburne agrees:

Now it is a basic principle of rationality, which I call the principle of credulity, that we ought to believe that things are as they seem to be (in the epistemic sense) unless and until we have evidence that we are mistaken ... If it seems to me that I am seeing a table or hearing my friend's voice, I ought to believe this until evidence appears that I have been deceived. If you say the contrary—never trust appearances until it is proved that they are reliable—you will never have any beliefs at all^(2010, 115).

In response, sceptical theists argue that although appearances are usually a good epistemic guide to how things are, this does not apply *specifically* to what God would do in such-and-such situation; there are context dependent defeaters, *when talking about God's motives*. After all, any individual worthy of the title 'God' will have the power and knowledge to do whatever God likes, which is also morally justified; it is plausible to suppose that in many instances, the relevant moral justification will be beyond our ken.

Sceptical theists accept the following (cf. McBrayer 2010, 611):

Humans are incapable of making justified, *all-things-considered* judgments as to what God would do in any particular situation; e.g. I cannot be sure, *all-things-considered*, that God didn't create me, five minutes ago, with apparent memories extending beyond that. Call this "*Epistemic Evil*"².

There is supposed to be *some* reason to accept that "*Epistemic Evil*", merely in light of the epistemic and cognitive gap which must exist between us and God; this gap is supposed to, at the very least, render "*Epistemic Evil*" plausible. After all, a being who is worthy of the title 'God' would have an awesome degree of power and knowledge that vastly exceeds ours. For instance, there could be any number of reasons that morally justify allowing suffering that appears gratuitous, in spite of the moral justification for allowing it. Crudely sketched, if there is an instance of suffering that *appears to be* gratuitous, but there is nevertheless good moral reason for

allowing it, God would be far more likely, *ceteris paribus*, to know about it than we would.

Consider the following example:

Suppose that Sam, the dog, has to go to vet for shots and a flea bath. As far as Sam can tell, there is no reason that could morally justify being subjected to such treatment; indeed, as far as Sam knows, trips to the vet often result in suffering. Matt, Sam's owner, knows exactly what justifies subjecting Sam to such trips: *it is in Sam's best interests to be vaccinated*. Thus, Sam's inability to understand Matt's reasons for taking him to the vet, and subjecting him to what cause Sam a good deal of suffering, does not indicate that there are no good reasons; that is, Sam would be ignorant of Matt's reasons regardless of whether or not Matt actually had any.

Putting aside that Sam is a dog, it would be epistemically irresponsible if Sam made *all-things-considered* judgments about Matt, given his paltry epistemic and cognitive resources, compared to Matt's. By a parallel line of reasoning, it would be epistemically irresponsible, to say the least, for us to make *all-things-considered* judgments about what God would do, under such-and-such conditions, given the epistemic and cognitive gap between us.

III

Just like God could allow apparently gratuitous suffering, without doing anything wrong, the same applies to God deceiving us, or allowing us to be placed in a situation that breeds ignorance; in other words, it could be that God has moral reasons, for instance, beyond our ken, to massively deceive us, for the greater good (or to prevent a greater evil). Clearly, atheists and theists accept that instances of suffering *tout court* do not provide evidence against theism. Thus, it could be that there is plenty of suffering in the world which is justified by moral reasons *beyond our ken*. As such, in the final analysis, such suffering may be compatible with theism.

There is a principle used to defend the theistic contention against disconfirmation by the appearance of gratuitous suffering; crudely sketched, the principle amounts to this:

It seems to be that x and this is incompatible with y; but, there is context, C, which renders x and y compatible. We cannot rule out C. Thus, x on its face is not reason to think that not-Y.

² Of course, theists could reject this way of cashing out the epistemic component of sceptical theism; however, they would have the burden to explain why the offered version is unacceptable, and why their version is better able to avoid the problem of epistemic evil.

Theists use this pattern of reasoning in response to the appearance of gratuitous suffering that seems, on its face, to disconfirm the theistic contention: if there is an epistemic and cognitive gap between humans and God, all bets are off with respect to apparent gratuitous suffering. By a parallel line of reasoning, something similar could be said for our epistemic inventory: although it seems that there are plenty of things we know (e.g. what we have for breakfast), it might turn out, in the final analysis, we are extremely ignorant. For one thing, it could be that God has reasons, beyond our ken, to deceive us massively. It is not clear how a sceptical theist, at least in principle, could deny that this is a serious threat.

The connection between sceptical theism, on the one hand, and general scepticism, on the other, might not be immediately obvious. But, if there is an individual who deserves the honorary title ‘God,’ this individual has the *ability* to massively deceive us. Further, humans cannot make justified *all-things-considered* judgments as to what God would choose to do with respect to our epistemic situation, i.e. this is similar to the claim that God could allow all sorts of suffering, for moral reasons, beyond our ken. Just as proponents of the problem of evil cannot rely on the *appearance* of gratuitous suffering to infer that there are instances of gratuitous suffering, in the face of sceptical theism, by a similar line of reasoning, epistemic realists cannot rely on appearances as a guide to how things are, as this would presume that God does *not*, all-things-considered, have reasons that are beyond our ken to deceive us. We cannot make justified *all-things-considered* judgments about our epistemic situation *if we accept sceptical theism*. If I cannot make justified *all-things-considered* judgments as to what God would do in this or that situation, then I cannot make a justified *all-things-considered* judgment as to what God would do with respect to my epistemic situation; this follows necessarily. For instance, it could be that God wants me to believe such-and-such, although such-and-such is false, because of reasons beyond my ken; e.g. it may turn out, if I believe falsely, then this will provide me an opportunity to develop moral character.

In a nutshell, just as the epistemic and cognitive gap that exists between us and God prevents us from making justified *all-things-considered* judgments as to what God would do vis-à-vis any instance of suffering merely in virtue of how ignorant we are, the same holds for epistemic situations: merely in virtue

of the epistemic and cognitive chasm that exists between us and God we cannot be sure that we are as epistemically well-off as we think.

Approach this as an argument. Call it *the Fooling-You Argument*:

1. If you had *all-things-considered* knowledge as to your epistemic situation³, God could not fool you with epistemic gerrymandering.
 2. But it *could be* that God is fooling you with epistemic gerrymandering.
- Thus,
3. You lack *all-things-considered* knowledge as to your epistemic situation⁴.

If we accept sceptical theism, it is not clear *where there is room*, philosophically, to deny the first or second premises. If we are not justified in making *all-things-considered* judgments as to what God would do, vis-à-vis our epistemic situation, it is not clear how we could rule out that God is fooling us. So, if the sceptical theist appealed to *prima facie* evidence that God would not do such a thing – perhaps, for instance, she has had religious experiences that she takes to be a good reason to reject the problem of epistemic evil – she would still not have a good reason to disregard the problem of epistemic evil *for the same reason* that she, qua sceptical theist, supposes that the *appearance* of gratuitous evil is no reason, on its own, to doubt the viability of theism.

Suppose that you took the evil demon hypothesis seriously. This is the claim that there is an evil demon that is responsible for your perceptual experiences. You conclude that there is little you can be said to know in the following sense: for anything you take yourself to *know*, you have a defeater for that epistemic candidate. The defeater is that things could seem to be such-and-such, even though this is not at all how they are. Surely, on the basis of your acceptance of the evil demon hypothesis, and your

³ It might turn out that you do not have the kind of knowledge that you would need in order to justifiably rule out the mere epistemic possibility of falling victim to a sceptical scenario.

⁴ If we want to rule out that God engages in massive trickery, such that all of our beliefs are false, we would need to know *all-things-considered* what God would do in *this* situation. However, sceptical theists hold that we cannot know this. Thus, sceptical theists cannot rule this out.

recognition of its ability to defeat the epistemic clout had by your perceptual experiences, there is little that you can be said to *unproblematically know* via your perceptual experiences. For any of those experiences, it could be that things only *appear* to be such-and-such, even though they are not such-and-such, *in light of your doxastic commitment to the evil demon hypothesis*. If the evil demon is inscrutable enough – as he surely is, right along with God – with the ability to massively deceive you, then it is hard to see how you could secure knowledge, rather than defeated beliefs, via your perceptual experiences. A similar point applies to those who accept sceptical theism as a way to undermine the evidential problem of evil.

IV

Before concluding, I want to deal with a serious challenge. There might be those who, just like Descartes, for instance, doubt that God would engage in such epistemic gerrymandering such that many, if not most, of our beliefs were false, because God is not a deceiver, by God's very nature. After all, if God is omni-benevolent, then by God's very nature, it is supposed that God cannot engage in something like deception, at least if we accept that deception is a kind of harm to those who fall under its spell.

Consider what Descartes says:

God [...] who possesses all those supreme perfections of which our mind may indeed have some idea but without understanding them all, who is liable to no errors or defect ... From this it is manifest that He cannot be a deceiver, since [...] fraud and deception necessarily proceed from some defect (1993, 72).

Plantinga agrees:

Does anything similar [epistemically speaking] apply to the Christian theist? [i.e. God engaging in deception] I think not. A Christian accepts theism, of course, but also specifically Christian beliefs that distinguish Christians from other theists. Now if Christian belief is true, it is eminently likely that the belief-producing processes that give rise to it are in fact reliable. According to the central teaching of Christianity, God has devised a way in which we human beings, despite our sinful condition into which we have fallen, can achieve a right relationship with God; this scheme involves the life, suf-

fering, death and resurrection of Jesus Christ, the second person of the trinity. *Now God would want us human beings to know about this scheme*, in particular because its efficacy involves, at least in the paradigmatic and central cases, *belief* in it. Thus, it is likely, given the essentials of Christian belief, that God would furnish a way in which human beings can know about this salvific scheme, and hence that belief-producing processes from which Christian belief emanates are indeed reliable (2002, 266 – emphasis mine).

The claim that God would not massively mislead us is susceptible to the following trilemma:

This claim is either a mere epistemic possibility, *prima facie*, or *all-things-considered* justified; but, none of these options is of much help to the sceptical theist when it comes to avoiding the sceptical consequences of their commitment to sceptical theism.

First, suppose that the God-deceiving-us suggestion is merely epistemically possible, i.e. for all we know, it could be that *p*. This is not much help to the sceptical theist. For the same reason that the *mere epistemic possibility* of sceptical theism does not undermine the evidential heft of apparent gratuitous suffering as a reason to reject theism (e.g. the mere epistemic possibility of an evil demon does not undermine the epistemic confidence I place in my perceptual faculties) the mere epistemic possibility that God is deceiving us does little to undermine the wide ranging epistemic consequences of accepting sceptical theism.

Second, suppose the claim that God does not deceive is *prima facie* justified; this is not a response to the problem of epistemic evil, just like the *appearance* of gratuitous evil does not block the sceptical theist response. The epistemic situation is similar in each case; that is, just like the sceptical theist response undercuts the evidential clout that would otherwise be enjoyed by the appearance of gratuitous suffering, so the possibility that God is messing with our epistemic situation for reasons beyond our ken, that we cannot rule out, undercuts the claim that God would not do such a thing. After all, on its face, it might seem as though God would not permit suffering either, but we realize that in the final analysis, it might be that the suffering God allows is conducive to bringing about greater goods, or preventing

worse evils; the same can be said for God engaging in massive deception.

Third, suppose that the claim is *all-things-considered* justified (i.e. there is nothing which could come along and defeat our justification). This would mean that sceptical theism, at least with respect to its epistemic component, is false. This is because, *by the sceptical theist's own lights*, we cannot make justified *all-things-considered* judgments as to what God would do in *any* particular situation, including the judgment that God would not massively deceive us. There is a similar line of reasoning for the claim that God *could* have reasons, beyond our ken, for massively deceiving us, just as there is for the claim that we cannot infer that there are instances of gratuitous suffering merely on the basis of such appearances.

Finally, Plantinga's doctrinal appeal has problems of its own:

1. Even if we assume that Plantinga rejects sceptical theism, his response *appears* to be in tension with sceptical theist commitments. These commitments, if taken on by the sceptical theist, result in a serious difficulty in making authoritative claims as to what God would do in this-or-that situation, at least in an *all-things-considered* sense.

2. When Plantinga claims that, '[now] God would want us human beings to *know* about this scheme, in particular because its efficacy involves, at least in the paradigmatic and central cases, *belief* in it' (266, emphasis mine), this suggestion falls prey to the trilemma I mentioned previously. In light of the problems already discussed, this sort of move is of little help to the sceptical theist. Further, his suggestion appears to assume that the Christian faith is what God had in mind *all-things-considered*. The epistemic status of this claim is tainted by the possibility – which we would have little reason to disregard, as sceptical theists – that God could massively trick us for *reasons beyond our ken*. There is no reason to suppose, without special pleading, that Christian tenets are exempt from the epistemic threat that sceptical theism poses *for those who accept sceptical theism*.

3. Even though it might seem that sceptical theism effectively undercuts the evidential basis for the problem of evil, this approach to the problem of evil is secured at a steep epistemic price: *it serves to cast a sceptical shadow over just about everything we take ourselves to know*. Thus, sceptical theism might not be the best way for theists to deal with the problem of evil.

V

The worry I have been defending, throughout the paper, amounts to this: if we cannot infer that *there really are* instances of gratuitous suffering, on the basis of how things appear, because of sceptical theist considerations, we cannot conclude that *we really have knowledge* on the basis of how things appear, for similar reasons. Put differently, if sceptical theism has a defeating quality when it comes to inferring that there are instances of *apparent* gratuitous suffering, then it has the same quality when it comes to other kinds of inferences, *where God may have an intervening role*. In those cases, sceptical theism serves as a kind of undercutting defeater. However, appearances, as a kind, are a prominent source of knowledge as to how the world hangs together. As a result, we are left with the choice of either accepting sceptical theism as a way to undermine the evidential problem of evil, in which case we must contend with the problem of epistemic evil, or, we must throw out sceptical theism as a way to avoid the epistemic problem of evil, in which case we must deal with the evidential problem of evil. Either way, this is a difficult position for theists to be in.

References

- Alston, W. (1991) The inductive argument from evil and the human cognitive condition. *Philosophical Perspectives* 5:29–67
- Bergmann, M. (2001) Sceptical theism and Rowe's new evidential argument from evil. *Noûs* 35 (2): 278–296
- Descartes, R. (1993) *Meditations on First Philosophy in Focus*. S. Tweyman (ed.) Routledge
- Draper, P. (1989) Pain and pleasure: An evidential problem for theists. *Noûs* 23 (3): 331–350
- McBrayer, J. (2010) Sceptical theism. *Philosophy Compass* 5 (7): 611–623
- Plantinga, A. (2002) Reply to Beilby's Cohorts. In Beilby, J (ed.) *Naturalism Defeated?* Cornell University Press: 204–276
- Rowe, W. (1979) The problem of evil and some varieties of atheism. *American Philosophical Quarterly* 16 (4): 335–341
- Swinburne, R. (2010) *Is There a God?* Oxford University Press
- Tucker, C. (forthcoming). Seeming and Justification: Introduction. In Tucker, C. (ed.) *Seeming and Justification*. Oxford University Press

CHOMSKY AND FOUCAULT ON HUMAN NATURE – A PERSPECTIVE FOR RECONCILIATION

Ana Smokrović

Philosophy Department, Rijeka
e-mail: asmokrovic@gmail.com

Abstract:

Chomsky and Foucault are closer to each other in their views than it is often assumed. This paper focuses on the “Chomsky-Foucault debate”, and in particular on the most philosophical topic discussed in the book, namely the question of whether there is a human nature. This paper argues that the contrast is more one of focus and stress than one of deep philosophical disagreement. This fact is of interest for the present day continental-analytic debate, given the importance of the two authors for the respective schools of thought. A constructive approach should look for bridges and enhance the dialogue of the two schools; Chomsky and Foucault are excellent authors in this regard, both as role models and as foci of inquiry.

Key words: human nature, post-structuralism, knowledge, innateness, justice.

How close are Chomsky and Foucault in their views and recommendations?¹ I want to argue that they are closer to each other than it is often assumed. I found them to be very similar in their political motivation, although they are divergent in their characterization of some of the intellectual means for implementing it. The issue is important for philosophy and “theory”, since it concerns the views of the central post-structuralist author, Foucault, and the views of the thinker who has been inspiring analytic philosophers for half a century. In this paper I want to focus on “Chomsky-Foucault debate”², and in particular on the most philosophical topic dis-

cussed, namely the question whether there is a human nature. *The Chomsky-Foucault Debate on Human Nature* triggered my interest in the work of both Chomsky and Foucault some years ago, because in it these two influential thinkers bring up many questions concerning the existence of human nature, knowledge, the possibility of creating a scientific theory, the question of justice, civil disobedience, and the idea of an alternative future society.

In the first part of the debate, which is our philosophical topic here, Foucault and Chomsky debate the existence of human nature and notions of knowledge and science. While “human nature” for Chomsky stands for innate structures of the human mind which guide social and intellectual behaviour, for Foucault human nature is a mere concept created within our civilization. “Knowledge” to Chomsky represents human’s cognitive precognitions or innate structures for creating knowledge and shaping scientific theories, while for Foucault “knowledge” stands for the organization of a particular body of claims in a particular historical moment; he addresses the cultural conditions surrounding a particular body of claims in a particular historical period. What is interesting to me is that in the debate they are not opposed to one another but simply observing particular concepts from different perspectives: the concept might be the same, but conceptions differ. (Regarding the second, political element of the debate, Chomsky is focused on the creation of an alternative society based on human nature that will ensure human freedom. However, in this part of the debate Foucault is restrained and maintains the position of the present historical moment and focuses on disclosure of political power which controls and shapes the social body. Regarding the question of resistance present in an act of civil disobedience, these two intellectuals disagree in theory. For Chomsky civil

¹ I wish to thank Anna Loutfi and David Weberman from Central European University, Budapest, for their help with the manuscript from which the present paper derives.

² See Chomsky, Noam, and Foucault, Michel 2006. *The Chomsky-Foucault debate on Human Nature*. NY: The New Press.

disobedience is necessary in order to put pressure on the government, whilst for Foucault it is just an act within the state framework, so it is not so meaningful. On the question of “justice” they also disagree inasmuch as, “justice” for Chomsky is an essential characteristic of human nature and any future society that should be built on it, while for Foucault it is, again, just a concept created within our civilization; although in his critique of power which creates inequalities, it seems that he has a strong sense of justice. Finally, while Foucault is really careful regarding his vision of future society, Chomsky is optimistic concerning the creation of a better society in which human freedom will be imperative; there is, for Chomsky, a better world that humans should aspire to.)

The Chomsky-Foucault Debate on Human Nature

In *The Chomsky-Foucault Debate on Human Nature* Chomsky and Foucault can be seen as belonging to two different traditions within the left. Chomsky agrees with the anarchist tradition which argues for the existence of a universal human nature and hopes for a society that can be built upon it. Foucault argues against the existence of human nature and views ideas concerning justice or human nature as constructivist, claiming that they are just cultural constructs that can be deconstructed. Within this debate of such intellectual depth, they show in which different directions the Left can go in its political thinking. The debate itself was set up in a tense political moment during the seventies; a moment characterized by a pronounced reaction against the establishment. The questions raised in the debate are prominent again in today's historical moment which is also high-strung and characterized by a strong economic crisis of the neo-liberal system.³

Going a few decades back to the 1960s and 1970s, both Chomsky and Foucault were well known left-orientated intellectuals deeply involved in politics; they are, so to say, “public intellectuals”. Their work in academia was intertwined with their analysis of public discourse, activism and the critique of contemporary political systems, and both argued for social change that was seen as a necessary move at that particular moment. Namely, the

debate took place in 1971, which was a period of crisis and the critique of existing order was articulated through the politics of the “New Left”. The “New Left” was a Marxist movement spreading within Europe as well the United States, advocating social change, with Herbert Marcuse, associated with the Frankfurt school of critical theory, as a leading figure.⁴ The New Left was a cultural as well as political protest: “the newly awakened sensibility of the rebels is a political factor; the real fight was a political fight...” (Ehrmann 1970: 383).

Let me now pass to the debate which is our main topic. Its kernel is the television debate which took place in 1971 in the Netherlands. Foucault's and Chomsky's approaches towards the social sciences in the debate are different. Chomsky looks at humans from a cognitive perspective and gives a positive stand on human nature which for him represents the cornerstone for the vision of a future society based on justice. On the other hand, Foucault is skeptical towards the idea of human nature which is, along with justice, just a concept created within our civilization. However, both Foucault and Chomsky have a common political agenda: both are leftists who give a profound critique of contemporary social order based on liberal capitalism, not contradicting one another, but simply observing the particular concepts like human nature or knowledge from different perspectives within a leftist framework. Maybe the best description for this debate comes from the introduction of Fons Elders, a Dutch philosopher, who introduces these two great intellectuals with the following words:

Perhaps the best way to compare both philosophers would be to see them as tunnellers through a mountain working at opposite sides of the same mountain with different tools, without even knowing if they are working in each other's direction (Chomsky, Foucault 2005: 1).

³ To Wendy Brown “neo-liberal” stands for the arrangement characterized by the dominance of the free market which imposes an imperative of production and profit onto political and social life, as well as the state (Brown 2006: 694).

⁴ The “New Left” was opposed to earlier leftist movements associated with the Marxist-Leninist ideology carried out in the Soviet Union. At the time, Peter Kropotkin was raising warnings about the “red bureaucracy” referring to Bolsheviks and Lenin, who, in the October revolution in 1917 and the overthrow of the Czar and the existing regime established the Soviet Union, and after the death of Lenin in 1924, Stalin, who started building his authoritarian rule, notorious for its coercive bureaucracy and strong police. The New Left was distancing itself from this kind of Stalinism..

To start with the basic concept “human nature” which can be seen as a set of the characteristics common in all human beings; these characteristics are considered to be natural or intrinsic to humans and independent of culture, that is, from the external surroundings. Firstly, the philosophical aspect of the debate begins with the fundamental question of whether we are “are we the product of all kinds of external factors, or if, in spite of our differences, we have something we could call a common human nature, by which we can recognize each other as human beings” (Elders in Chomsky, Foucault 2005: 2). Chomsky argues that human beings are creative, and “creativity” for him means that a human being has the possibility to say what they mean and to understand what people say to them and this collection of abilities is called “knowing a language” (Chomsky, Foucault 2005: 2). So, “knowledge” for Chomsky indicates mastering language in an unconscious way. In order to explain his view, he begins with the creativity which exists in a child’s ability to understand new sentences that they have never heard before. At issue is here is the gap between a child who has a small quantity of data that somehow results in a complex and sophisticated knowledge (Chomsky, Foucault 2005: 3). Furthermore, there is the phenomenon of individuals who have varied experiences in language (they have different stimuli; let’s say that one is exposed to a richer language stimuli than another) who inevitably reach congruent levels of knowledge; i.e. two English speakers, although they have different experiences with the language, can have the same knowledge of the English language, that is, they will be able to understand each other. Thirdly, people have different experiences which result in different limitations in a wide range of languages, and yet, a child can operate within the system of language (Chomsky, Foucault 2005: 3). On the basis of these three arguments, Chomsky concludes that there is instinctive and innate knowledge; that this knowledge arises from a schematism according to which from a small amount of data it is possible to acquire a sophisticated knowledge; and moreover, that this instinctive knowledge is a fundamental constituent of human nature (Chomsky, Foucault 2005: 4). As Chomsky claims in the debate “...this mass of schematisms, innate organizing principles, which guide our social and intellectual and individual behaviour, that’s what I mean to refer to by the concept of human nature” (Chomsky, Foucault 2005: 4). It is clear here that Chomsky is concerned with the cognitive aspect

of human beings; basically how humans are capable of creating language, a capability that is a universal human characteristic which represents the basis for his claim that there is such a thing as a universal human nature.

On the other side, Foucault does not see the concept of human nature as scientific. In his explanation he does not refer to cognitive aspects of human nature and in that understanding opposes Chomsky. Rather, Foucault talks about human nature as a concept within a particular science, so that basically they are talking about human nature from completely different points and not contradicting one another. Foucault states that “[i]n the history of knowledge, the notion of human nature seems to me mainly to have played the role of epistemological indicator to designate certain types of discourse in relation to or in opposition to theology or biology or history” (Chomsky, Foucault 2005: 7). In his elaboration, he compares two concepts – “life” and “human nature”. The concept of “life” which in the seventeenth and eighteenth century was used in studies of nature: the concept of “life” was used only in order to fix positions within the classification of human beings. “Life” was used simply to delineate biology as a science and to describe its object of study, rather than to describe the living beings. At the end of the eighteenth century, transformations in biological knowledge demonstrated new concepts in the science and also propagated particular notions which were used in order to delimit a certain type of scientific discourse (i.e. “life”) (Chomsky, Foucault 2005: 6). So, for Foucault, the notion of life is a meta-concept which is used in order to describe other biological notions (i.e. reproduction, assimilation of food; any notion which concerns the living beings); it is not used in the discourse of living beings per se; it is a concept defined by science, an indicator for that which biology will define as a living being (i.e. something which reproduces, eats, dies, etc). Since “life” in itself is not a scientific concept but a concept defined by science in order to indicate something else, and “human nature” is an analogue to the concept of “life” in biology, what follows is that “human nature” also is not a scientific concept for Foucault.

I find Foucault’s answer to be exceptional, though I do not find it to be a contra argument for Chomsky’s notions of human nature. Foucault and Chomsky are opposed to the question of the mere existence of human nature; however, in their argumentation they are not contrary because they simply

argue from different positions. Foucault talks about the constitution of science which defines its concepts, in this case science defines “life” and science also defines “human nature”. Chomsky does not talk about the constitution of science or the organizing of the concepts that are used in the process of that constitution; Chomsky himself says that he talks about neural networks which are the property of human cognitive systems thanks to which a child can obtain a sophisticated system of language. This is something biologically given and unchangeable and, according to Chomsky, it is a constituent element of human nature (Chomsky, Foucault 2005: 7). Furthermore, Chomsky agrees with Foucault’s claim that “life” is a concept within the biological sciences which can be used as an instrument in the process of a systematization of knowledge and he poses the following question in a pacifist tone: if the concept of life has taken biologists to the scientific understanding of life, could the concept of human nature lead to the scientific understanding of humans? In Chomsky’s words:

It seems to me that one might speculate a bit further [...] and ask whether the concept of human nature or of innate organizing mechanisms, or of intrinsic mental schematisms or whatever we want to call it, I don’t see much difference between them, but let’s call it human nature for shorthand, might not provide for biology the next peak to try to scale, after having – at least in the minds of the biologists, through one might perhaps question this – already answered to the satisfaction of some the question of what life is (Chomsky, Foucault 2005: 7).

To conclude with the concept of human nature, Chomsky and Foucault can be seen as “climbing the same mountain from the opposite side” in the way that they are not juxtaposed to one another, but are simply making different claims. Chomsky talks about a universal human ability to acquire a system of language which can be called “human nature”, while Foucault talks about concepts around which a particular science is shaped, in which the claims are classified in a particular way in a particular historical period. Essentially, Chomsky who is in his work principally concerned with language, talks more about the biological and psychological precognitions of human beings: how they create knowledge, and shape scientific theory. Foucault, on the other hand, studied the history of ideas or systems of thought,

the history of science, and is focused on the cultural treatment or the organization of a particular body of claims. He talks about “ideological” interpretation and readings of what human beings are and what they succeed in knowing, and he is more focused on the critique of political power which controls and shapes the whole social body.

The debate proceeds from Chomsky’s and Foucault’s claims on the existence of human nature towards knowledge. When talking about knowledge, Foucault and Chomsky start with the subject and the question of creativity. Foucault admits that he has given little thought to the creativity of individuals and he continues with the notion that he and Chomsky are in different positions apropos knowledge (Chomsky, Foucault 2005: 15). In this context of knowledge and creativity, linguists before Chomsky were focused on the process of language formation and the rules present in that process rather than on the innovation present in language. Chomsky, a major critic of behaviourists, is focused on creativity in a cognitive sense, or creativity at a “lower level” as a normal human act, as he explains it (Chomsky, Foucault 2005: 19-20), which I have mentioned above.⁵ In Foucault’s words:

Mr. Chomsky has been fighting against linguistic behaviourism, which attributed al-

⁵ Chomsky claims that behaviourism is only concerned with behaviour and the control of behaviour (Chomsky 2005: 32). George Graham explains behaviourism as a doctrine according to which “behaviour can be described and explained without making ultimate reference to mental events or to internal psychological processes” (Graham 2010). According to Graham, behaviourism sees human behaviour to be constituted in relation to external elements, not internal ones. He distinguishes three types of behaviourism: methodological behaviourism focused on psychology and whose object is the behaviour of humans and animals (not on their mental conditions). Secondly, there is psychological behaviourism which is a branch of psychology eager to analyse behaviour in relation to different exterior physical stimuli. Thirdly, there is analytical or logical behaviourism which attempts to decode human reactions in different situations (Graham 2010). Chomsky was famously one of the harshest critiques of behaviourism arguing that behaviourism denied the creativity of the subject by claiming that human reaction is conditioned by external stimulus which creates a reflex (but why then do people have different reactions to same stimulus?) Chomsky, in his generative grammar, argues for the universals of human linguistic structure, unlearned but known, which permit and assist children to acquire human languages.

most nothing to the creativity of the speaking subject; the speaking subject was a kind of surface on which information came together little by little, which he afterwards combined. In the field of history of science, or, more generally, the history of thought, the problem was completely different (Chomsky, Foucault 2005: 15).

In other words, in the history of thought, which was Foucault's area of interest, more emphasis was placed on individual creation, while the rules of creation were left in the shadows (Chomsky, Foucault 2005: 22, 23). Essentially, Foucault is impatient with creativity because it was too emphasized in the process of "discovery" within his area of interest. He describes how the history of knowledge had to obey two claims. The first is that of "attribution", which imposed that each discovery had to be situated and dated, and had to have an inventor responsible for it; Foucault calls it "the principle of the sovereignty of the subject" (Chomsky, Foucault 2005: 16). The second claim does not refer to the subject but to the truth which is not constituted in history, a truth that is hidden and waits to be unveiled (Chomsky, Foucault 2005: 16). Taking into account these two elements in the history of knowledge, Foucault concludes that knowledge in his area of interest (the history of science/thought) is always negative in relation to truth, or in Foucault's words:

It isn't difficult to see how these two claims were adjusted, one to the other: the phenomena of collective order, the "common thought", the "prejudices" of the "myths" of a period, constituted the obstacles which the subject of knowledge had to surmount or to outlive in order to have access finally to the truth; he had to be in an "eccentric" position in order to "discover" (Foucault, Chomsky, 2005: 16).

Regarding these two imperatives, Foucault asks the question, "what if understanding were a complex, multiple, non-individual formation, not 'subjected to the subject' which produced effects of truth?" (Chomsky, Foucault 2005: 17). Foucault proposes a different perspective on knowledge according to which the generation of knowledge could be seen as collective, and not an individual practice. It is obvious that when Foucault talks about "understanding as a collective totality of rules allowing such and such a knowledge to be produced in a certain period" (Chomsky, Foucault 2005: 17), he

speaks about an "episteme"; the three big structures – Renaissance, the classical period, and nineteenth century – which replace each other, but are separately internally fixed, which he famously introduced in *The Order of Things: An Archaeology of the Human Sciences*. He demonstrates his point in the example of medicine which in the period from the 1770-1830 was completely transformed, which means that a completely new grid with new rules, decisions and limitations has occurred. That is to say, a new inner logic of medical understanding has occurred, or an entirely new "grille" (Chomsky, Foucault 2005: 18). In the history of knowledge, according to Foucault, there are two lines of inquiry: "according to one, one has to show how, under what conditions, and for what reasons the understanding modifies itself in its formative rules, without passing through an original 'inventor' discovering the 'truth'; and according to the other, one has to show how the working of the rules of an understanding can produce in an individual new and unpublished knowledge" (Chomsky, Foucault 2005: 18).

It seems that creativity is connected with rules for Chomsky and Foucault, and that they are talking about the rules from different perspectives, which again is not contradictive. Foucault talks about them from the social point of view: episteme, or rules are in a juxtaposed position towards the creativity of the subject, that is, a certain understanding is always bounded within certain rules of a particular period. Meanwhile, for Chomsky, who speaks from the linguistic point of view, the rules are the condition of the creativity of the subject, meaning that a person who has mastered the system of rules knows how to use a language. Again, interestingly, they are not directly opposed, just speaking from different aspects. For example, if I was moderating this debate, I would ask Foucault what does he think about the human ability of acquiring a language and is that not a universal human ability that can be seen as human nature, and if not, how should that be described?

The debate shifts from creativity and knowledge to the possibility of creating a scientific theory from a small amount of data. To Chomsky that kind of achievement resembles the process in which a child reveals its own language. He claims the following:

If we really want to develop a theory of scientific creation, or for that matter artistic creation, I think we have to focus attention precisely on that set of conditions that, on one hand, delimits and restricts the scope of our possible knowledge, while at the same

time permitting the inductive leap to complicated systems of knowledge on the basis of a small amount of data. That, it seems to me, would be the way to progress towards a theory of scientific creativity, or in fact towards any question of epistemology (Chomsky, Foucault 2005: 25).

Chomsky's idea that, in order to develop a scientific theory, one has to focus on the limits of our possible knowledge and then to make an inductive leap drives Foucault to make a critique of the idea of "progress" and "growth": both of these concepts arise from within a European understanding which, according to Foucault, turned out to be a universal understanding. He points out the principle of convergence which allows for many different ways of making several types of knowledge appear simultaneously, not just from the one direction of "progress" (Chomsky, Foucault 2005: 26, 27). Chomsky agrees with Foucault because he was not implying the idea of scientific progress as a matter of accumulated knowledge:

I didn't think that scientific progress is simply a matter of the accumulated addition of new knowledge and the absorption of new theories and so on. [...] But it is precisely because of this property of our minds, which in detail we don't understand, but which, I think, in a general way we can begin to perceive, which presents us with certain possible intelligible structures, and which in the course of history and insight and experience begins to come into focus or fall out of focus and so on; it is precisely because of this property of our minds that the progress of science, I think, has this erratic and jagged character that you [Foucault] describe (Chomsky, Foucault 2005: 27, 38).

Chomsky is approaching Foucault's claims about the rules which determine a particular knowledge when he describes behaviourism, which is, according to him

a negation of the possibility of developing a scientific theory. That is, what defines behaviourism is the very curious and self-destructive assumption that you are not permitted to create an interesting theory. [...] Behaviourism is the arbitrary insistence that one must not create a scientific theory of human behaviour; rather one must deal di-

rectly with phenomena and their interrelation... (Chomsky, Foucault 2005: 34).

Chomsky looks at behaviourism not as a science, but as a set of limitations on the construction of a scientific theory, and it looks to me that this characterization of behaviourism sees its requirements as a grid (in Foucauldian terms) imposed. Here, in this point, they finally address an issue from the same perspective: questioning the rules of science that determine a particular subject-matter, in this case behaviourism, which deprived the creativity of the subject by focusing simply on their behavior, behaviour which is entirely conditioned by external conditions and has nothing to do with the internal matter of the subject.

One can conclude that, in the context of science, Chomsky talks about the cognitive aspects of human beings present in the process of creating a particular collection of knowledge, a process of creation that Foucault is skeptical of:

Where perhaps I don't completely agree with Mr. Chomsky is when he places the principles of these regularities, in a way, in the interior of the mind of human nature. [...] But to say that these regularities are connected, as conditions of existence, to the human mind or its nature is difficult for me to accept. [...] Perhaps the point of difference between Mr. Chomsky and myself is that when he speaks of science he probably thinks of the formal organization of knowledge, whereas I am speaking of knowledge itself, that is to say; I think of the content of various knowledges which is dispersed into a particular society, permeates through society, and asserts itself as the foundation for education, for theories, for practices (Chomsky, Foucault 1005: 29, 30).

So, while Chomsky is talking about the intrinsic aspect of the human mind, Foucault is talking about a "second phase" which follows from Chomsky's cognitive one. Again, they are not opposed in their claims, or in Chomsky's words:

I think that illustrates very nicely the way in which we're digging into the mountain from opposite directions, to use your [Elders'] original image. That is, I think that an act of scientific creation depends on two facts: one, some intrinsic property of the mind, another, some set of social and intellectual conditions

that exist. And it is not a question, as I see it, of which of these we should study; rather we will understand scientific discovery and similarly any other kind of discovery, when we know what these factors are and can therefore explain how they interact in a particular fashion, my particular interest, in this connection at least, is with the intrinsic capacities of the mind; yours [Foucault's], as you say, is in the particular arrangement of social and economic and other conditions (Chomsky, Foucault 2005: 32).

With the notions regarding the constitution of science the philosophical aspect which focused principally on the question of human nature is finished. The debate now turns toward politics which begins with Foucault's explanation of the importance of politics, which is

probably the most crucial subject to our existence, that is to say the society in which we live, the economic production within which it functions, and the system of power which defines the regular forms and the regular permissions and prohibitions of our conduct. The essence of our life consists, after all, of the political functioning of the society in which we find ourselves. [...] Not to be interested in politics, that's what constitutes a problem (Chomsky, Foucault 2005: 37).

Beginning with Chomsky, well-known for his anarcho-syndicalism, or libertarian socialism, who argues that there is a connection between freedom, dignity and creativity. Following from this, there is a particular social organization in which those fundamental human characteristics could be achieved (Chomsky, Foucault 2005: 42). The connection between anarcho-syndicalism and free, dignified, and creative human nature is that, according to Chomsky, "a fundamental element of human nature is the need for creative work, for creative inquiry, for free creation without arbitrary limiting effects of coercive institution, then, of course, it will follow that a decent society should maximize the possibilities for this fundamental human characteristics to be realized" (Chomsky, Foucault 2005: 37). In short, for Chomsky, anarcho-syndicalism is an optimistic vision of future society: "a federated, decentralized system of free associations, incorporating economic as well as other social institutions" (Chomsky, Foucault 2005: 38). And thanks to the progress present in technology, in this kind of system workers could

form independent assemblies and would no longer represent just a wheel in the productive process (Chomsky, Foucault 2005: 39). Chomsky gives this vision of a future society because he sees it as a task. Essentially, he argues that there are two intellectual tasks. The first is focused on the creation of a just future society, while the second which detects the nature of power, oppression, terror and destruction (Chomsky, Foucault 2005: 41).

On the other hand, in this story concerning human nature and future society, Foucault is much more careful: "...it is difficult to say exactly what human nature is. Isn't there a risk that we will be led into error?" (Chomsky, Foucault 2005: 44). He does not have a vision of a better society but he firmly maintains the position of the present historical moment, claiming that to him it seems that "the real political task in a society such as ours is to criticize the working of institutions, which appear to be both neutral and independent" (Chomsky, Foucault 2005: 41). He is concentrated on the disclosure of power which shapes society. As a true Marxist, he argues that state and its institutions are managed by a ruling class (Chomsky, Foucault 2005: 40, 41).

In this political part of the debate the question of the disturbance of social order comes into focus, in particular in relation to a concrete activity which threatens the social order. Foucault and Chomsky are both known as activists but they have different stances towards civil disobedience. To Chomsky civil disobedience threatens the social order, but with the possibility of fascism passiveness is also a risky act (Chomsky, Foucault 2005: 44, 45). For Foucault, civil disobedience is not an act of resistance as it is for Chomsky. For Foucault civil disobedience is "a legal, obligatory behaviour in violation of the commands of the state, which may not be legal" (Chomsky, Foucault 2005: 46). So it appears that civil disobedience for Foucault is ineffective because in practicing it one still exists in the framework of the commands of the state, while Chomsky sees the act of civil disobedience as a pressure directed towards the state.

Civil disobedience leads us to the question of justice. Chomsky opens up in this dialogue. He states very clearly that people are not in a position of creating a system of ideal justice and an ideal society because people are too limited and too biased; however, what people are in a position to do is imagine and organize a more just society (Chomsky, Foucault 2005: 50). He separates "legality" and "justice" and claims that people should act in accordance with

the law if the law includes justice; but when a juridical apparatus is not organized around justice, people should confront that kind of system (Chomsky, Foucault 2005: 51). Foucault has different ideas regarding justice; to him, “justice” as well as “human nature” are “notions and concepts which have been formed within our civilization, within our type of knowledge and our form of philosophy, and that as a result form part of our class system...” (Chomsky, Foucault 2005: 58).

Foucault views are, I would say, relativistic. He claims that the proletariat does not go into war contra ruling class because of justice, but because of power; he claims that “one makes war to win, not because it is just” (Chomsky, Foucault 2005: 51). I find his notions on justice interesting because although a relativist, it seems to me that Foucault nevertheless has a strong sense of justice which is evident in his critique of undemocratic political systems. Nevertheless, his relativistic notions are evident in his claim that in the war which the proletariat wages against the ruling class *spiritus movens* is not justice but a fight for power (Chomsky, Foucault 2005: 51). My first question is, generally speaking, what is the goal of the proletariat’s fight in the first place? I mean, if the geometry of power is fairly designed in a society, that is, if society is built on justice, would the proletariat go to war for more power just in order to win more power? On a more grounded level, if we take into consideration the concrete example of the Soviet Union, which constituted a socialist state where the proletariat did overthrow the ruling class and established a centralized and authoritarian state greatly based on violence and control, Foucault’s caution makes sense. Foucault even goes a step further by claiming that “justice” “in itself is an idea which in effect has been invented and put to work in different types of societies as an instrument of a certain political and economic power or as a weapon against that power” (Chomsky, Foucault 2005: 54). Basically, “justice” is a concept shaped by the subordinated and functions as a justification for oppression. Furthermore, in a classless society that kind of concept would be unnecessary (Chomsky, Foucault 2005: 54).

Chomsky answers in a simple way by claiming that there is some basis of justice which resides in fundamental human qualities and that there is a “real” notion of justice that is so grounded together with true humanly valuable concepts such as love, kindness, sympathy, solidarity and so on (Chomsky, Foucault 2005: 55). Chomsky says that the “social

revolution that you’re trying to achieve is in the ends of justice, is in the ends of realizing fundamental human needs, not merely in the ends of putting some other group into power, because they want to” (Chomsky, Foucault 2005: 57). Here it is evident that Chomsky talks from a universal perspective while Foucault is much more careful because he takes into consideration the situation with socialism in Europe and Russia, where, in one moment, socialism took a wrong turn.

To conclude, behind every political philosophy there is an implicit or explicit stand regarding human nature. Today’s neo-liberal politics are based on the notion that human nature is negative and that everybody should be seen as an ‘enemy’ or, in today’s terminology, a ‘competitor’, in a world where private profit is imperative. The contemporary world is shaped with regard to profit; humans are subjected to renting themselves in markets and are subjected to a neo-liberal ethos which insists on individualism and competitiveness. Although anarchism can be seen as utopian and naive because it presumes that humans have a positive human nature, meaning that they are benevolent and have a sense of justice and altruism, it should be seen as a tool for raising consciousness against neo-liberal ideology, especially in this historical moment when we are witnessing an enormous worldwide economic crisis. In this kind of political situation, Chomsky’s and Foucault’s critique of social order, as well as Chomsky’s alternative vision of a society, and Foucault’s precaution, should be seriously taken into consideration.

Conclusion: Bridging the Continental-Analytic Gap

I have argued that the difference between Chomsky and Foucault, dominant in their debate, is more a difference of focus than a clash of contrary views. Chomsky argues from the cognitive aspect of human beings, focusing on innate structures which create knowledge, while Foucault is more focused on cultural conditions determining a particular body of claims in a certain period which is formed as “knowledge”. In the political context of the debate they differ regarding their future vision of society. While Chomsky proposes an idea of society based on rules which will guarantee freedom for humans, Foucault does not want to enter into speculations about the future; he remains in the present historical moment. Concerning justice, they differ inasmuch as Foucault has a relativist notion regarding justice, although he also has a strong sense of justice in his

analysis of power, while to Chomsky the idea of justice is the basis of human nature and a future society.

In a further in-depth analysis of the constitutive concepts of the debate within the greater scope of their work, they converge on a number of points. In the analysis of power and its *modus operandi*, Foucault argues that power runs through the whole social body and in that flow it creates inequalities; basically the society in which we live today is a successor of the bourgeois socio-economic order which is divided hierarchically and is undemocratic. Chomsky shares this criticism of power which for him is in the hands of governments and elites who control the rest of the society. Essentially, in this critique of society Chomsky and Foucault can be seen as converging. To Foucault the technology of power regulates the complete social body using different rules, while to Chomsky the masses are being regulated primarily through media which is possessed by those in power. This is also one of their points of convergences. Furthermore, both of them were (and Chomsky still is) passionate activists who shared the idea of resistance to power structures. To Chomsky civil disobedience is a way of rebellion against the existing socio-political order, and although Foucault claimed that this kind of resistance is not really meaningful because it is an act within the state's framework (however, in his work he is not really clear what resistance is; he just claimed that it is undistinguishable from power), in practice he was a vigorous activist. The critique of knowledge, which is represented as neutral and truthful, coming from a formal institution, is also a point of convergence in their work. Namely, Foucault argues for genealogy as a possible anti-science towards verified knowledge from formal institutions. Similarly, in his study of counterrevolutionary subordination, Chomsky is critical towards the liberal scholarship which is devoted to the maintenance of the values and social order of liberal bourgeois democracy. In addition, both of them are critical towards intellectual elitism, as well as those on the Left.

The point in which they do split is on the question regarding the future. While Foucault is cautious and does not want to speculate about the future, remaining focused on the disclosure of political power which shapes the social body, Chomsky gives a vision of a future society, a better society based on justice. On the political side, taking into consideration the sad socio-political context of the crises of our time, the differences within the Left, in this case

between Chomsky and Foucault, become evanescent. This means that not much energy should be spent in finding the theoretical divergences in their work and pushing them further, as both perspectives should be taken into consideration in order to finally achieve a Left hegemony in a political context. Their brilliant critique of power and its *modus operandi*, their critique of institutions and knowledge (which are far from objective and neutral as usually represented), their ideas regarding resistance to power structures and the importance of it, Chomsky's idea of a better society and Foucault's incisive caution regarding this all should be taken into consideration in thinking about a more democratic society.

On the philosophical side, it is important to stress their convergence since it promises a more fruitful dialogue between continental and analytic philosophy. After all, as we already stressed, Foucault is the central post-structuralist author, whereas Chomsky is the thinker who has been inspiring analytic philosophers for half a century. Their dialogue should be a paradigm of how the two schools of thought can come to mutual understanding.

References

- Bennet, Tony, et al. (2005) *New Keywords: A Revisited Vocabulary of Culture and Society*. Malden: Blackwell Publishing.
- Bertani, Mauro and Alessandro Fontana, eds. (2003) *Michel Foucault: Society must be defended. Lectures at the College de France 1975–1976*. New York: Picador.
- Brown, Wendy. (2006) "American Nightmare: Neoliberalism, Neoconservatism, and De-Democratization." *Political Theory*. (34):6. 690–714.
- Chomsky, Noam (1967) *On Resistance*. Retrived March 23rd 2012 (<http://www.chomsky.info/articles/19671207.htm>).
- Chomsky, Noam, and Michel Foucault. (2006) *The Chomsky-Foucault debate on Human Nature*. NY: The New Press.
- Edgley, Allison. (2009) "Manufacturing Consistency: Social Science, Rhetoric and Chomsky's critique." *Westminster Papers in Communication and Culture*. 6(2): 23–42.
- Enhrmann, Henry W. (1970) "The agony of Utopia." *Polity*. 2(3): 380–391.
- Eribon, Didier. (1992) *Michel Foucault*. Cambridge MA: Harvard University Press.

Faubion, James D., ed. (2000) *Power: Essential Works of Foucault, 1954–1984, Volume III*. London: Penguin group.

Foucault, Michel. (1980) *History of Sexuality Volume 1: An Introduction*. New York: Vintage Books.

Gordon, Colin. (1980) *Power/Knowledge: Selected Interviews and Other Writings, 1972–1977* by Michel Foucault. New York: Pantheon.

Graham, George. (2010) "Behaviorism". The Stanford Encyclopedia of Philosophy. Retrived May 5th 2012 (<http://plato.stanford.edu/archives/fall2010/entries/behaviorism/>)

Hall, Stuart, and Anthony McGrew (1992) "The question of cultural identity". *Modernity and Its Futures*. Cambridge: Polity Press. 596–632.

Hardt, Michael, and Antonio Negri. (2000) *Empire*. Cambridge MA: Harvard University Press.

Hindess, Barry (2007) "Marxism". Pp. *A Companion to Contemporary Political Philosophy*. Goodin, Robert E. et al. Oxford: Blackwell Publishing. 383–403.

Lovett, Frank. (2010) "Republicanism". The Stanford Encyclopedia of Philosophy. Retrived May 17th 2012 (<http://plato.stanford.edu/archives/sum2010/entries/republicanism/>)

Miller, James. (1993) *The Passion of Michel Foucault*. Cambridge MA: Harvard University Press.

Miller James. (1993) "Foucault's Politics in Biographical Perspective". *Salmagundi*. 97:30–44.

Pateman, Barry, ed. (2005) *Chomsky on Anarchism*. Oakland, CA : AK Press.

Quinton, Anthony, and Anne Norton. (2007) "Conservatism" in *A Companion to Contemporary Political Philosophy*. Goodin, Robert E. et al. Oxford: Blackwell Publishing. 285–312.

Resch, Robert Paul. (1989) "Modernism, Post-modernism, and Social Theory: A Comparasion of Althusser and Foucault". *Poetics Today*. 10(3): 511–549.

Sartre, Jean-Paul. (1948) *Existentialism and Humanism*. London: Methuen & Co. Ltd.

Stevenson, Leslie, and David L. Haberman. (1988) *Ten Theories of Human Nature*. New York: Oxford University Press.

THE ROLE OF RATIONALITY IN THE FORMULATION AND THE COMPLIANCE WITH THE PRINCIPLES OF JUSTICE

Peter Cholakov

Institute for the Study of Societies and Knowledge
Bulgarian Academy of Sciences, Sofia, Bulgaria
e-mail: p_cholakov@yahoo.com

Abstract:

The function of rationality in *A Theory of Justice* (1971), which is of paramount importance for John Rawls' (1921–2002) project, is often criticised as ambiguous. David Gauthier, for example, claims that Rawls develops principles for recipients who essentially share his intuitions of morality, without managing to prove their validity. In *Political Liberalism* (1993), *Justice as Fairness* (2001) and other writings Rawls himself embarks upon the task to throw more light on this issue, making the Kantian distinction between 'rational' and 'reasonable'. I intend to demonstrate that in *A Theory of Justice* the formulation and the compliance with the principles of justice are based on the interaction between the rationality, represented in the idea of the good, and the sense of justice of individuals.

Key words: Rawls, principles of justice, veil of ignorance, rationality, reasonability, congruence.

In the modern world it often seems that the moral values resemble monuments with lost and forgotten histories. The vanishing spell of religion, substituted by the miracles of modern science, deprives moral norms and notions of their roots and puts in doubt their self-evidence. This bitter conclusion represents philosophers as archaeologists seeking to explain and legitimate the existence of some remains from the past. It makes no exception for major philosophical categories such as justice. Thus, if rationality is the leading principle through which we organise our present secularised knowledge, it is natural to attempt to use it as a means of reconstructing and reformulating our understanding of the concepts of morality.

The object of analysis of this paper consists of an exploration and discussion of the role of rational-

ity in the formulation and the compliance with the principles of justice, suggested in the liberal political philosophy of John Rawls (1921 – 2002). Rawls' *A Theory of Justice* (1971), as well as the works of other prominent social and political thinkers (Weber, Harsanyi, MacIntyre, Gauthier), illustrates the significance of rationality as an intellectual tool used in the formulation of moral notions and political theories¹. Despite this similarity, the aforementioned theories use rationality in a remarkably different way.

John Rawls writes *A Theory of Justice* with the conviction that rationality itself is not a sufficient basis for the development of a moral theory. In order to construct such a theory, we also need to implement our moral intuitions. Thus the main goal of Rawls is to *rationalise our intuitions of justice*. Although *A Theory of Justice* has an army of enthusiastic proponents, the function of rationality in it is *perceived by some scholars as ambiguous*. Most of Rawls' critics, such as David Gauthier for example, claim that in *A Theory of Justice* rationality plays only a supplementary role, and Rawls develops principles for recipients who essentially share his intuitions of morality without managing to prove the validity of these intuitions. The problem is illustrated by the fact that in *Political Liberalism* (1993),

¹ Max Weber famously claims that Modern rationality, as the main factor for *Die Entzauberung* – “the removing of the world” – was an artefact of religion (Weber 1996). According to Alasdair MacIntyre the notions of morality are contextually determined. Thus, their recovery from the crisis, caused by their being spread by rationality doubts in their universal validity, is an extremely strenuous task because we often lack the very fundament of these notions. In *Morals by Agreement* (1986) David Gauthier states that he developed a moral theory for self-interested individuals, based entirely on rationality.

Justice as Fairness (2001) and other writings Rawls himself tried to throw more light on this issue, making the Kantian distinction between ‘rational’ and ‘reasonable’².

The main thesis of this paper consists in the claim that in *A Theory of Justice* the formulation and the compliance with the principles of justice are both based on the interaction between rationality, represented in the idea of the good, and the sense of justice of the individuals. In his subsequent works Rawls has described the concepts of rationality and reasonability as complementary ideas for the solution of the above problems (Rawls 1993: 52/ 2001a: 6). We will focus mainly on *A Theory of Justice* and discuss the part that rationality plays in its interaction with the sense of justice and reasonability.

The main thesis is explored in the three parts of the present work. *Part one* analyses rationality and the sense of justice in Rawls’ original position, as it is defined in *A Theory of Justice*. We claim that the distinction between reasonability and rationality in Rawls’ subsequent writings could help understand better the meaning of the original position presented in *A Theory of Justice*. *The second and the third parts* discuss the role of rationality for the stability of the well-ordered society. This significant problem is underexplored in the scientific literature. We aim to examine it thoroughly while dividing it into two separate questions. *The second part of the paper* is devoted to rationality and the question of how a person develops a sense of justice. *The third part* explores how this sense is supported in the well-ordered society. In order to clarify the second problem we start our analysis with Rawls’ view of goodness as rationality.

1. Choosing the principles of justice

Rawls defines his understanding of *rationality* in Chapter III of *A Theory of Justice*, where he examines the rationality of the parties involved in the original position. Rationality is perceived as: “a.) taking effective means to ends with unified expectations and objective interpretation of probability; b.) as above but without unified expectations and using the principle of insufficient reason” (Rawls 1999: 126)³.

The sense of justice is “an effective desire to apply and to act from the principles of justice and so from the point of view of justice” (p. 497). Rawls

focuses his attention on distributive justice. Its subject is “the way in which the major social institutions distribute fundamental rights and duties and determine the division of advantages from social cooperation” (p. 6). Rational parties decide about the principles of justice behind the veil of ignorance and these principles are adopted as a result of rational choice. That is why Rawls perceived the theory of justice as the most significant part of the theory of rational choice⁴. The choice under the veil of ignorance should be fair *ex deffinitio* from which follows the famous interpretation of “justice as fairness”.

In *Political Liberalism* Rawls revises this relation between rational choice and the theory of justice. He claims that the quoted above dictum from *A Theory of Justice*, regarding rational choice, must not be interpreted in such a way that the way that in this treatise the principles of justice are merely an artefact of the rationality of the parties (an example of moral theory that claims to be entirely based on rationality, in which Rawls is criticised, is that of David Gauthier⁵). Rawls states that within the idea of fair cooperation reasonable and rational are “complementary ideas”. Each is “an element of this fundamental idea and each connects with its distinctive moral power, respectively with the capacity for a sense of justice and the capacity for the conception of the good” (1993: 52). While the conception of the good provides agents with their own goals, which they pursue in cooperation, the sense of justice guarantees compliance with the principles of cooperation. The latter idea is common for *A Theory of Justice*, *Political Liberalism* and *Justice as Fairness*. It supports our thesis that the formulation and the compliance with the principles of justice are an artefact of the interaction between rationality, represented in the idea of the good, and the sense of justice as a moral power of reasonable individuals. Note that contrary to Rawls’ statements, we claim that the two moral powers are not only complementary but also interdependent. There are at least three cases in *A Theory of Justice* supporting such a view:

4In *Morals by Agreement*, Gauthier claims that in fact Rawls is inconsistent with this definition of the theory of justice, and instead, *petitio principii*, is supporting arguments for the principles of justice on intuitions. Hence Rawls fails to substantiate the rational compliance with these principles (Gauthier 1986: 4).

5 To “choose rationally, one must choose morally. [...] Morality, we shall argue, can be generated as a rational constraint from the non-moral premises of rational choice. Neither Rawls nor Harsanyi make such a claim” (Gauthier 1986: 4).

²On this distinction see the first part of the article.

³All citations from the *Theory of Justice* are from 1999 edition.

the morality of authority, the arguments for congruence and the construction of the original position. We discuss them below.

Adjusting the Kantian definition for the purposes of a political concept of justice, Rawls associates reasonability⁶ with: “first, the willingness to propose and honour fair terms of cooperation, and second, with the willingness to recognise the burdens of judgement and to accept their consequences” (1993: 49). To say that one is rational means that he is following his intellectual powers while pursuing a certain goal. Reasonable persons desire “for its own sake a social world in which they, as free and equal, can cooperate with others on terms that all can accept” (p. 50) – this statement represents the core of the idea of reciprocity.

In *Political Liberalism* and other works the distinction between rationality and reasonability is used by Rawls in the explanation of the stability of a political concept of justice in a pluralist society. In order to understand better the role of rationality in the original position, we may use the principle of elimination; we could start from the intuitive fundamentals of the principles of justice, and see what remains for rationality. There are several different levels/stages on which the parties use their moral intuitions in the original position:

- Because of the reasonable restriction of the “veil of ignorance” each person (party) possesses only a rough idea of his plans for life: “he does not know how to tailor principles to his advantage. The parties are effectively forced to stick to general principles, understanding the notion here in an intuitive fashion” (Rawls 1999: 113). For example, the principles naturally are unconditional, universal in application, public and etc. (pp. 114-18).
- The parties have the desire to choose fair principles for society because they *ex initio* do not have the moral vice of envy (p. 124).
- The capacity for the sense of justice of the parties “insures that the principles chosen will be respected” (p. 125).
- In accordance with these levels, there are certain levels of the implementation of rationality:

- The parties formulate the draft-principles using not only their intuitions but also a sketch of their rational plans of life (p. 124).
- The parties decide on the principles using rational choice as a method. The agents are rational because they will not enter into an agreement they cannot keep, or can do so only with great difficulty. Therefore “in assessing conceptions of justice, the persons in the original position are to assume that the one they adopt will be strictly complied with” (p. 126)⁷.
- Rationality is a means of persuasion of the recipients of Rawls’ theory to follow its principles (see below).

It is obvious from the above analysis that neither the construction of the original position, nor the formulation of the principles of justice is possible without *the interaction* between the rationality and reasonability of the parties.

The remarks on rationality lead us to *its ‘comprehensive’ level* – the whole theory is an attempt to rationalise our intuitions of justice as recipients of Rawls’ treatise. If we put ourselves in the thought experiment⁸ of the “veil of ignorance”, in the following interaction between the chosen principles of justice and our intuitive moral premises *under the reflective equilibrium*, we will not simply rationalise our considered judgements – we could be convinced as *rational agents* to act morally. In Rawls’ interpretation, in the reflective equilibrium “one is to be presented with all possible descriptions to which one might plausibly conform one’s judgements together with all relevant philosophical arguments for them. [...] A person’s sense of justice may or may not undergo a radical shift” (1999: 43). Hence on *the basis of reasonable restrictions the rational agents, who follow their good, could uphold their sense of justice*. That is why we perceive that it is more plausible to *interpret the two moral powers as interdependent* rather than treating them as autonomous complementary units.

The contract between the parties is indeed imaginary, but it could be used as a criterion for our empirically existing institutions and relations – it “moves us closer to the philosophical ideal, it does

6 In his article ‘Kantian Constructivism and Moral Theory’, Rawls defines ‘reasonable’ as “a morally substantive matter beyond the bounds of rational choice”; see Rawls 1980 for further details.

7 This of course does not mean that Rawls automatically solves the problem of compliance.

8 For the role of the veil of ignorance as a device of ‘making vivid’ our intuitions, I am indebted to Dr. Matt Matravers, professor at the Department of Politics, the University of York, UK.

not, of course achieve it” (p. 43). The chosen principles of justice will be impartial and will reflect the interests of everybody. These are as follows:

First principle. Each person is to have an equal right to the most extensive total system of basic liberties compatible with a similar system of liberty for all.

Second principle. Social and economic inequalities are to be arranged so that they are both:

- (a) to the greatest benefit of the least advantaged, consistent with the just savings principle, and
- (b) attached to offices and positions open to all under the conditions of fair equality of opportunity. (p. 266)

The original position also defines the principles requiring compliance with the concept of justice. Rawls states (p. 95) that in order to establish a complete conception of right, the parties in the original position are to choose in “a definite order not only a conception of justice but also principles to go with each major concept falling under the concept of right” (e.g. ‘fairness’, ‘fidelity’ and etc.). The concept of right regulates social systems and institutions, individuals and the law of nations. One of the principles for the individuals is fairness. It is of crucial importance for the problem of compliance because all “requirements that are obligations” derive from it (p. 96). This principle holds that a person is required to do her part as required in the rules of an institution when two conditions are met: “first, the institution is just (or fair), that is, it satisfies the two principles of justice; and second, one has voluntarily accepted the benefits of the arrangement or taken advantage of the opportunities it offers to further one’s interests” (p. 96).

2. Stability: acquiring a sense of justice

The distinction which we established between rationality and the sense of justice in the original position sheds light on the fact that the principle of fairness is based on the rationality of the parties⁹. They prefer concepts of justice which are easier to comply with because the observation of the duty of

compliance is an absolute premise for stability. The very idea of stability is connected with that of the well-ordered society designed to

advance the good of its members and [be] effectively regulated by a public sense of justice. [...] Since a well-ordered society endures over time its conception of justice is presumably stable: that is when institutions are just (as defined by this conception), those taking part in these arrangements acquire the corresponding sense of justice and desire to do their part in maintaining them (Rawls 1999: 398).

The problem of stability could be divided *into two important questions*: ‘how we acquire a sense of justice’ and ‘how we uphold this sense in a well-ordered society’. Rationality plays a substantial role in the answers provided to both issues. It follows from the above definition of stability that in a well-ordered society we will acquire a sense of justice in the process of socialization. Rawls perceives this process as the main stimulus for the person’s moral development. He distinguishes between three stages of this development leading to a sense of justice: *the morality of authority*, *the morality of association*, and *the morality of principles* (pp. 405-20). Three corresponding psychological laws regulate these stages. We will discuss only the first stage, because it is marked by the transformation of the rational egoism of the child into moral feelings. The first stage is a *condicio sine qua non* since the development achieved in it makes the following two stages possible.

A sense of justice is acquired gradually by the younger members of society as they grow up (p. 405). In this first stage, according to Rawls, the family has a leading role in socialization and the precepts formulated by the parents are based on principles of justice. Rawls, following Rousseau, formulates the psychological principle that the “child comes to love the parents only if they manifestly love him” (p. 406). Although the child has – as a potential – the ability to love, her actions are motivated at the beginning only by a rational self-interest (the loving parents fulfil all her desires and uphold her ego). On the basis of the parents’ love, the child develops strong affection towards them. From this moment her love does not have a rational instrumental explanation, because the child no longer perceives her parents as means to fulfil egoistic ends. Then the child starts to internalise the parents’ pre-

⁹ This does not mean that the citizens have the duty to comply with the demands of the legal and political order under all circumstances. For example, in order to eliminate the malfunctions of the institutional system the citizens have – under certain precise conditions – the right of civil disobedience (1999: 326).

cepts (based on the principles of justice), which she felt as burdens before. And, most importantly, the child develops the sense of guilt when she fails to comply with them.

The analysis of the morality of authority *demonstrates a lack of coherence* in Rawls' concept of rationality and reasonability. While Rawls, as we have already seen, claims that these two notions are complementary but independent – it is apparent that there is a certain *interdependence between the ability to pursue our good and the sense of justice*. In the case of the morality of authority, the basis for the sense of justice is only possible on the grounds of the interaction between the child's rational self-interest, on the one hand, and her instincts and desires, on the other hand.

In his response to the second question, regarding stability, Rawls claims that it is rational (as defined by the "thin" theory of the good) for those in a well-ordered society to affirm their sense of justice as regulative of their plan of life. He attempts to show that the disposition to take up and to be guided by the standpoint of justice accords with the individual's good (p. 497). This accord is expressed by the term "congruence". In order to understand its meaning we need first to present briefly the main points in Rawls's theory of the good.

3. Stability: congruence

Rawls distinguishes between *two theories of the good*. We need a "thin" theory "to explain the rational preference for primary goods and to explicate the notion of rationality underlying the choice of principles in the original position" (1999: 349). This theory also examines whether having and maintaining a sense of justice is a good (in the thin sense) for the members of the well-ordered society¹⁰. If within theory "it turns out that having a sense of justice is indeed a good, than a stable society is as stable as one can hope for" (p. 350). Hence, the thin theory regards both the formulation of principles of justice and the compliance with them. The "full" theory of the good comes after the "thin" one, because it is based on the established notion of justice and the concept of right.

Rawls starts the thin theory by arguing that that once we establish that

an object has the properties that it is rational for someone with a rational plan of life to

want, then we have shown that it is good for him. And if certain sorts of things satisfy this condition for persons generally, then these things are human goods (p. 351).

In order to understand the role that rationality plays in the formulation of the concepts of good (in the thin theory) and right, we have to emphasise the fact that the definition of good is morally neutral. One may say of a man "that he is a good spy, or a good assassin, without approving of his skills" (p. 354). Consequently, we use the concept of rationality to clarify the idea of good that the parties in the original position have; *rationality on its own – without interaction with the sense of justice – cannot lead us to the concept of right*. We perceive the good as moral value in the full theory, because it is defined there by the principles of right. Therefore, for example, in the full theory the good judge is the one who follows the principles of justice (p. 355).

The term "rational plan of life" plays an important part in Rawls's view of the relationship between rationality and justice. We already postulated that in the original position the parties have *an idea* of their good, represented in the sketches of their rational life plans. Rawls gives two definitions of the rationality of an individual's plan of life. First, it is rational if, and only if

(1) it is one of the plans that is consistent with the principles of rational choice when these are applied to all the relevant features of the situation, and (2) it is that plan among those meeting this condition which would be chosen by him with full deliberative rationality, that is with full awareness of the relevant facts and after the careful consideration of the consequences. Secondly, a person's interests and aims are rational if and only if, they are to be encouraged and provided for by the plan that is rational for him (p. 359).

From this definition it follows that the rational plans of life of various individuals will be quite different, because they depend on the person's specific endowments and circumstances.

Rawls mentions two types of principles of rational choice. These represent the content of the notion of rationality. *The counting principles* (effective means, greater likelihood and inclusiveness) are relevant only for short-time questions. Therefore in order to represent how our choice depends on the circumstances on the basis of which we choose, Rawls adds the principle of postponement and two

¹⁰ This is the society regulated by the principles of justice.

principles regarding time (principles of continuity and of rising/declining expectations). We could name this group of principles ‘*dynamic principles*’.

Following the idea of Sidgwick (Sidgwick 1907, cited in Rawls 1999: 366) Rawls also introduces the rational plan of life chosen with deliberative rationality. This is the plan that would be decided upon as the outcome of thorough reflection in which “the agent reviewed, in the light of all relevant facts, what it would be like to carry out these plans and thereby ascertained the course of action that would best realise his most fundamental desires” (Rawls 1999: 366). This type of rational plan supposes that the individual makes a choice judging the different types of facts that would allow him to make a decision for which he ‘would not be sorry’ in the future.

On the grounds of the above premises, Rawls concludes that a person’s good is determined by “the rational plan that he would choose with deliberative rationality from the maximal class of plans” (p. 372). He claims that it is rational for the parties in the original position to want different types of primal goods (friendship, personal affection, social cooperation, etc.) because they are absolutely necessary for the formulation of any rational plan of life. Rawls concludes that the list of primary goods can be accounted for by “the conception of goodness as rationality in conjunction with the general facts about human wants and abilities, their characteristic phases and requirements of nature, the Aristotelian Principle, and the necessities of social independence”¹¹ (p. 381). Of course this conclusion, which proves the validity of the thin theory, gives us a *carte blanche* to use the full theory of the good and to imply all constraints of right to our rational plans of life. It follows that in the well-ordered society (according to the full theory) an individual’s concept of the good is guided by the principles of justice.

In Chapter IX, at the end of *A Theory of Justice*, Rawls presents *three main arguments for the congruence* between the concepts of right and good. According to the first of them, the citizens of the well-ordered society will have the desire to act

fairly, if they possess an effective sense of justice (p. 498). Once they fully accept the principles of justice as regulative for their personal constitution, it is safe to assume that they will hardly change their ends and start to act unjustly. Hence their rational plans of life will be subjected to the sense of justice. The ‘real problem’ of congruence is to provide arguments for a person who only accepts the thin theory of the good and does not implement the restrictions of right. In order to solve this, Rawls develops the second and the third arguments. They both illustrate the role of rationality in the matter of congruence.

According to the second argument, it follows from the Aristotelian Principle and its companion effect that participating in the life of the well-ordered society is a great good (p. 500). The companion effect states that the well-ordered society is the perfect base for the development of human abilities (this development is desirable according to the Aristotelian principle). But because this society is a social union of social unions and given the fact that “our potentials and inclinations far surpass what can be expressed in any one life, we depend upon the cooperative endeavours of others not only for the means of well-being but to bring to fruition our latent powers” (p. 500). It follows that as rational beings we will want to participate in the life of the well-ordered society and, hence, we have to accept the principles of justice on which this society is regulated. This argument supports our claim that the two moral powers – the ability to formulate and pursue a concept of the good and the sense of justice, are not only complementary, but also interdependent. *As rational beings we are willing to uphold our sense of justice.*

This idea is expressed in the third argument for congruence. According to it acting justly is something that people will want to do as free and equal rational beings. The desire to act justly and the desire to express our nature as free moral persons turn out to specify what is, in essence, the same desire. When “someone has the true beliefs and a correct understanding of the theory of justice, these two desires move him in the same way” – to uphold the two principles of justice (p. 501). I will name the latter argument for congruence ‘the argument of understanding’. It supports our idea of the ‘comprehensive level of rationality’, outlined above, according to which the original position and the whole theory of justice could serve as a source of persuasion for a certain shift in the views of a given rational recipient. Once we understand the theory we

11 It is beyond the limits of the present work to discuss these arguments. I will only mention Rawls’s definition of the Aristotelian Principle, because it is relevant to congruence. According to this principle: “other things equal, human beings enjoy the exercise of their realized capacities (their innate or trained abilities), and this enjoyment increases the more this capacity is realised, or the greater its complexity” (1999: 374).

will be persuaded as rational persons to uphold (or to develop) a sense of justice. The argument of understanding further upholds the existence of interdependence between the two moral powers.

The third argument for congruence brings up two new questions. *First* of all, is it a contradiction to this argument that sometimes our sense of justice could act against our good as rational beings? Rawls' astute and witty response is that "the loves that may hurt the least are not the best" (p. 502). In other words, as true friendship and affection could sometimes be connected with self-imposed emotional burdens and even sacrifice, so could be the bidding of the sense of justice, because its origin is a social origin (we should remind ourselves here of the morality of authority and the other stages of moral development). We should further remember that the sense of justice is a "desire to conduct oneself above all else" (p. 503).

The second issue concerns the case in which a person does not perceive it as valuable to uphold her sense of justice (and therefore it is against her good). Now, should the others oppress and make this person follow the principles of justice? Rawls' unsatisfactory answer is that the principles adopted in the original position are binding for everybody and hence for such individuals "their nature is their misfortune" (p. 504). This issue is connected with the question of pluralism, which is underdeveloped in *A Theory of Justice*.

The thesis established in this paper, according to which there is both interaction and interdependence between the two moral powers, is not equivalent to a presumption that our sense of justice stems from rationality. It is obvious, for example, from the morality of authority that the fundamentals of the sense of justice are derived not only from the rational selfishness of the child, but also from her (potential) ability to love. Rationality is the soil on which, with the help of love and friendship, the sense of justice is acquired. What we claim is that *the two moral powers support each other in the choice and compliance with the principles of justice*.

On the basis of the analysis of role of rationality in *A Theory of Justice* we could distinguish between three different levels on which it is implemented:

- 1.) It is an integral part of theory of good and as such it is significant for the formulation of the principles of justice.
- 2.) It is a necessary, but not a sufficient, basis on which we develop our sense of justice.

- 3.) It is implemented in the arguments, which prove that a concept of justice could be stable in the well-ordered society.

On each of these levels rationality and its corresponding moral power interact with reasonability and the sense of justice.

The student of rationality in *A Theory of Justice* must take into consideration the elaboration of this notion in Rawls' other works. The development of the concept of rationality is based on the gaps in *A Theory of Justice*. One of the prominent issues overlooked in this famous project is that if the two principles have a strong regulative role in the well-ordered society this will effectively undermine any opportunities for pluralism and diversity of views. And the last two follow from the principles of justice (1993: Introduction, p. xvii). In order to resolve this internal conflict Rawls restates his theory in *Political Liberalism* and *Justice as Fairness*. His main task there is to show how political liberalism is possible among the variety of reasonable comprehensive doctrines. The notions of rationality and reasonability are significant at all levels of the reformulated theory (for example the idea of the burdens of judgement, that follows from reasonability is in the core of the concept of pluralism) and, especially, for the problem of stability. The latter is upheld on the basis of the sense of justice and the overlapping consensus (1993: 163-68; 2001a: 180-98). The complexity of the analysis of the development of Rawls' views on rationality in his subsequent works raises further impediments to the attempts to achieve a thorough understanding of his project in order to establish an objective evaluation of it. This intriguing issue justifies a separate exploration.

References

- Barry, B. (1995) 'John Rawls and the Search for Stability', *Ethics*, 105, pp. 847-915. University of Chicago Press.
- Gauthier, D. P. (1986) *Morals by Agreement*. Oxford University Press.
- Hill, Th. (1994) 'The Stability Problem in Political Liberalism', *Pacific Philosophical Quarterly*, 75, pp. 333-352, Blackwell Publishers.
- Kukathas, C. and Pettit, P. (1990) *Rawls: A Theory of Justice and its Critics*. Polity Press.

- Kymlicka, W. (2002) *Contemporary Political Philosophy*. Second edition. Oxford University Press.
- MacIntyre, Al. (1985) *After Virtue*. Second edition. Gerald Duckworth & Co. Ltd.
- Matravers, M. (2000) *Justice and Punishment*. Oxford University Press.
- Matravers, M., Mendus, S. (2003) 'The Reasonableness of Pluralism', in C. McKinnon and D. Castiglione, *The Culture of Toleration in Diverse Societies: Reasonable Tolerance*, pp. 38–53. Manchester University Press.
- Mulhall, S. and Swift, A. (1996) *Liberals and Communitarians*. Second edition. Blackwell.
- Rawls, J. (1980) 'Kantian Constructivism in Moral Theory', *Journal of Philosophy*, 77, pp. 515–72.
- Rawls, J. (1993) *Political Liberalism*. Columbia University Press.
- Rawls, J. (1998) *Collected Papers*. Samuel Freeman (ed.). Harvard University Press.
- Rawls, J. (1999) *A Theory of Justice*. Revised edition. Oxford University Press.
- Rawls, J. (2001a) *Justice as Fairness. A Restatement*. The Belknap Press of Harvard University Press.
- Rawls, J. (2001b) *The Law of Peoples: with "The Idea of Public Reason Revisited"*. Cambridge, Massachusetts: Harvard University Press.
- Vallentyne, P. (1991) 'Gauthier's Three Projects', in Vallentyne (ed.), *Contractarianism and Rational Choice*. Cambridge University Press.
- Weber, M. (1996) *Protestant Ethic and the Spirit of Capitalism*, translated by Talcott Parsons. Roxbury Publishing Company.
- Wenar, L. (1995) 'Political Liberalism: an Internal Critique', *Ethics*, 106, pp. 32–62.

THE STATUS OF PHILOSOPHY – SOURCES OF THE SPINOZAN CONCEPTION ON THE FREEDOM OF PHILOSOPHY

Gabriela Tănăsescu

The Institute of Political Sciences and International Relations
of the Romanian Academy
e-mail: gabriela.tanasescu@ispri.ro

Abstract:

The paper aims to indicate the complex genealogy of Spinoza's conception of the autonomy of philosophy and its political status. In essence, the study highlights the main directions of the history of thought – Western, Hebrew, Arabic and Greek – which inspired Spinoza in his approach to the great dramatic nature of the “separation” of philosophy from religion.

Key words: reason, rationalism, philosophical rationality, liberation of philosophy.

The assumption of this study is that the Spinozan approach to the autonomy of philosophy and the argumentation of its political status is not only the culmination of the “free”, “rationalist cogitation” of Western modernity, but also the apogee of an intellectualist direction of Hebrew rationalism and of a significant influence of medieval Arabic thought¹. The study also starts from the assumption, located in the lineage of Leo Strauss' interpretation, that Spinoza's perspective on the status of philosophy was marked by a specific understanding, Jewish and Islamic, of the principle of faith and by a specific understanding and reporting to Revelation. As such, the study outlines the Jewish and Arabic succession through which Spinoza “discovered” and used Greek thought, mainly Platonic, in the effort to argue for the

autonomy of philosophy and to confer it a political status.

Libertas philosophandi

Spinoza is the author of the most radical endeavor to “liberate the philosophy”², to “separate” philosophy from religion and, so circumscribed, to set the issue of faith. Considered as one of the most dramatic episodes of the history of philosophy – but which, in Strauss' expression, helped to shape the “antagonism” that constitutes the “core” of Western spiritual and intellectual history³ –, Spinoza's approach of “separating” philosophy from religion theoretically has involved the annihilation of any possibility of synthesis between “being a philosopher” and “being a theologian”, “to be a philosopher” or to situate yourself in philosophical rationality is a condition equal in Spinoza only “to open up to the challenge of theology”. Whithal, Spinoza's “separation” of philosophy from religion and “hugging” of philosophy states, in an updated platonic manner, the “separation” of the intellectual condition of the “wise” from that of the “profane”⁴,

² My book, *Spinoza – rațiune și libertate*, contains the extended analysis of this theme.

³ Leo Strauss, “The Dialog between Reason and Revelation. Progress or Return?”, in vol. *The Rebirth of Classical Political Rationalism. An Introduction to the Thought of Leo Strauss*, Essays and Lectures by Leo Strauss, Selected and Introduced by Thomas L. Pangle, Chicago and London, The University of Chicago Press, 1989, p. 270.

⁴ Spinoza made a distinction between *prudenterum*, those “wise”, “cautious”, “few”, “rational”, “judicious” men, who have a “philosophical attitude of mind” and who “live for truth”, and *vulgus*, the “common people” who are “prevent[ed] [by prejudices] from giving their mind to philosophy”. See the *Letter 30, To the noble and learned Henry Oldenburg, from B.d.S.*, in *The Letters*, in vol. *Spinoza, Complete Works*, with Translations by

¹ I reassume here the hypothesis and one of the conclusions presented in my book *Spinoza – rațiune și libertate. Libertatem philosophandi*, București, Editura Institutului de Științe Politice și Relații Internaționale, 2010, which provides the basis of a future project.

one unable of intellectual connection to initiatory philosophical concepts and themes, but emotionally attached to the Scripture that “teaches only piety” and brings “very great comfort to mankind”⁵.

The “liberation” of philosophy – *libertatem philosophandi*⁶ – followed in *Tractatus Theologico-Politicus* the strategy of an “unprecedented” criticism⁷ of the Biblical texts and aimed at replacing

the role of religion in public life by philosophy, which meant implicitly the affirmation of the freedom of conscience and expression. “The grounding of freedom to philosophize by criticizing the religion”⁸ was realized by Spinoza using politics as an instrument⁹. The separation of philosophy from religion or the founding of the political status of philosophy occurred thus mediated, by a thematic adjoining of political to theological, a juxtaposition which, even from the title, induces the possibility to situate the political at the same level with the theological and to conceive of it as an autonomous field.

Although Spinoza’s conception on democracy, state and tolerance – which, actually, occupies a narrow place in the economy of *TTP* – is inestimable not only in relation to its epoch but also within the history of political thought, his major interest was to emancipate philosophy, to impose it as an area in which reason has “absolute freedom” and in which knowledge is guided by the “light of the mind” and by the search for truth¹⁰.

In this “liberation of philosophy” and its takeover of the role of theology Spinoza used as a tool the “separation” of reason from piety, the separation of those who choose science from those who choose faith – faith without reason, without free judgment; namely simple “obedience” – a separation aiming at the theoretical annihilation of theology and

Samuel Shirley, Edited, with Introduction and Notes, by Michael L. Morgan, Indianapolis/Cambridge, Hackett Publishing Company, Inc., 2002, pp. 843-844. This distinction also appears in the *Theological-Political Treatise’s* Preface as the difference between *Philosophe Lector*, the “learned reader”, the “rational man guided by reason” or by the “teachings of the light of reason” and which has a “liberal approach to philosophical questions”, and *vulgus*, “the multitude”, “the masses” that, being “not guided by reason”, are “victims of the emotional attitudes” and, ordinarily, misinterpret. See in this regard Benedicti de Spinoza *Opera quae quotquot reperta sunt recognoverunt J. Van Vloten at J. P. N. Land, Editio tertia, Tomus secundus, Hagae Comitum, Apud Martinum Nijhoff, MCMXIV, Tractatus theologico-politicus, Prefatio*, pp. 91-92, and Spinoza, *Complete Works*, ed. cit., pp. 393-394. Besides, the word *caute*, incised with a rose on Spinoza’s ring, signifies, as considers Yovel, the great intellectual solitude of those who, like a *marrano*, “possessed a secret metaphysical truth, the genuine key to salvation and the worthwhile life, which the *multitudo* would never grasp...” Therefore, *caute* means “not only prudence, but a sense of depth and intimacy of the rational truth, which can hardly be shared by the vulgar...”, a “rational wisdom” which “thus becomes esoteric”. See Yirmiyahu Yovel, *Spinoza and Other Heretics*, vol. I: *The Marrano of Reason*, Princeton, Princeton University Press, 1989, pp. 30-31.

⁵ Benedicti de Spinoza, *Tractatus Theologico-Politicus*, ed. cit., p. 250, XV; p. 257, XV; *Theological-Political Treatise*, ed. cit., p. 520, XV, p. 526, XV.

⁶ The title of Spinoza’s work is *Tractatus Theologico-Politicus, continens dissertationes aliquot, quibus ostenditur, libertatem philosophandi non tantum salva pietate et reipublicae pace posse concedi; des eandem, nisi cum pace reipublicae ipsaque pietate, tolli non posse*.

⁷ According to the anecdote reproduced by John Aubrey in *Brief Lives, chiefly of contemporaries, set down by John Aubrey, between the years 1669 and 1696*, when Spinoza’s *Tractatus Theologico-Politicus* first came out, in 1670, Hobbes – “a notoriously bold man”, in Leo Strauss’ words – said that “he durst not write so boldly”. See Edwin Curley, “‘I Durst Not Write So Boldly’ or How to Read Hobbes’ *Theological-Political Treatise*”, <http://umich.edu/emcurley/spinoza>. Nevertheless, Leibniz properly considered, already in September 1670, that Hobbes’s *Leviathan*, “a work monstrous even in what its title suggests... has also sown the seeds of that very smart critique which this bold man [*homo audax*] carries out against sacred scripture”. See “29. Leibniz an Jakob Thomasius, 23. September (3. Okt.) 1670”, Gottfried

Wilhelm Leibniz, *Sämtliche Schriften und Briefe. Reihe 2. Philosophischer Briefwechsel. 1663-1685*, Erster Band, ed. Akademie der Wissenschaften, Berlin, Akademie Verlag, 1923, p. 66. See also Edwin Curley, “*Homo Audax*: Leibniz, Oldenburg and the TTP”, *Studia leibnitiana Supplementa: Leibniz’ Auseinandersetzung mit Vorgängern und Zeitgenossen*, ed. by Ingrid Marchewitz & Albert Heinekamp, Stuttgart, Franz Steiner Verlag, 1990, pp. 278-279. Leo Strauss himself confessed that he has “leaned very heavily” in his analysis of “these things” on Spinoza because, beyond the criticism of biblical texts, in what concerns “the modern criticism of Revelation” and the “incomprehensibility of God’s essence”, Spinoza is the “man who tried to force” the limits of the modern dispute, being “the most extreme, certainly...” See Leo Strauss, *op. cit.*, pp. 267-268.

⁸ As Karl Löwith shown in *Gott, Mensch und Welt in der Metaphysik von Descartes bis zu Nietzsche*, Göttingen, Vandenhoeck & Ruprecht, 1967, p. 207.

⁹ See in this regard Steven B. Smith, “What Kind of Democrat was Spinoza?”, *Political Theory*, Vol. 33, No. 1, February 2005, p. 7.

¹⁰ Spinoza, *Theological-Political Treatise*, in *Complete Works*, ed. cit., p. 526sq, XVI; p. 523sq, XV. In the *Letter 30* Spinoza uses the expression: “to live for truth”. See *The Lettres*, in *Complete Works*, ed. cit., p. 844.

the denial of the legitimacy of its essential function: the revelation of the truth about God and world nature.

Spinoza reached ultimately to the annihilation of theology by denying the dependence of theology on reason and, as a consequence, by its complete separation from philosophy. Formally, the “separation” of philosophy from theology meant not to reduce theology to philosophy. Essentially, the “separation” was the camouflage for locating theology outside the criteria of logical legitimation, the correctness of thought, and denying it a rational foundation or validation of “intellectual knowledge”. The argument for which “neither is theology ancillary to reason nor reason to theology”¹¹ was used finally to conclude not only the impossibility of validating the fundament of theology by reason, but also on the irrelevance of validating it by faith.

Spinoza’s approach to the “liberation” of philosophy and the freedom of conscience – which, in the reception of his work, along with the issue of God’s nature, has surpassed all other problems – had as major lines of inspiration the Western-European “renewing” of the critical reflection of the XVIth and XVIIth centuries as well as traditional Jewish criticism¹², and that of *marrano* origin¹³. The

originality of Spinoza’s conception – the “system or the labyrinth of his thought process” – is derived from the elements and motifs of the critical thinking of these epoch and of the history of thought which were capitalized on by him to different extents. In this regard, as Stanislaus Dunin-Borkowski has argued, at the time that Spinoza’s system was in the process of elaboration, the ideas he developed were neither new nor original, but they became, through him, what in the history of philosophy is considered an original creation¹⁴.

Libertas philosophandi appears, therefore, as the culmination of a development of “free cogitation” which passed from the “free exam” to fideism, deism, pantheism and atheism, with all the cortege of attempts to impose specific ways of “philosophical religiosity” or “metaphysical mystics”. Within such a circumscription, the Spinozan conception of the “liberation of philosophy”, with its implicit conception of the freedom of conscience, appears as a unique and radicalized reflection on the spirit of the most “bold” critical considerations on religion and biblical texts. These critical considerations included the attempts at the theological and religious renovation of the Catholic Church, as in the Erasmian *Philosophia Christi*, as well as of the Lutheran reformed churches, Hobbes’s positivist attitude towards religion, and the nonconformist approaches that attempted to develop a “route of rationalism” undertaken, in the XVIth century, by Rabelais, Bonaventure Des Périers, Raimundo Sabunde, Montaigne, Jean Bodin, and, in the first half of the XVIIth century, Giordano Bruno, Galileo Galilei, Pierre Gassendi, Descartes (including the controversy generated by its “methodical doubt”), Saint-Évremond, Cyrano de Bergerac, and to the experiments carried out in the *London Royal Society*¹⁵. Also, Spinoza’s conception appears as a reflection of *marrano* thought – due mainly to Uriel da Costa and Isaac de La Peyrère – and, in general, of some Jewish and Arabic lineages.

¹¹ Spinoza, *Theological-Political Treatise*, ed. cit., p. 520.

¹² Of the Old Testament, the Talmudic, rabbinic and Kabbalistic literature, with reference to the Word of God, the Spirit of God and the divine presence, to the original substance, the eternal and infinite God (*En Sof*), who have competed in shaping a “metaphysical Hebraism” in his work. See in this regard, for example, from a rich bibliography which put in connection this theme with various concepts present in Spinoza’s thought, Warren Zev Harvey, “Spinoza’s Metaphysical Hebraism or Hebraic Metaphysics”, and Warren Montag, “‘That Hebrew Word’. Spinoza and the Concept of *Shekhinah*”, in vol. *Jewish Themes in Spinoza’s Philosophy*, edited by Heidi M. Ravven and Lenn E. Goodman, Albany, State University of New York Press, 2002, pp. 107–115 and pp. 131–147.

¹³ *Marranos* were the Sephardim from the Iberian Peninsula forced to Christianize under the threat of expulsion (after 1492), but have kept their Hebrew identity and continued to practice Judaism in secret. The term designated the “secret Jews”, the “crypto-Jews”, who did not cease to “meditate on Hebrew writings”, on the *Decalogue* in particular. They have replaced the cult with the study of the “Book” or of the Bible and the “irrational religious emotion of the cult” with the “rational fraternal ethics”. Yovel states that the experience of *Marranos* created the “tendency to prefer the inner heart to external works”, considered as “false, empty shells”, because in the *Marrano* families with a Judaizing history what mattered was what “could be felt with the heart”, the “direct

awareness of the spiritual truth”. See Yirmiyahu Yovel, *op. cit.*, pp. 25.

¹⁴ See in this regard Stanislaus de Dunin-Borkowski, “Benedict Spinoza”, in *The Catholic Encyclopedia*, vol. 14, New York, Robert Appleton Company, 1912, <http://www.newadvent.org/cathen/14217a.htm>.

¹⁵ References to the subject matter of this paper are in the first chapter, *Spinoza, praecursoris et contemporalis*, of the book *Spinoza – rațiune și libertate*, ed. cit., especially on the pages 15–101.

Occidentali et Hebraei Praecursatoris

From the perspective of Western thought, a substantial intellectual encouragement and source of inspiration for Spinoza was Erasmus' paradigm of the "Christian rationalization" of Christianity and theology through their synthesis with philosophy, which meant the comprehension in them of the values of human emancipation. In terms of the empowerment of philosophy, Erasmus was definitely important for Spinoza because he expressed a typical Renaissance understanding of the "return to the Greeks": not as the "revenge of Greek thought against Christian dogma", but as an attempt to trace the path of a philosophical-theological synthesis, desirable in the "tumultuous" context of the affirmation of a "worldview, absolutely metaphysical and even with mystical predispositions, which appreciated Plato not through the criticisms of the philosophers, but from the angle of cosmology and theology, and... which has woven in an overview of Platonic elements, elements of Greek cosmology and Christian mysticism, Homeric myths and Jewish Kabala, natural science and medieval scholastics"¹⁶. This eclectic context decanted two concepts which are detectable also in Spinoza's thought: reason and freedom. They were decanted, prior to their appearance in Erasmus' works, in Marsilio Ficino's book *Theologia platonica* which – through the concept of human self-knowledge, which was decisive to discover the divinity hidden in man – influenced Pico della Mirandola and his trust in man's power to define himself by the "reflex act of the intellect"¹⁷. By defining the human condition as a reflection of reason or of a free act of thinking, Pico della Mirandola already announced Spinoza's conception. Assuming the thesis that man consists in being his own act involves accordingly the assumption that

"philosophy itself" teaches man "to depend on his own conscience"¹⁸, because consciousness "purif[ies] our spirit" and prepares it to be "flooded with the light of natural philosophy" and, ultimately, "through the knowledge of divine things, to reach perfection"¹⁹. Man's freedom of thought and conscience, man's love for learning and searching the truth through reason is what gives "spiritual health", human superiority and perfection, and, as has been desired throughout the scholastic tradition, human responsibility. Spinoza radicalized Pico della Mirandola's thesis, specifically reflected in Erasmus, and argued for a mental attitude opened to philosophy relying exclusively on the values of secular humanism, which I think can appropriately be considered as being to some extent of Jewish extraction and placed in a Talmudic genealogy.

In this perspective it must be emphasized that Spinoza drew from "much older" sources, Jewish and *Marrano*, in many ways being a "logical culmination" of what already was asserted by illustrious predecessors²⁰. Among them, the most important "practitioners of the rationalist approach of the Hebrew philosophy" were Maimonides, Gersonides and Hasdai Crescas, with other medieval Hebrew philosophers whose heterodox ideas have had a significant impact on Spinoza²¹, particularly Solomon ibn Gabirol, Bahya ben Joseph ibn Paquda, Judah ha-Levi, Rabbi Abraham ben Meir ibn Ezra [during the pre-maimonidian period], Shem-Tov ben Joseph ibn Falaquera, Judah ben Abravanel (Abrabanel) sau Leo Hebraeus.

Wolfson has shown, in this respect and in a particularly convincing way, that the primary sources of Spinoza's philosophical knowledge and the "main trunk on which was grafted the entire philosophical knowledge that he acquired later" was Hebrew literature. These Hebrew sources were the

¹⁶ Erwin Panofsky, *Ideea*, translation in Romanian from German by Amelia Pavel, București, Editura Univers, 1975, p. 29.

¹⁷ Giovanni Pico della Mirandola, *Conclusiones sive theses 900: Romae anno 1486 publice disputandae, sed non admissae. De hominis dignitate*; in Romanian: G. Pico della Mirandola, *Raționamente sau 900 de teze propuse spre a fi dezbătute în public, la Roma, în anul 1486, dar neacceptate. Despre demnitatea omului*, Translation by Dan Negrescu, Introductory study by Gheorghe Vlăduțescu, București, Editura Științifică, 1991, p. 64. The 900 sentences were originally collected under the title *Conclusiones philosophicae, cabalisticae et theologicae*, with an introduction, *Oratio de dignitate hominis*, which became one of the most famous works of Renaissance character.

¹⁸ *Ibidem*, p. 135.

¹⁹ *Ibidem*, p. 126.

²⁰ See for this issue, among others, Steven Nadler, *Spinoza's Heresy. Immortality and the Jewish Mind*, Oxford, Oxford University Press, 2004, chap. "Immortality and Eternity"; Yirmiyahu Yovel, *Spinoza and Other Heretics*, vol. I: *The Marrano of Reason*, Princeton, Princeton University Press, 1988, chaps.: "Heretic and Banned", "Spinoza, the Marrano of Reason", "The Split Mind: New Jews in Amsterdam".

²¹ One of the "explanations" of this reference of Spinoza's to "direct and immediate" sources, but also of the mediated ones, is due to Austryn Harry Wolfson, "Behind the geometrical method", in vol. *The Philosophy of Spinoza*, Cleveland and New York, Meridian Books, The World Publishing Company, 1962, pp. 3–32.

“matrix” in which he formed his plan of general ideas – according to Wolfson, the other sources appear only as “introductions” –, the matrix from which he extracted the main illustrations and from which he mainly cites, but also that which challenged his first “doubts”²².

As pointed out by Steven B. Smith, Spinoza was not a faithful mediator, believer, or loyal transmitter of this tradition. Rather, he was a profound and far-reaching critic of it²³. I think that his Jewish heritage, notably the heterodox one, was probably Spinoza’s most important source of inspiration and, equally, of critical reflection. An overwhelming role in this respect was played by Maimonides’s thesis that the intellect is the highest good as a solution to overcome superstitions through reason, rational life and intellectual love of God. Maimonides attempted to show how a religion informed by rational thought can avoid the “Scylla of irrational faith and the Charybdis of rational unbelief”²⁴ – what Strauss called “impossible obedience and flagrant transgression”²⁵ – and what to Spinoza was the impossible. Consequently, Spinoza avoided Maimonides’ attempt to reconcile Aristotelian philosophy and science (“the dictates of reason”) with the teachings of the Torah and with the rabbinic tradition (philosophical tradition and biblical tradition) – in *Dalal al-Ha’irin* (in Arabic), *Moreh Nebukim* in Hebrew (*Stalker wandering*) – his allegorical interpretation of the sacred texts, his apophatic theology (“negative theology”) and his subsumed conception of the nature of God, the creation of the world and the free will. By the report of the philosophical tradition the biblical tradition is identified with an abstract vision and the universalistic tendency of ancient philosophy has as its counterpart the concrete, practical, communitarian thought of the Judaism until

Maimonides. The “Philosophical revolution” proposed by Maimonides targeted “mentalist Judaism”, a complementary application of the rabbinical model, of the reactionary, fundamentalist vision, and of a critical universalistic thought of Greek origin, starting not from the necessity of establishing the validity or non-validity of Aristotle’s thought, but from the Bible, from the possibility that the Bible “support such a theory”. Part of the profound reaction caused by the thesis of the divine intellect and of humanity being saved if it becomes an intellect, the thesis of religion as preparation for abstract philosophical thought, arises as the emergence of Kabala. With this, interpretation of the Bible appears as a mediatory between the “concrete, communitarian vision and a more theoretical, more abstract one”, as an “interrogation” which does not eliminate the preponderant character of [the “paramount importance” of] the concrete, the role of textuality on semantics or the allegorical thought of Christian origins²⁶.

But the “contradictions” and the “improbabilities” of the Old Testament analyzed by Maimonides generated in Spinoza “subtle doubts and incentives of the mind” and revealed his high propensity to “sensible issues” that can be interpreted, under the conditions of a sufficient knowledge and wisdom, only by “crossing the boundary” between explicit, detailed knowledge, and an allusive, indirect, insinuating one, implying, by a subtle way of instruction and of revelation, “truth”.

As a post-Maimonidian philosophy Spinoza could follow the attempt to reconcile philosophy with religion and to demonstrate that the Bible and the Talmud are in accord with philosophy (Shem-Tov ben Joseph ibn Falaquera). Spinoza was also interested in a number of ideas from the philosophy of of this time: a pleading for the study of philosophy, a desire for control of the body and soul, an analysis of the different degrees of human perfection (Shem-Tov ben Joseph ibn Falaquera), and a new neo-Platonic attempt to treat the Platonic ideals of beauty, wisdom, perfection and to define love in philosophical terms (ben Judah Abravanel).

Arabes et Graeci Praecursoris

²⁶ See on this topic the sketch of a refined introduction in the specificity of the report: philosophical tradition – Jewish tradition drawn by Moshe Idel [as part of the debate “Christianity and Judaism”], *Revista* 22, nr. 227, 31 iulie 2007, p. 7.

²² *Ibidem*, p. 12 sqq. According to Wolfson, „with the exception of the new sciences, his reading in Latin supplied him merely with a new vocabulary for old ideas” (p. 13).

²³ See Steven B. Smith, “How Jewish Was Spinoza?”, in vol. *Piety, Peace, and Practice of Esoterism*, edited by Paul J. Bagley, Historical Library, Volume 47, Kluwer Academic Publishers, 1999, p. 191.

²⁴ Steven Nadler, “Spinoza vs. Maimonides. For the Future of Judaism”, *Jewish Daily Forward*, published December 02, 2009, issue of December 11, 2009, www.forward.com/articles/120111/.

²⁵ Leo Strauss, “Literary Character of the *Guide for the Perplexed*”, in vol. *Persecution and the Art of Writing*, Chicago & London, The University of Chicago Press, 1988, p. 52.

Beyond Jewish influences, Spinoza also assimilated contributions from Arab thinking, as evidenced by the “striking similarities, down to the smallest detail” between his later thinking and that of al-Farabi’s²⁷ and his commentator Ismail Dahiya’s writings, as the link between Spinoza’s conception and “Ibn Tofail’s strange work, the story of *Hai Ibn Yoktan*”²⁸.

From the perspective of the Jewish sources of Spinoza’s conception of philosophy it needs to be pointed out that medieval and pre-Renaissance Jewish philosophy was concerned with systematically harmonizing the new ideas with the Rabbinic tradition which was marked by the “rediscovery” of Greek thought achieved through the Arabic translations and comments beginning in the IXth century. Having had access, in Hebrew, to Aristotle’s work incorporated in Averroes’s commentaries and to Plato through al-Farabi’s commentary, Spinoza had enough terms of comparison between “the three philosophical literatures” – Hebrew, Latin and Arabic. For Spinoza, then, from the perspective of a XVIIth century thinker, with a broad cultural propensity, the three literatures represented, as Wolfson highlights, a common tradition. Despite the differences, this common tradition reflected a purely theological and

dogmatic concern with a “common origin”: “Greek philosophy, in the midst of which stood Aristotle”. Wolfson showed in this regard that “the same Greek terminology lay behind the Arabic, Hebrew and Latin terminology, and [that] the same scientific and philosophic conceptions formed the background of all those who philosophized in Arabic, Hebrew, or Latin”, the three philosophic literatures representing “in fact one philosophy expressed in different languages, translatable almost literally into one other”²⁹. Beyond these terminological and ideatic correspondences between the philosophy of Arabic, Hebrew and Latin expression, there were at least two Arabic-Jewish “specificities” which influenced considerably Spinoza’s conception of philosophy: the principle of faith and the social and political status of philosophy.

It must be understood that the Arabic “mediation” between Greek philosophy and Jewish thought was possible because between Judaism and Islam there is a similarity regarding the principle of faith. Leo Strauss emphasizes in this regard that the fundamental difference between Christian scholasticism and Arabic and Hebrew philosophy is reflected in their literary sources and in their understanding of and relation to Revelation. Thus, to Jews and Muslims the Revelation “has the character of Law (*torah, shari’a*)”, of training” rather than Faith, “their reflections on Revelation” being “not a creed or a set of dogmas, but a social order, if an all-comprehensive order, which regulates not merely actions but thoughts and opinions as well. Revelation thus understood lent itself to being interpreted by loyal philosophers as the perfect law, the perfect political order”. In this sense, *falāsifa* [the philosophers, in Arabic] “attempted to arrive at a perfect understanding of the phenomenon of Revelation”³⁰, which, in the case of al-Fārābī, ended with the distinction – made for the first time in the Islamic philosophy – between philosophy, as source of truth and guide in life, and theology, as source of symbolic transmission of truth and, simultaneously, with the idea of the prevalence of reason over Revelation as source of truth.

From my point of view this statement is fundamental to assessing the Jewish vein and its resonance with the Arabic one in Spinoza’s philosophical thinking, whose metaphysics culminates in the “rational” rejecting of transcendence and in its “metamorphosis” in ethics

²⁷ The fundamental thesis of Abū Naṣr Muḥammad al-Fārābī (Alpharabius) (c. 870-950), founder of the Islamic Neo-Platonism, was “absorbed” in Islamic philosophy, but also in the Hebrew and Christian ones.

²⁸ Stanislaus von Dunin-Borkowski (1912), “Benedict Spinoza”, in *The Catholic Encyclopedia*, New York, Robert Appleton Company, retrieved January 17, 2003 from New Advent: <http://www.newadvent.org/cathen/14217a.htm>. Abu Bakr ibn al-Tufail’s novel, *Hayy ibn Yaqdhan* (in Latin, *Philosophus Autodidactus*; in English *The Self-Taught Philosopher or The Improvement of Human Reason: Exhibited in the Life of Hai Ebn Yokdhan*) – “The living, awakening son” –, inspired by the title of the novel written by Avicenna a century earlier (late eleventh century), is a *itinerarium mentis*, is an initiation story or the story of an “autodidact wild child”, of the “path of scientific knowledge and self-discovery”, of the “systematic process of discovering the truth through rational inquiry”. The philosophical themes of the novel, particularly the epistemological ones, are recognizable not only in Spinoza’s theory of knowledge and in his critique of religion, but also in his own way of life – of austerity, solitude, serenity – favourable to the search for truth. They are also recognizable in Western philosophy in Thomas Hobbes, John Locke, Isaac Newton, Jean-Jacques Rousseau, and Immanuel Kant in varying formulas and degrees. Spinoza encouraged the Dutch translation of the novel, its publication in 1672 being due to his friend Johannes Bouwmeester.

²⁹ Harry Austryn Wolfson, *op. cit.*, p. 10.

³⁰ Leo Strauss, Introduction, in *op. cit.*, pp. 10-11.

and whose biblical analysis culminates in the rejection of Revelation and in a (democratic) conception of the “good” state and the “best” order that makes freedom possible.

In Jewish and Islamic philosophy, the search – presumed by faith – for the perfect law and perfect political order was guided by “essential works”³¹, such as Plato’s *Republic* and *Laws*³². This fact involved a distinct content of Judaic and Islamic political philosophy as against Christian scholasticism and a change of its weight and importance in the whole of philosophy.

Through al-Fārābī – considered by Maimonides “the greatest among the Islamic philosophers and truly the greatest philosophical authority after Aristotle”³³ –, the whole of philosophy was presented through the medium of a political frame or construction, as far as both the study of God and of the universe, as well as the city – as these appear in the two parts of his most important work³⁴ – have as the final aim of their investigation “the essence of human perfection and happiness”³⁵.

According to al-Fārābī, happiness consists in a certain science and a certain way of life, the science in question proving to be “the science of the essence of every being” and “the art which supplies that science proves to be philosophy”, and the

corresponding way of life is that supplied by “what proves to be the royal or political art” in which the “philosopher and the king prove to be identical”. As such, the happiness produced is due to philosophy as “the highest theoretical art”, supplemented not by the religion or the Revelation, but by politics, in its Platonic sense, by the “royal art as the highest practical art”³⁶.

This “alliance” between the political and philosophical is in fact also Spinoza’s fundamental concern³⁷, the stake of the “liberation of philosophy” and the motivation of “Free State” from the *Theological-Political Treatise*, and of the “free man” – or the supreme intellectual freedom which makes possible the internalization of *natura naturanta* [of the supreme whole] and, through it, beatitude – from *Ethics*. The “liberation” that Spinoza has reserved for philosophy followed the path on which al-Fārābī “walked”, inspired in his turn by “Plato’s way”, namely the attempt to “combine” the “way of Socrates” and the “way of Thrasyarchus”. The “way of Socrates”³⁸ designated the option for the nonconformist, intransigent confrontation with the false opinions and the faulty lifestyles of the fellows in the city (the “conflict” with the crowd) that ultimately brought death, thus a “way” appropriate for philosopher’s relations with

³¹ In Christian scholasticism the place of essential works was occupied by Aristotle’s *Politics*, Cicero’s writings and the Roman law.

³² If in Western Europe they were “discovered” in the XVth century, they had been translated into Arabic in the IXth century and commented on by al-Fārābī.

³³ This was so on the strength of the extended commentaries on Aristotle’s work, especially on logic, and on the strength of his own logical system. He tried to “naturalize the philosophical learning of the Christians with Islam. Combining Aristotelian cosmology and psychology with the Neoplatonic metaphysics and a curious political Platonism, he made a unique contribution to the political thought of Islam”. See F. W. Zimmerman, *Al-Fārābī Commentary and Short Treatise on Aristotle’s De Interpretatione*, London, Oxford University Press, 198, p. IX, apud Derek E. Hines, “The Influence of Plato and Aristotle on Alfarabi”, p. 3. [dspace.nitle.org/bitstream/handle/10090/5295/Hines.pdf](https://www.dspace.nitle.org/bitstream/handle/10090/5295/Hines.pdf). Nevertheless, he has been much less known and valued in Christian scholasticism than Avicenna and Averroes.

³⁴ *āra ahl al-madīna al-fādila* (*Book of Opinions of the people of the Ideal City or The Virtuous City*).

³⁵ al-Fārābī refers to “the happiness of this world in this life”; “the ultimate happiness in the other life” appears as distinct in the second part (that devoted to Plato’s philosophy) of *The Virtuous City*, but it is irrelevant inasmuch as the immortality of the soul is not mentioned nor is the Platonic conception of life after death.

³⁶ Leo Strauss provides an excellent interpretation of the forked evolution of medieval philosophy, of a Judeo-Arab cultural and ideological framework, “perfect” for absorbing the foundation of Greek moral and political philosophy. Strauss focused mainly on the mediating role of the Arab al-Fārābī in this evolution, namely that of “processing” and commenting on the Platonic and Aristotelian work. See Leo Strauss, *op. cit.*, pp. 7–21.

³⁷ Unlike al-Fārābī, this was developed by Spinoza in two of his works written in the same period, *Theological-Political Treatise* and *Ethics*.

³⁸ Al-Fārābī made a distinction between “the way [the “science and art”] of Socrates” – which appears in the *Laws* and that is just a part of the “way” of Plato – i.e. the “scientific investigation of justice and the virtues” – and the other part, namely “the way [the “science and art”] of Timaeus” [obviously from the dialogue *Timaeus*, which proves to be “the art of Plato”] – “the science of the essence of every being”, “the science of the divine and of the natural things”. Both *Timaeus* and *Laws* belong to the category of scholarly dialogues, those of “old age”, in which Plato returns to his own doctrines developed in the dialogues of “mature” age, deepening and changing them. Only in the “late dialogues” is there an “eclipse” of Socrates. If in *Theaetetus* and *Philebus* Socrates appears in the consecrated role of “guiding”, in *Parmenides*, *Sophist*, *Statesman*, *Timaeus* he is no longer the central character or is just a “mute spectator”, and in the *Laws* is absent.

the elite. In a distinct manner, the “way of Thrasymachus”³⁹ represented the option for “justice” understood as the “advantage of the stronger” or of “mutual benefit” and was proper for a philosopher’s relation to the crowd.

These different “ways” reflect different attitudes towards the city, either non-compliance and death (the case of Socrates), or a compliance which ensures the safety and preservation of life. Al-Fārābī saw in Plato’s synthesis, in the “correction of the way of Socrates”, the path to avoid conflict with officials, with the crowd and, *in extremis*, to avoid the fate of Socrates. Plato’s option, as Strauss summed it up, involved in parallel the “provisional acceptance of the accepted opinions” or a conformation to them, including those opinions of the religious community, but also the “suggestion of opinions which, while pointing toward the truth, do not too flagrantly contradict the accepted opinions” and, especially, the “gradual replacement of the accepted opinions by the truth or an approximation of the truth”. The Platonic solution counted for al-Fārābī as the “replacement of the philosopher-king, who rules openly in the virtuous city, by the secret kingship of the philosopher who, being a ‘perfect man’ precisely because he is an ‘investigator’, lives privately as a member of an imperfect society which he tries to humanize within the limits of possible”⁴⁰.

Such an option assumed implicitly that the philosophers were not exponents of the society or of some parties, but of the “interests of philosophy”, which was animated, in fact, by the desire to “protect the highest interests of mankind”⁴¹.

The general character of the activity of the philosopher, of *falāsifa*, drawn by al-Fārābī by way of interpreting Plato’s synthesis reflected the precarious social and political status of philosophy in these societies – ancient and Islamic, contemporary of Plato – in which freedom in instruction (doctrinal freedom) and research did not exist, where there was, rather, a fear that philosophy can harm those whose faith was not strong enough. In this respect Strauss considered that al-Fārābī’s comments – included mainly in the work devoted to Plato and then reassumed in the writings of al-Fārābī’s disciples – indicated the causes of the need to dissimulate the “severe teachings” and to reassume the “antiquated or forgotten” distinction

between exoteric and esoteric: the lack of “harmony between philosophy and society”, the “non-recognition of the right of philosophizing” or the legitimacy of philosophy, the “grave danger” which threatens philosophy and philosophers.

The exoteric teaching of a *falāsifa*, most obviously and most commonly in a context dominated by the “lack of the freedom of philosophizing”, constituted the “exteriority”, the ‘armor’ that philosophy needed for political reasons. It represented the “form in which philosophy became visible to the political community”, “the political aspect of philosophy” or the “political” philosophy. Strauss considered it to be not accidental that al-Fārābī presented the whole of philosophy in a political framework and that the title “the two philosophies” [by which was known his treatise *al-Jam’ bayna ra’yay l-hakimayn* (*The Book about the Harmony in the Doctrines of the Two Philosophers: the divine Plato and Aristotle*)] suggests the difference between “the two doctrines”: the “interior” and the “exterior”. In this respect I think that al-Fārābī followed, as Maimonides would later, the Platonic principle – the “unwritten teaching”, namely the “unwritten doctrine”, stands for the fundamental metaphysical teachings transmitted “only to those who are trusted” and kept secret to the general public in order “not to be subjected to an indecent and degrading treatment”⁴² and not to endanger the philosopher.

In Spinoza’s work the precarious social and political status of philosophy is reflected in his historical perspective: the dissimulation of the truth, specific to “Plato’s way”, and the “secret kingship of the philosopher”, specific to the “way” of al-Fārābī, must be seen, in the manner in which Strauss did⁴³, through “the most important difference” between “Christianity, on the one hand, and Islam and Judaism, on the other” as indicating a difference in the status of philosophy. In Christianity philosophy became “part” of revealed theology, officially recognized as the preparation of those who studied the sacred doctrine. Therefore, in Christianity training in the domain of sacred doctrine involved a

³⁹ The content of this can be found in the *Republic*, one of the most important Plato’s dialogues of “maturity”.

⁴⁰ Leo Strauss, *op. cit.*, p. 16 sqq

⁴¹ *Ibidem*, pp. 17–18.

⁴² It must added that it was said of al-Fārābī that he read Aristotle’s *Physics* forty times (and two hundred times *On the Soul*) [See in this respect Ian Richard Netton, “al-Farabi, Abu Nasr”, www.muslimphilosophy.com/ip/.../H021.htm] was familiar with the statement of the most important witness of the existence of Plato’s “unwritten doctrine” [in *Physics*, 209 b].

⁴³ By assuming an approach subsumed to the sociology of philosophy.

considerable amount of philosophical competence. As a result, in the Christian world, the official recognition of philosophy implied also its ecclesiastical supervision. By contrast, in Judaism and Islam, the sacred doctrine, which represented at least in the first instance the official interpretation of Divine Law (*Talmud* or *fiqh*), “had much less in common with the philosophy.”

In fact, Talmudic competence could be acquired completely without philosophical instruction, so that even “Jews of the philosophic competence of Halevi and Maimonides took it for granted that being a Jew and being a philosopher are mutually exclusive”. In this context, “the status of philosophy was, as a matter of principle, much more precarious in Judaism and in Islam than in Christianity”, these differences explaining the “partial collapse of philosophic inquiry in the Islamic and in Jewish world, a collapse which has no parallel in the Western Christian world”⁴⁴. Strauss pointed out in this connection that in Islamic and Jewish societies, often associated in terms of their comprehensive regulation of morals, worship of the divine, the theater, etc., together with the “totalitarian society” of the Greek city, philosophy was essentially just a private and trans-political activity and the “philosophical life”, a “life of the hermit”.

The most eloquent example of the status of philosophy and of its reception was, according to Strauss, that the work of Maimonides, *Moreh Nebukim*, considered as the equivalent of Thomas Aquinas’s *Summa Theologiae*, was more greatly appreciated, even apart from the authority of Aquinas’s work, in Christianity. In contrast, the equivalent of the *Summa Theologiae* in Judaism “could be considered” the *Mishne Torah*, i.e. Maimonides’s work of coding the Hebrew Law. Also quite revealing in this regard is the difference between the beginning of the two philosophical works: the *Summa* begins with an interrogation of the need of sacred doctrine outside philosophical disciplines, i.e. a “justification of the sacred doctrine before the tribunal of philosophy”, while the *Moreh* begins with a “diffuse commentary on a Biblical verse” (Genesis 1. 27) which reflected “the urgency of the legal justification of philosophy, i.e., a discussion in legal terms of the question whether the Divine Law permits or forbids the study of philosophy, rather than a philosophic justification of

the Divine Law or of its study”⁴⁵. Moreover, as Strauss pointed out, Maimonides was also in agreement with the observation that the Hebrew tradition emphasized “God’s justice” and not “God’s wisdom”, and set as the Jewish equivalent of philosophy or theology certain elements of *aggadah* (or Legend, the “unwritten teaching”), part of Hebrew doctrine generally considered to be less formal and authoritative than the *halakhah* (the “written teaching”).

Unlike Maimonides, Spinoza asserted “bluntly” that the Jews “despise philosophy”⁴⁶, thus creating an oppositional relation: traditional Judaism *versus* philosophy, Jerusalem *versus* Athens. But unlike Erasmus, who motivated, through *Philosophia Christi*, both “Athens and Jerusalem”, philosophy and religion, Spinoza radically settled the solution exclusively on behalf of philosophy by transcending religion, both Jewish and Christian, and by secularizing their themes.

References

Curley, Edwin. “I Durst Not Write So Boldly” or How to Read Hobbes Theological-Political Treatise”, <http://umich.edu/emcurley/spinoza>.

Curley, Edwin (1990) “Homo Audax: Leibniz, Oldenburg and the TTP”, *Studia leibnitiana Supplementa: Leibniz’ Auseinandersetzung mit Vorgängern und Zeitgenossen*, ed. by Ingrid Marchewitz & Albert Heinekamp, Stuttgart, Franz Steiner Verlag.

Dunin-Borkowski, Stanislaus von (1912) “Benedict Spinoza”, in *The Catholic Encyclopedia*, Vol. 14. New York, Robert Appleton Company, retrieved January 17, 2003 from New Advent: <http://www.newadvent.org/cathen/14217a.htm>.

Harvey, Warren Zev (2002) “Spinoza’s Meta-

⁴⁵ *Ibidem*, p. 20. Strauss also states that the reasons invoked by Maimonides in order to prove that certain truths concerning the divine things must be kept secret were used by Thomas Aquinas to demonstrate that the rational truth about divine things is in need of being divinely revealed.

⁴⁶ In the *Theological-Political Treatise*, ed. cit., p. 503, XI, when he referred to the Apostles who preached to the Jews and have adapted “the novelty” of the Gospel’s doctrine “to the character of their contemporaries”, as the people who “despised philosophy”. Teaching “a religious doctrine free from all philosophic speculation”, “none of the Apostles did more philosophising than Paul, who was called to preach to the Gentiles”.

⁴⁴ Leo Strauss, *op. cit.*, p. 18 sqq

physical Hebraism or Hebraic Metaphysics”, in vol. *Jewish Themes in Spinoza's Philosophy*, edited by Heidi M. Ravven and Lenn E. Goodman (2002), Albany, State University of New York Press, pp. 107–115.

Idel, Moshe (2007) “Christianity and Judaism”, *Revista* 22, nr. 227, 31 iulie 2007.

Leibniz, Gottfried Wilhelm (1923). *Sämtliche Schriften und Briefe*. Reihe 2. Philosophischer Briefwechsel. 1663–1685, Erster Band, ed. Akademie der Wissenschaften, Berlin, Akademie Verlag.

Löwith, Karl (1967) *Gott, Mensch and Welt in der Metaphysik von Descartes bis zu Nietzsche*, Göttingen, Vandenhoeck & Ruprecht.

Mirandola, G. Pico della (1991) Raționamente sau 900 de teze propuse spre a fi dezbătute în public, la Roma, în anul 1486, dar neacceptate. Despre demnitatea omului, Translation by Dan Negrescu, Introductory study by Gheorghe Vlăduțescu, București, Editura Științifică.

Montag, Warren (2002) “‘That Hebrew Word’. Spinoza and the Concept of Shekhinah”, in vol. *Jewish Themes in Spinoza's Philosophy*, edited by Heidi M. Ravven and Lenn E. Goodman (2002), Albany, State University of New York Press, pp. 131–147.

Nadler, Steven (2004) *Spinoza's Heresy. Immortality and the Jewish Mind*, Oxford, Oxford University Press.

Nadler, Steven (2009) “Spinoza vs. Maimonides. For the Future of Judaism”, *Jewish Daily Forward*, published December 02, 2009, issue of December 11, 2009, www.forward.com/articles/120111/.

Netton, Ian Richard. “al-Farabi, Abu Nasr”, www.muslimphilosophy.com/ip/.../H021.htm

Panofsky, Erwin (1975) *Ideea*, translation in Romanian from German by Amelia Pavel, București, Editura Univers.

Smith, Steven B. (1999) “How Jewish Was Spinoza?”, in vol. *Piety, Peace, and Practice of Esotericism*, edited by Paul J. Bagley, Historical Library, Volume 47, Kluwer Academic Publishers, 1999.

Smith, Steven B. (2005) “What Kind of Democrat was Spinoza?”, *Political Theory*, Vol. 33, No. 1, February 2005.

Spinoza, Benedicti de. *Opera quotquot reperta sunt recognoverunt J. Van Vloten at J. P. N. Land*, Editio tertia, Tomus secundus, Hagae Comitum, Apud Martinum Nijhoff, MCMXIV, Tractatus theologico-politicus, Prefatio.

Spinoza (2002) *Complete Works*, with Translations by Samuel Shirley, Edited, with Introduction and Notes, by Michael L. Morgan, Indianapolis/Cambridge, Hackett Publishing Company, Inc.

Strauss, Leo (1988). “Literary Character of the Guide for the Perplexed”, in vol. *Persecution and the Art of Writing*, Chicago & London, The University of Chicago Press, 1988.

Strauss, Leo (1989) “The Dialog between Reason and Revelation. Progress or Return?”, in vol. *The Rebirth of Classical Political Rationalism. An Introduction to the Thought of Leo Strauss, Essays and Lectures by Leo Strauss*, Selected and Introduced by Thomas L. Pangle (1989), Chicago and London, The University of Chicago Press.

Tănăsescu, Gabriela (2010) *Spinoza – rațiune și libertate*. Libertatem philosophandi, București, Editura Institutului de Științe Politice și Relații Internaționale.

Wolfson, Austryn Harry (1962) “Behind the geometrical method”, in vol. *The Philosophy of Spinoza*, Cleveland and New York, Meridian Books, The World Publishing Company, 1962, pp. 3–32.

Yovel, Yirmiyahu (1989) *Spinoza and Other Heretics*, vol. I: The Marrano of Reason, Princeton, Princeton University Press.

Zimmerman, F. W. (1988) *Al-Fārābi Commentary and Short Treatise on Aristotle's De Interpretationibus*, London, Oxford University Press.

Discussion

Pro and Con Discussion Regarding the Tenets of the Hermeneutic Philosophy of Science (Continuation)

Since the late 1960s studies in the hermeneutic philosophy of science have been gaining currency. In looking at this development from a historical perspective, one can today recognize an independent philosophical tradition. Scientific research is committed to the universe of interpretative phenomena – so the tradition's guiding tenet goes – since it is itself a hermeneutic process. To put it in an extended formulation, those phenomena which philosophical hermeneutics unveils in the being of history, language, and art can also be identified in natural science as a particular mode of being-in-the-world. This is a radical claim that provokes a double confrontation. On the one hand, the hermeneutic philosophy of science opposes the mainstream philosophical picture of science which is spelled out predominantly in terms of objectivism, epistemological representationalism, and cognitive essentialism. On the other hand, a conflict with traditional philosophical hermeneutics is inevitable. On this traditional

enterprise, a constitutive view of interpretation might be integrated into a theory of scientific communication, but by no means into a theory of scientific research and knowledge. Due to this double confrontation, the hermeneutic philosophy of science brings into play several interesting debates of general philosophical importance. This tradition is a target of criticism from positions as different as neo-scholastic ontology and neo-positivist epistemology. The publication of Professor Arvin Voss's article documents our desire to initiate on the pages of the Balkan Journal of Philosophy an ongoing pro and contra discussion regarding the tenets of the hermeneutic philosophy of science. To be sure, Professor Voss's elegant criticism of the hermeneutic view about the status of science's theoretical objects as well as his rehabilitation of essentialism concerning the constitution of scientific knowledge will meet a counter-criticism devised by the exponents of an interpretative turn in the philosophy of science.

ON THE DIVERGENT BEING OF SCIENCE'S THEORETICAL OBJECTS

Dimitri Ginev

Sofia University "St. Kl. Ohridski"

Sofia, Bulgaria

e-mail: dimiginev@abv.bg

Abstract:

This paper differentiates between science's intentional theoretical objects and purely idealized theoretical objects. One identifies the former by inquiring into the functions they fulfill in a dynamic system (say, the system of chemical reactions in a metabolic pathway), whereas the latter get introduced by means of mathematical idealizations. Regardless of the epistemological differences, the existence of both types of theoretical objects is projected

upon possibilities that are to be appropriated in a research process. The paper addresses the potentiality-for-being of the intentional theoretical objects. Though the functions which such an object fulfills are throughout experimentally recognizable, its being never becomes transformed into an actually accomplished presence.

Key words: covariant realism; hermeneutic realism; cognitive existentialism; intentional theoretical object.

Robert Crease is a leading figure in the Continental philosophy of science. For many years I have enjoyed reading his excellent books and papers. It is my great privilege to be in a position of responding to his cogent and subtle comments on my position of cognitive existentialism. In this rejoinder I will start right away with a discussion of Professor Crease's conception of formal indication. It is this conception that provides him with a reference system for evaluating doctrines of cognitive existentialism.

I

According to Prof. Crease, perception does not constitute a privileged standpoint in our efforts to approach the status of scientific objects. Perception is always embedded in contextualized practices, and we are able to perceive something "to be real" if we find it to behave in a predictable way within a particular practical context. His version of hermeneutic realism clearly distinguishes phenomenological constitution from all kinds of (social, instrumental-technological, and cognitive) constructionism. At the heart of this version (called also "covariant realism") is the view that science's theoretical objects (like "electron" or "regulative gene") are invariant under transformation. (In fact, this view is a specification of the noetic-noematic correlation in transcendental phenomenology. One is to state that this specification extends phenomenology's constitutional analysis to cover the "scientific objects' life".¹ This is why it has still much to do with privileging consciousness's intentionality in the constitution of meaningful objects. Husserl's belief that through the method of "free variation" the unfamiliar object will be transformed into a familiar one also plays a prominent role in Crease's covariant realism.²) Sci-

ence's theoretical objects may look different from different perspectives (or, when they are involved in different contexts of doing research). Yet, with regard to each particular theoretical object, the descriptions of "what we see from these perspectives flow together in an orderly way when described by the right set of transformations" (Crease 2009, 227). Making use of covariance as a constraint (in a manner similar to the way in which a theory's basic equations get formulated in a covariant form) allows one to recognize the reality of science's theoretical objects.

Against this background, Crease goes on to rehabilitate early Heidegger's idea of formal indication as a "reckoning with a being-on-the-way". Of course, this idea belongs to the repertoire of Heidegger's hermeneutics of facticity, and it is originally formulated in a discourse devoted to pre-theoretical experience. Nonetheless, Crease is convinced that formal indication can be generalized as an approach to scientific entities qua objects that are in a state of being constantly interpreted and anticipated. What one gains by adhering to this approach is the chance to "make explicit the temporality of the grasp" of the objects. (I am not quite sure whether for Crease this temporality is tantamount to the temporality of scientific objects' ongoing constitution.) Conceptualizing the scientific object in a formally indicative manner – so the argument goes – offers "a path of research in advance". The object is never grasped as a pure presence, and hence, it is never totally given. It can be only grasped as having an infinite number of profiles that are potentially to be attained. Scientists hear the language of their discourse as indications allowing them to pick out phenomena which become bodily present in instrumentally mediated ways with possibilities of appearing otherwise given other mediations (Crease 2009, 228). The investigation of science's theoretical objects comes to the fore as an unceasing anticipation of phenomenal profiles that constitute the objects. No doubt, Crease's covariant realism based on the formally indicative approach to scientific objects

¹ The historical roots of this specification are to be traced in my view back to Fritz London's phenomenological approach to the ontology of the objects defined implicitly in the deductive theories. (Fritz London, "Über die Bedingungen der Möglichkeit einer deduktiven Theorie", *Jahrbuch für Philosophie und phänomenologische Forschung*, VI(1923), 335-384.) Kostas Gavroglu makes the case that many of the ideas elaborated by London in his later treatment of superconductivity and superfluidity as macroscopic quantum phenomena are informed by his early phenomenological wanderings. (Kostas Gavroglu, *Fritz London: A Scientific Biography*, Cambridge 1995.)

² Crease appeals also to Cassirer's celebrated conception that the perceptual world always possesses a structure that cannot be reduced to an aggregate of scattered sensations. He has good reasons to invoke that conception in which

the "tendency toward invariants" is attributed to the perceptive acts as their principal characteristic. Yet what Cassirer does not address in his conception is that the structural features of the perceptual world (like all structures formed in human existence) are fore-structured contextually and horizontally. Stressing this fore-structuring does not make the structuralist approach to perception meaningless. It only provides an argument for the priority of the hermeneutic approach in studying the constitution of the perceptual world.

puts into play a promising position in the hermeneutic philosophy of science. However, in the few papers devoted explicitly to formal indication as an approach to science's theoretical objects the position of covariant realism is still underdeveloped and its depiction suffers from conflating too many issues of different sorts involved in the attempts at reformulating the Heideggerian concept in terms of philosophy of science.³ Nonetheless, there are good reasons for Crease's insistence that the development of this concept would help one to resolve two aporias in the hermeneutic philosophy of science which cognitive existentialism does not succeed to cope with. The aporias concern respectively the appropriate way of addressing the autonomy of inquiry (more specifically, its irreducibility to changing configurations of scientific practices) and the account of how theoretical objects arise through practices (Crease 2012, 126). To be sure, in several respects Crease's approach might suggest better solutions than cognitive existentialism. In spite of this, I should like to express some disagreements with this approach.

It seems to me that formal indication serves the function of a regulative ideal in a sense close to that of German classical idealism. It finalizes "in principle" the research process which, in my view, is always in an open horizon. (I am saying "in principle" because for Crease the research process in each scientific domain is a potentially ceaseless process. Formal indication brings into play a Fichtean regulative ideal whose achievement is always an "unending task".) The formally indicative approach assigns an implicit teleology to the research process that is not congruent with the very idea of hermeneutic openness.

There are *ad infinitum* ways of contextualizing a theoretical object of science, but each of them is constrained by a (hidden) covariance that ideally regulates (thereby providing a formal indication to) the whole process of the object's contextual interpretations. Formal indication refers to a covariance that is steadily in an upcoming state. In each contextual interpretation of a theoretical object formal indication announces itself as a covariance that is to be expected. *Post festum* with regard to the accomplished interpretations the covariance might be re-

constructed as a group of transformations. Yet *ante factum* it is only recognizable as a "feeling of anticipation". This is why formal indication manages to temporalize the research process (as being predicated on contextual interpretations of theoretical objects). In the various historical situations the formal indication is characterized by different groups of transformations. But since the formal transitions from one to another group are always possible, formal indication works constantly against incommensurability implied by scientific changes. Crease treats the concept of covariance (and "covariant realism") as referring not to the particular groups of transformations but to the ongoing transitions. (If I correctly understand him, formal indication is to be unearthed by reconstructing the diachronic transformations of the groups of synchronic transformations. However, this formulation invokes a structuralist discourse that Crease hardly would accept.)

The regulative ideal of formal indication is a kind of idealization generated by the intrinsic dynamics of scientific research. Thus, the essential value of idealization gets rehabilitated in Crease's conception. Yet this is a quite peculiar rehabilitation. Crease does not regard idealization as an initial trans-empirical codification of possible empirical content that is liable to conceptual-explanatory and experimental articulation. This is why his conception fundamentally differs from the established idealizational conceptions of science.⁴ The regulative ideal cannot be disentangled from scientific practices and accordingly, stylized as a de-contextualized procedure that provides the absolute point of departure of scientific investigation. Thus considered, idealization does not contradict the tenets of hermeneutic contextualism. Nonetheless, putting the regulative-constraining ideal first (and thus, allowing the interpretation of science's theoretical objects to succumb to a kind of idealization) restricts the scope of the hermeneutic philosophy of science significantly. In focusing on the constraining structure achieved through idealization, covariant realism bears resemblance to structural realism (and other branches of scientific structuralism). Yet unlike the latter, the covariant realist does not confine the constraining

³ Without a clear differentiation between the issues which the formally indicative approach handles *en bloc* several undesirable ambiguities may come into play. Crease's confusion of the problematic of theoretical terms' semantic with the problematic of theoretical objects' ontological status is a case in point.

⁴ Thus, the conception worked out by Leszek Nowak and the other members of the Poznan school places emphasis upon the idealizing assumptions which supposedly simplify the picture of the investigated reality. Making these (counterfactual) assumptions explicit enables one to formulate procedures of idealizations that determine to a great extent the logic of doing research.

structure to the formal-conceptual organization of scientific knowledge, trying then to demonstrate that this structure characterizes reality itself. For Crease, the constraining structure as a regulative ideal takes place in the dynamics of scientific practices. The classical subject-object distinction does not presumably play any role in covariant realism. Thus considered, Crease's conception is by no means something like an epistemological generalization of Carl Gustav Jacobi's canonical transformation theory that aims at "deriving" the theoretical objects' status from studying the transformation properties from basic physical theories.

I think that Crease quite intentionally keeps intact a residual Platonism in his conception. Though he insists on the contextual existence of science's theoretical objects, he assumes (again in a classical transcendental-epistemological sense) the Platonic existence of ideal (de-contextualized and outside any horizon) indicators that operate in the real horizons and contexts, thereby making possible "references to what we encounter" (Crease 2012, 128). Paradoxically enough, de-contextualized indicators are liable for the contextualization of scientific objects. Yet the price for this "partial restoration of essentialism" has to be paid by making concessions to non-hermeneutic doctrines, and confusing hermeneutic-practical issues with epistemological-semantic problems. Indicative in this regard is Crease's confusion of theoretical terms with theoretical objects. As already mentioned, he even goes on to substitute the discussion of the objects' status with a discussion of the dual semantics of theoretical terms. By ascribing a kind of semantic parallelism (double meaning of each theoretical term – meaning in a theoretical framework and meaning in the laboratory), he tacitly (and in contradiction with the declared tenets of covariant realism) accepts a traditional epistemological framework.

According to the version of hermeneutic realism I support, science's theoretical objects are entities that exist in particular modes of being-in-the-world. Thus, the constitution of the world meaningfully within these modes has to be at stake in determining the theoretical objects' ontological status. Accordingly, the principled difference between covariant realism and hermeneutic realism consists in the difference between the figure of immanent transcendence and the figure of the regulative ideal. It is a difference that involves a contradistinction between the free choices of possibilities in an open horizon and the choices guided by an anticipation of what

will come into being that is constrained in a covariant manner. In the latter case, the range of possibilities that can be appropriated is "somehow" always already finalized.

The crux of the matter in Crease's conception is the ambiguity involved in the claim that "the real must not be the function of representation but of our experiences". If under "our experiences" he understands a back-and-forth between meanings embedded in a theoretical framework and meanings constituted in the laboratory, then, I am afraid, covariant realism would land in a Hacking-like interventionism. On the alternative I try to put forward, hermeneutic realism is the view that reality gets always revealed in characteristic hermeneutic situations whereby these situations are inherent to reality. All "experiences" and back-and-forth movements between theoretical and practical meanings get specified by characteristic hermeneutic situations (in the sense in which this notion is construed in my book *The Tenets of Cognitive Existentialism*). Furthermore, I do not agree with the claim that hermeneutic realism is not a constructivism "because the objects that we perceive to be real exhibit invariances that are not under our control" (Crease 2009, 224). This could be an argument against constructivism from the scientific structuralism's register, but not from that of hermeneutic philosophy.

In the remainder I will basically proceed by discussing a historical example whereby my aim will be to flesh out the claims already raised.

II

Imagine a biochemist in 1903 who is engaged with identifying an object called at that time an "enzyme". Let us empathically adopt her standpoint. Being educated (but also indoctrinated) within the established professional background knowledge, "our biochemist" (I will use this expression hereafter) attributes to the term "enzyme" a quite definite content. The neologism "enzyme" had been coined 26 years earlier by Willy Kühne. In fact, he coined this word as a result of a sharp criticism he leveled at the speculations on intracellular ferments. By introducing the neologism, he emphasized the essential difference between "organized ferments" (supposedly living microorganisms) and "unorganized ferments" (enzymes). On Kühne's view, enzymes are agents that can function outside the living cell. The possibility of intracellular enzymes was not ruled out

by him. The distinction between ferments (produced by the “living protoplasm”) and enzymes (nonliving substances) is still widely accepted in 1903. Though contradicting to a certain extent experimental results obtained in the last decade of 19th century, the professional ideology of theoretical biology at that time suggests that synthesizing and transporting materials are realized by biochemical reactions within cells but without any catalytic participation of “enzymes” in these reactions. (By “professional ideology” I refer to various suggestions inspired by different theoretical and non-theoretical motives, but not having the character of proven research results. The professional ideology is the system of beliefs of a certain *Denkkollektiv* in Ludwik Fleck’s sense.)

For analyzing the status our biochemist ascribes to this theoretical object, we have to take into consideration her historical situation. We have to ask in the first place the question of what she knows about “enzymes” at that time. For sure, she is acquainted with Edward Buchner’s discovery that the cell juice extracted from macerated yeast could ferment sugar to alcohol and carbon dioxide.⁵ Notoriously, this was the discovery of the soluble enzyme “zymase” responsible for a process of cell-free fermentation. She knows also that there is enough evidence for believing that this enzyme is also an active agent inside the living cell.⁶ Our biochemist is well acquainted with the theory of intracellular metabolism proposed by Eduard Pflüger in 1875. Yet this was only a chemical version of the protoplasm theory, and by no means an alternative to it. On the predominant belief (again as a part and parcel of the prevailing professional ideology), the attempts at extracting “unorganized ferments” from organs of the animal body that *in vitro* can bring into being the activities of the organs are doomed to failure. Despite this belief, in 1903 our biochemist realizes that the postulates of protoplasm theory are swept out of theoretical biology. (As a consequence, there is a growing conflict between the established professional ideology and new theoretical assumptions and experimental results.)

Our biochemist is looking for a theory that can explain chemically vital processes (digestion, metabolism, respiration, and assimilation) by ascribing

each chemical change occurring in the living cell to a specific intracellular enzyme (Kohler 1973, 185). She is sympathetic to the belief (typified most succinctly by Franz Hofmeister) that all vital reactions are mediated by intracellular enzymes). Our biochemist is enthusiastic not only about Buchner’s discovery, but also about some initial experiments giving preliminary evidence that intracellular enzymes carry out oxidation. She has heard (but still is not convinced of the information’s reliability) that Gabriel Bertrand had extracted a hydrolytic enzyme that hydrolyzes a substrate to an oxidizable form. She had also taken notice of the discovery of the first enzyme co-factor in 1897.

Quite recently (and almost by accident) she had read in *Berichte der deutschen chemischen Gesellschaft* an article written by Emil Fischer in 1894 in which the first structural explanation of how enzymes work was suggested. Based on her investigation about the reception of Fischer’s “key and lock model” (as providing this explanation), our biochemist knows that the model is not valid for all enzymatic reactions. Nonetheless, by employing this model enzymology (and more generally, biochemistry) gets a welcome explanatory independence. The development of the idea that the “geometrical structure of molecules” is responsible for the specificity of enzymatic catalysis encouraged research practices aimed at the elimination of the difference between the chemical activity of living cells and of chemical reagents (in particular, reagents distinguished by molecular asymmetry) (Futon 2006, 71).

Although our biochemist is still too much indebted to old-fashioned views related to the tradition of Liebig-style “physiological chemistry”, her novel views are slowly taking shape due to the outcome of a vigorous public controversy that took place in the years 1897-1902.⁷ At issue was the plausibility of the central dogma of protoplasm theory, according to which protoplasm (allegedly identified as a single molecule) carries out all key vital functions. (This controversy called into question not only the premises of protoplasm theory, but the professional ideology as well.) The exponents of protoplasm theory insisted that the active agent in yeast juice was not a specific enzyme but shreds of living protoplasm. Our biochemist continues to hesitate in 1903 as to whether to get rid of the view that (a) enzymes are located outside the cells, and (b) are able to catalyze only hydrolysis. Yet she regards the discovery of

⁵ On the reception of this discovery see Kohler (1972). The first oxidative enzyme had already been found (by Yoshida) before Buchner’s discovery.

⁶ On the profound impact of Buchner’s discovery on the development of enzymology see Bechtel and Abrahamsen (2007, 6).

⁷ See in this regard Kohler (1972, p. 327).

zymase as challenging the mainstream of research work in physiological chemistry. Our biochemist is highly interested in the ongoing research work of Sergei Vinogradsky and Martinus Beijerinck on microbial fermentations that provides arguments for dispensing with (a) and (b).

Furthermore, she had read in *Zeitschrift für physiologische Chemie* some reports on enzymes (probably intracellular) that catalyzed oxidations and syntheses. Some colleagues of her go so far to claim that there is a specific intracellular enzyme for each metabolic reaction. She is, however, skeptical about this radical assumption since nobody might verify it. Nonetheless she is enthusiastic about attributing to the enzymes (as the only entities that are the biochemistry's object *sui generis*) more significant physiological functions. She has good professional reasons for that: The success of the "physiological chemistry of enzymes" is the surest way to successful emancipation of biochemistry both as discipline and profession.⁸ She is proud of the appearance of the first true biochemical journal (*Biochemisches Zentralblatt*) one year earlier. Like her colleagues she looks at this fact (alongside the formation of the Departments for Biological Chemistry at Pasteur Institute and Lister Institute in London) as clear signs of a successful cognitive and social institutionalization of the discipline which is no longer treated as no man's land between chemistry and medicine, or as a trend within physiological chemistry.

In addition to the news related to an intrinsic transformation in the tradition of physiological chemistry our biochemist is literally bombarded with interesting news obtained in neighboring domains (the emancipating from cytology new domain of cell biology in the first place) that soon will prove their significance for various kinds of explorations in enzymology. Thus, she just had gotten the news that mitochondrion is a genuine organelle and not an artifact of technical devices. She follows with a great interest the ongoing work of organic chemists who had made headway by focusing on simpler compounds in studying metabolically important structures like hydroxyl, carboxyl, and amino groups.⁹ Our biochemist is surprised that the most vehement support for the existence of chemical agents that catalyze physiological processes outside living cells

has come from Emile Duclaux, Emile Roux, and other of the closest disciples of Pasteur. They acknowledge that "Buchner's enzyme" is the first to cause changes in the carbon chains and rearrangements of groups. (This looks most surprising in her eyes since Pasteur had been an orthodox adherent to the view that only the living protoplasm may catalyze physiological processes including fermentation.)

Two years ago our biochemist has begun to study Richard Neumeister's attempt at the generalizing of Buchner's idea. Neumeister suggested in 1897 that the fermentation was not effected by a single substance, but by a more complex set of cell constituents (Fruton 2006, 77). In 1901 Franz Hofmeister presented the initial version of the biochemical enzyme theory of metabolism. The community of our biochemist celebrates Hofmeister's initiative as the first genuine theory in biochemistry in which the concept of enzyme is introduced by definition. Closely related to that theory are Hugo Fischer's arguments (published in 1902) against the old-fashioned distinction between vital and enzyme action. Due to these arguments the view has gained currency in several leading laboratories that enzyme theory should entirely replace protoplasm theory. Yet the accumulation of new results and the resistance of old ideas have created growing crisis in the community's professional ideology. Our biochemist agonizes about the choice between neovitalism (as supported by theories of physiological chemistry that do not make room for the independent activities of particular enzymes) and chemical reductionism.

III

Against the background of this retrospection, I would like tentatively to categorize "enzymes" as species of an *intentional theoretical object*. By definition, an intentional theoretical object is an entity that presumably exists by fulfilling a class of functions that can be fully identified in their empirical details. The object *per se*, however, is never identifiable in its total being. It has always a potentiality-for-being. Only symptoms of its existence get revealed and represented in a situational-contextual-intentional manner. In scientific research such an entity gets originally introduced for the sake of filling an "ontic gap" in a domain that is in a process of theoretical and empirical articulation. This entity is constantly (intentionally) given with regard to the

⁸ See Kohler (1973); Fruton (1999, pp. 144-160).

⁹ See Holmes, F.C. 1992. *Between Biology and Medicine: The Formation of Intermediary Metabolism*, Berkeley, CA: Office for History of Science and Technology, University of California at Berkeley, pp. 54-58.

intended possibilities of its identification, reading and representation. It is inscribed on a (constantly transcending the research process's situatedness, though in each particular research situation finite) horizon of such possibilities.

Science's intentional theoretical objects have to be distinguished from purely idealized theoretical objects. The latter are introduced solely by means of mathematical idealizations. They do not refer to still "unknown existing entities" whose existence progressively might be specified, but to structures (and objects embedded in those structures) that in their ideality transcend the factuality of what is under empirical study, thereby acquiring a counterfactual status: In order to be empirically valid for all phenomena envisioned, delimited, and circumscribed by a theory (and to offer unique and unambiguous models), the theory's postulates must have a trans-empirical status. This is why the objects these postulates speak about are characterized by the same status. It is the covariance of the theory's formalism (as a mathematical codification of the postulates) that ensures the trans-empirical status. (In other words, the covariance guarantees a kind of morphism between theoretical models and data models.¹⁰) Working out mathematical idealizations (and idealized theoretical structures/objects) is a requisite for the introduction of covariance. Thus, for instance, the way in which one relates the metric of space-time continuum to the covariant form of Maxwell equations depends on how one idealizes the behavior of the electromagnetic field.

There is a long tradition (typified most clearly by Hermann Weyl) in theoretical science as well as in philosophy of science of relating covariance and the symmetry principles to the epistemological conditions and criteria of objectivity. As Brading and Castellani (2003, 18) argue, there is a simple rationale behind the efforts to connect spacetime symme-

tries specifically with objectivity: What is objective should not depend upon the particular perspective under which it is taken into consideration, i.e. objective is what remains invariant with respect to the transformation group of reference frames. Peter Kosso provides a more sophisticated argument for the connection between symmetry and objectivity. According to him, every symmetry principle is a principle of objectivity. He states that the argument is "not that values of properties are invariant and hence objective; rather it is relations among the values that are covariant and hence objective" (Kosso 2003, 418). To achieve this objectivity under covariance (as a paraphrase of "invariance under transformation") amounts to disentangling these relations from their contextual interpretations. Symmetry based on covariance is the most effective means of de-contextualizing entities that are mathematically constituted as objects (covariant relations among values) of research. Being thus de-contextualized, the theoretical objects' existence is entirely succumbed to laws written mathematically in covariant form. Yet this de-contextualization is (a) only a particular scientific practice (or a set of such practices, including practices of idealization) that (b) takes indispensably place within a broader contextual configuration of (experimental, formal, measuring, semiotic, instrumental, calculative, etc.) practices – in particular, a configuration of practices aiming at a contextual interpretation of idealized theoretical objects.

It is my contention that Crease's formally indicative approach has to be concretized with regard to the difference between intentional theoretical objects and purely idealized objects. It is worth mentioning that the former may or may not receive a de-contextualization through covariant mathematical interpretation. The intentional theoretical object "enzyme" as discussed so far received historically such an interpretation when covariant mathematical equations of enzymatic kinetics appeared for the first time. The subsequent linearization of these equations by means of various mathematical plots contributed additionally to this de-contextualization. From the viewpoint of the mathematical theory of enzymatic kinetics, "enzymes" acquired also the status of idealized theoretical object. Yet this status hinged heavily on their status as an intentional theoretical object, i.e. as an object that exists through its contextual interpretations within configurations of scientific practices.

¹⁰ On the standard version of the view of scientific structuralism (in particular, the positions of structural realism and semantic realism), the relation between the two types of models is a relation of isomorphism since only this kind of morphism is a map that preserves the relevant structure as it is revealed by the respective theory. Yet this purely mathematical criterion (in mathematics things are structurally represented by models only up to isomorphism) is too strong for empirical theories (even those in mathematical physics). This is why several authors (Brading, French, Ladyman, Bueno) have undertaken attempts at relaxing the relation of isomorphism in recent years. The general tendency in these attempts consists in replacing isomorphism with another type of morphism.

IV

Let us return now to our biochemist. After 1903 she will become involved (due to her wholehearted devotion to the new domain's articulation) in all of the main configurations of scientific practices in which "enzymes" will be further identified, read off, and represented (through conceptual, mathematical, semiotic, and experimental means) as species of a theoretical object *sui generis*. In a nutshell, our biochemist will participate in practices of material, structural, formal and functional identification of enzymes after the ultimate rejection of protoplasm theory. She will address in 1905 the problem of whether physiological synthesis is due exclusively to enzymes. From a historical perspective, it is this problem that promoted the need of enzymes' fourfold identification. The question we have to ask is whether the respective fourfold research process can be described in terms of Crease's covariant realism.

To begin with, in the first decade of the 20th century our biochemist will have been involved in practices of getting a quantitative picture of enzymatic kinetics, i.e. in the search for a formal identification of enzymes. In a full agreement with Crease's conception, the initial formal identification (i.e., identification in terms of quantifiable variables/parameters within a mathematical formalism) had been anticipated by Emil Fischer's work on the intermediate formation of enzyme-substrate complexes. It is the active mass of such a complex that determines the rate of the overall reaction. This observation opened the avenue to formalizable and quantifiable investigation of enzymatic kinetics. In other words, it was Fischer's approach to molecular compatibility within an enzyme-substrate complex that promoted to a great extent the search for mathematical plots of enzymatic kinetics. By the end of the first decade of the last century the main research interest was shifted to a certain extent from the protein structure of enzymes to the kinetics of enzyme-catalyzed reactions. The goal of the emerging search for enzymes' formal identification consisted in establishing predictive relationships between reaction velocities and experimental variables such as enzyme and substrate concentration. At stake was the velocity of these reactions. Attempts at the formation of enzyme rate equations (each of them viewed as a quantitative expression in concentration of reactant or product molecules in time) set the stage for mathematical modeling in enzymology. The most important outcome of these efforts was the

Michaelis-Menten formalism which still provides the standard mathematical apparatus in biochemistry.

Enzymes' material identification refers first and foremost to the proteins that can serve a catalytic function. In the second decade of the last century there began an intensive inquiry into the various intermediate molecules and enzymes that catalyzed each step in the sequences of possible metabolic pathways. Our biochemist will become familiar with practices of isolating the components of these sequences (enzymes, substrates, cofactors), practices of isolating biochemical intermediates, and practices of synthesizing unstable compounds for use in enzyme assays. During 1908-1914 she will be participating Otto Warburg's research team in studying the effect of cyanide and ethyl urethane on the oxygen uptake. This research work led to the material identification of the "respiratory enzyme". At stake was the possibility of deriving the enzymatic activity from the catalytic ability of inorganic lower oxidation. In searching for a material identification, our biochemist will be able to identify oxidases as enzymes that transport molecular oxygen. She will become convinced that the most important problem in enzymes' material identification is that one of how the *in vitro* experiments and solutions can be used in understanding the *in vivo* reality.

Enzymes' structural identification started when Franz Hofmeister enunciated in 1902 the view that the protein molecule is mainly built-up from amino acids. This view opened the avenue to investigating how the differences in amino acid composition of various proteins render them susceptible to degradation by the action of digestive enzymes of the gut. In the eyes of our biochemist the success of this investigation will be the birth of enzymology as a molecular science. By the end of the second decade of the 20th century she will participate in a research team studying enzymes' molecular structures through X-ray crystallography. Later on she will be interested in protein secondary structures. Finally, she will approach the enzymes' geometrical structures that control how the components of biochemical reactions interact. She will be looking for various spaces of representation of how the substrate molecule is complementary in shape to that of an enzyme's active site.¹¹

¹¹ Here I am using the expression "spaces of representation" in Rheinberger's (1995) sense.

Under “enzymes’ functional identification” I mean the precise differentiation of the functions they fulfill. To the initial functional identification of enzymes belongs the search for establishing to what extent the action of an enzyme is affected by the products of the reaction it catalyzes. In studying chemical mechanisms of enzyme-catalyzed reactions, our biochemist will elaborate on experimental practices of exploring an alternative metabolic pathway for a certain reaction, lowering the activation energy, and reducing the reaction entropy. In the 1930s these practices were of prime importance for enzymes’ functional identification, especially in the studies of how enzymes fulfill their functions through the energy changes in chemical reactions. Our biochemist will be among the pioneers who will study the regulatory and inhibitory functions of enzymes.

Was our biochemist’s research work in 1903 and afterwards guided by “formal indication” in Crease’s sense? Yes and no. Yes, insofar as a hidden covariance prompted a rapprochement of trends of inquiry that had been originated in different institutional places and oriented toward different research aims. No, inasmuch as this rapprochement was constantly accompanied by the emergence of research possibilities whose appropriation tolerated and stimulated practices, views, and ideas in each particular trend that were non-transferable and untranslatable in the other trends. Each of the trends of identifying “enzymes” as species of a theoretical object *sui generis* progressively widened its horizon of research possibilities, thereby deconstructing the existing covariance-as-constraint. What formally indicative approach does not take into account is the historical fact that the new possibilities arising out of the appropriation of existing possibilities of the theoretical object’s contextual interpretations constantly undo temporarily imposed constraints as related to a certain group of transformations. (I am not going to claim, however, that Crease’s conception precludes the opportunity of looking at formal indication as an interpretative process that deconstructs established conceptual, mathematical, experimental, etc. representations of a theoretical object.)

To sum up, there are in the beginning of the last century at least four “formal indications” of how to address the status of “enzymes”. Further research in enzymology (after the First World War) additionally pluralizes the formal indications. In the period 1920-1950 there are quite complicated processes of divergence and convergence among these indications.

Covariant realism is entitled to give account of the processes of convergence. But what about the processes of divergence (and even discrepancies) among the trends of (material, structural, functional, and formal) interpreting and representing enzymes? I am not able to figure out how covariance might cope with the fact that these trends are distinguished by divergent formal indications. The interpretative-ontological unity becomes threatened, if this intentional theoretical object would have been constituted meaningfully in interpretative contexts and configurations of practices that belong to divergent trends of inquiry.

In enzymology – to continue the discussion of the historical example – there appeared three strategies of overcoming (or at least, dealing with) the destructive divergences in the contextual interpretations of “enzymes”. In fact, these were three types of reaction to the object’s interpretative-ontological “demolition” by scattering its existence in an increasing number of interpretative contexts and configurations of scientific practices. These reactions are (i) the “reconceptualization” of enzymology, (ii) the creation of interfield theories which bridge two domains of research, and (iii) the splitting up of domains. All of the three types of reaction took place in the post-war research of enzymes. I will briefly address these strategies/reactions since, in my view, they are not on a par with the formally indicative approach to science’s theoretical objects.

The strategy of creating interfield theories in theoretical biology was observed and described for the first time in the mid-1970s. In Darden and Maull (1977), interfield theories function to provide a solution to a characteristic type of theoretical problem of how to construct explanatory models for the relations between the domains. What these authors suggest is an epistemological approach that is to be complemented by a practical-hermeneutic approach to the interfield theories. Cognitive existentialism treats these theories as theoretical practices that transmit in reciprocal manner and recast meaningfully theoretical objects (and other “epistemic things” in Rheinberger’s sense) existing in the established domains, thereby constituting intermediary objects and epistemic things. Thus considered, the interfield theories promote new interpretative contexts of reading and representing the intermediary objects. Prominent examples for such theories in the post-war research of enzymes are the thermodynamics of enzyme-catalyzed reactions, the theory of the role of active site flexibility in enzyme catalysis (the

theory which transmitted and re-constituted theoretical objects between molecular biology and biochemistry), and the theory of allosteric regulation.

It is the discrepancies between various lines of representing and reading enzymes that make the ultimate identification of enzyme as an object problematic. These discrepancies are the main reason for the under-determination of enzymes' ontic status. The efforts to eliminate such discrepancies provoked the second reaction called by William Bechtel (1984) "reconceptualizations" of enzymology.¹² (In a case study Bechtel manages to demonstrate how reconceptualizations in the sub-domains of vitamin research and respiratory metabolism facilitated the discovery of the connection between the B vitamins and the respiratory coenzymes.) The splitting up in sub-domains is a strategy that takes place when tenacious (and deteriorating the domain's image) controversies remain unsettled for a long period. In biochemistry this strategy is typically illustrated by the controversy between Peter Mitchell's chemiosmotic theory of the electron transport in respiratory metabolism and the chemical theory about this metabolism.

I am not quite sure whether covariant realism can be extended to cover the strategies of reconceptualizations, the creation of interfield theories, and the domain's splitting into sub-domains. The doctrine of covariant realism presupposes a kind of research process's linearity (corresponding to a kind of linearity in the theoretical object's interpretation and representation) that gets violated in these strategies which unavoidably leads to an "ontic pluralization" of the theoretical object's existence. To cut a long story short, the post-war multifarious (and multi-linear) research of enzymes is much better characterized not by appealing to a single formal indication but to a multi-directional appropriation of possibilities.

Finally, let me express my impression that covariant realism is a too optimistic version of hermeneutic philosophy of science. For adjudicating the justifiability of this impression I will first enumerate the necessary conditions which Crease states for establishing the reality of an intentional theoretical object. On the first condition, the real is what scientists can confidently and even unavoidably reach. The second condition refers to the discordance in-

herent in each present contextual interpretation of a theoretical object. In each context of inquiry the object is "not fully grasped, with hints of another, deeper order just over the horizon." (Crease 2009, 227) According to the third condition, the reality of a theoretical object "cannot be the function of a single theory, or one historical moment" (Crease 2012, 129). In their unity these conditions have to provide a sufficient ground for a successful identification of the theoretical object's reality. Now, I should like to show that even when the three conditions are fulfilled the theoretical object may prove to be a "non-real entity".

I will again proceed by means of a historical example.¹³ The intentional theoretical object I am going to discuss is the "high-energy intermediates of oxidative phosphorylation". In trying to explain what this expression means, I will start with the last component: Oxidative phosphorylation is a highly efficient metabolic pathway of releasing energy in which the mitochondria use their structure, energy, and enzymes to reform adenosine triphosphate (ATP) as the central unit of energy in the cell. ATP is the molecule that supplies energy to metabolism. The energy for the synthesis of ATP is released stepwise via an electron transport chain. A central problem for more than twenty years in biochemistry after the Second World War was the transfer of energy from the electron transport chain to ATP. According to the dominant research program of bioenergetics in the 1950s the energy is transferred from molecule to molecule. The adherents to this program hypostatized the existence of unidentified molecules with high-energy chemical bonds. These high-energy intermediates received at that time the status of a quite important intentional theoretical object. Its identification in the next two decades "would be the central task of the field" (Allchin 1997, 88). Prominent biochemists tried to achieve an interpretation of that theoretical object by means of experimental and instrumental practices of tearing apart the mitochondrion, isolating its components, and throwing them back in a test tube. These practices were subjected to the dogma of "scalar chemistry" – what matters is only the magnitude and not the direction of the reactions.

Within the configuration of scalar chemistry's practices several research possibilities about the object's existence became projected. In the early 1960s possibilities were appropriated that suppos-

¹² On Bechtel's approach, a reconceptualization consists in re-ordering the descriptive and theoretical apparatus used in characterizing the specificity of a domain of scientific research.

¹³ In what follows I draw on Allchin's (1997) case study.

edly specified properties of the high-energy intermediates. Theoretical and laboratory practices of making and proving predictions supported belief in the existence of this theoretical object. For instance, a line of such support was based on tracking radioactive atoms from molecule to molecule through reactions. Another cluster of projected research possibilities were related to the techniques (in particular, by using inhibitors) of forcing certain metabolic pathways to flow through what could be only the intermediates. Allchin convincingly argues that in this period the experimental foundation for believing in the existence of the high-energy intermediates of oxidative phosphorylation was quite rich: One believed that the intermediates exist but because of the complexity of the pathway leading to ATP their reality is not yet identifiable. Leading biochemists announced in the mid-1960s that these intermediates are no longer hypothetical entities. Yet one is to recognize that “all the evidence amassed for a high-energy intermediate molecule was only for concluding that there was an intermediate energy step of some kind” (Allchin 1997, 90). The intermediates had never been found within a certain configuration of practices, though various projected possibilities had been appropriated. They became progressively effaced. Today one is to conclude that the high-energy intermediates of oxidative phosphorylation do not possess reality but they still exist as inscribed in a horizon of possibilities. (The *belief* in their existence is intact in a small dissident group of the scientific community.) Appropriating these possibilities and projecting novel possibilities in new configurations of practices might change the state of affairs with their identification. The incremental disappearance (or better, effacement) of an intentional theoretical object is to be interpreted in terms of cognitive existentialism – an initiative that I will not, of course, undertake here.

Because covariant realism is a too optimistic position regarding the interpretation and identification of science’s theoretical objects, I am a little bit skeptical that the approach of formal indication might address properly the problematic of this effacement. In fact, Crease’s three necessary conditions for establishing the reality of an intentional theoretical object seem to be fulfilled in the historical existence of the object “the high-energy intermediates of oxidative phosphorylation”. For a long period, biochemists believed that these entities can be confidently reached. The research process was constrained by the formal indication of supposed

properties and functions of the intermediates. The laboratory inquiry was guided by an anticipation of getting details about these properties and functions. The discordances in the particular contexts of reading and representing the high-energy intermediates became surmounted due to theoretical and experimental corrections in the interpretation that approximated the regulative ideal. In all configurations of scientific practices devoted to the identification of these entities biochemists were able to reason empirically about the specificity of the intermediates. In a full agreement with covariant realism, they managed to show that the intermediates’ reality gets proved by triangulating from numerous experiments. Surprisingly enough, biochemists cultivated their belief that experimental results confirm this reality by advocating otherwise different theoretical positions. Nonetheless, at the end of the day “the high-energy intermediates of oxidative phosphorylation” proved to be (to use Allchin’s suitable expression) a “twentieth-century phlogiston”.

References

- Allchin, D. (1997) “A Twentieth Century Phlogiston”, *Perspectives on Science* 5, 81–127.
- Bechtel, W. (1984) “Reconceptualizations and interfield connections: The discovery of the link between vitamins and coenzymes”, *Philosophy of Science* 51, 265–292.
- Bechtel W. and Abrahamsen, A. (2007) “In Search of Mitochondrial Mechanism: Interfield Excursions between Cell Biology and Biochemistry”, *Journal of the History of Biology* 40, 1–33.
- Brading, K. and Castellani, E. (2003) “Introduction”, in Brading, K. and Castellani, E. (eds.), *Symmetries in Physics*, Cambridge University Press, 1–18.
- Crease, R. (2009) “Covariant Realism”, *Human Affairs* 19, 223–232.
- Crease, R. (2012) “Theory and Theoretical Objects in an Existential/Hermeneutic Conception of Science”, *Balkan Journal of Philosophy* 4/1, 121–130.
- Darden, L. and Maull, N. (1977) “Interfield Theories”, *Philosophy of Science* 44(1977): 43–64.
- Fruton, J. (1999) *Proteins, Enzymes, Genes: The Interplay of Chemistry and Biology*, Yale University Press.
- Fruton, J. (2006) *Fermentation: Vital or Chemical Process?* Brill.

Gavroglu, K. (1995). *Fritz London – A Scientific Biography*, Cambridge University Press.

Holmes, F.C. (1992) *Between Biology and Medicine: The Formation of Intermediary Metabolism*, Berkeley, CA: Office for History of Science and Technology, University of California at Berkeley.

Kohler, R. (1972) “The Reception of Eduard Buchner’s Discovery of Cell-Free Fermentation”, *Journal of the History of Biology*, 5/2, 327–353.

Kohler, R. (1973) “The enzyme theory and the origins of biochemistry”, *Isis* 64, 181–196.

Kosso, P. (2003) “Symmetry, objectivity, and design”, in Brading, K. and Castellani, E. (eds.), *Symmetries in Physics*, Cambridge University Press, 412–424.

London, F. (1923) “Über die Bedingungen der Möglichkeit einer deduktiven Theorie“, *Jahrbuch für Philosophie und phänomenologische Forschung*, VI, 335–384.

Rheinberger, H.-J., (1995) “From Microsomes to Ribosomes: ‘Strategies’ of ‘Representation’”, *Journal of the History of Biology*, 28/1, 49–89.

Book reviews

Modern World Issues from the Perspective of Anthropology

[(For Process Antropologia și problemele lumii moderne (*Anthropology Faced with the Problems of the Modern World*), by Lévi-Strauss, Claude, Preface by Maurice Ollender, Translation by Giuliano Sfichi, Iași, Polirom Publishing House,]

Marțian Iovan

Professor PhD

“Vasile Goldiș” Western University, Department of Social Sciences and Humanities, Arad, Romania
e-mail: miovan@uvvg.ro

Hailed as the father of modern social anthropology and as a major exponent of structuralism in the twentieth century, being part of the same family of thinkers as Michel Foucault, Roland Barthes, Louis Althusser, Jacques Lacan, Julia Kristeva, Edmund Leach etc., Claude Lévi-Strauss asserted himself, during 1950-1970, as the leader of this family of intellectuals, as the patriarch of French structuralism, whose name is featured in a veritable *pantheon* of social sciences (Eribon 1990: 91–98). In fact, structuralism is a current of thought that emerged in the early twentieth century, in language sciences, with the publication, by the Swiss scholar Ferdinand de Saussure, of his *Course of General Linguistics* (1909), in which language is envisaged as a general system of signs, based on interdependence, in which any change may cause changes to the whole. The structural and systemic approach spread, then, to numerous social and human sciences, including anthropology. Claude Lévi-Strauss is a continuator of linguistic structuralism, capitalizing on an inclination he had since childhood, as he confessed in an interview with Adelbert Reif: “From childhood I was bothered by the so-called irrational and I have always tried to find an Order beyond what we are told is Disorder.” In all his research and analyses, he looked systematically for Order.

His work, entitled *L’anthropologie face aux problèmes du monde moderne* (Paris: Éditions du Seuil), translated into Romanian and published, over

the same year, by Polirom, Iași, pertains to this cultural and scientific context, updating and synthesizing the great themes of his previous works, including: *The Elementary Structures of Kinship* (1949); *A World on the Wane* (1955), *Structural Anthropology* (2 vols., 1958, 1973); *The Savage Mind* (1962), *Mythologiques* (4 vols., 1964, 1966, 1968, 1971).

This work of Lévi-Strauss is actually the result of three conferences held in Tokyo in 1986, following the invitation of the Ishizaka Foundation, which were converted into three chapters of the volume, entitled: *I. The end of the cultural supremacy of the West; II. Three major contemporary issues: sexuality, economic development and mythical thinking; III. The recognition of cultural diversity.*

In this work Lévi-Strauss aims to explore “how anthropology regards the fundamental issues faced by humanity today” (p. 11), how these issues are formulated by anthropology from its distinctive perspective and what it is that sets apart its way of seeing contemporary world problems, in the hope of making them clearer. In his investigations, Lévi-Strauss makes a synthesis of all the main topics related to society, which had preoccupied him throughout his lifetime, from the standpoint of a new form of humanism, for anthropological structuralism is a “new democratic humanism” whose geographic area comprises all of the inhabited earth, and its method reunites procedures pertaining to all forms of knowledge: humanities and natural sciences.

Thus, anthropological humanism goes beyond traditional humanism in every way. All previous versions of humanism had been created for the privileged, as all came from privileged civilizations. Yet, anthropology proclaims, with multidisciplinary arguments, a humanism that is twice universal: first, it is the postulate of a new humanism that “nothing that is human can be alien to humans” (p. 62), and second, it uses methods and techniques borrowed from all sciences, and puts them to the service of human knowledge, thus calling for humanity’s reconciliation with nature, within the framework of generalized humanism. Regardless of how it is proclaimed, anthropology always aspires to the knowledge of the *total human*, based on her *productions* and *representations*.

From the position of this new anthropological humanism, Lévi-Strauss asks a series of questions: What does the “end of Western cultural supremacy” consist of, in a global society where everything seems to indicate that it tends toward a world civilization? How will cultures evolve in relation to races (to the genetic stocks of some populations), ideologies and religions? What dimensions will the development and progress of modern societies have from the standpoint of anthropological humanism? Can we apply judgments of value to past or contemporary cultures and to their directions of development? What should the new generations of governments, who are in a position to choose between different experiences, patterns, economic, social, cultural, ideological, or religious paradigms, do in order to ensure system-level balance or an optimal condition in the relationship between unity and diversity, between structure and the concrete phenomena through which it is manifested? What could the modern world learn from the experiences and lives of ancient societies (those that do not have writing and mechanical technologies), knowing that they have a history of hundreds of thousands of years (perhaps even two million years), while modern societies emerged three to four centuries ago? How can anthropology contribute to solving the great issues of contemporary society, such as economic, demographic, and ecological imbalances, artificial procreation, the “explosion of ideologies and integritisms”, the coexistence and clash of religions, the links between scientific thinking and mythical thinking? etc.

Through its specific methods anthropology can make original contributions to the answers to these questions, first and foremost by providing a better

understanding of these issues. This is because anthropology is a subject that is dedicated to studying the “human phenomenon”, particular facts, the lifestyles of savage peoples and human societies, through specific methods, processes and techniques, in order to discover their common denominator, the constants of the human condition. Observation, comparison, ethnographic survey, studying traditional, archaic societies in different ways “provides the only means to understand how people lived together for a historical period that might be equivalent to 99% of the entire duration of human life and, geographically speaking, until recently, on three quarters of the Earth’s inhabited surface” (p. 30). The anthropologist discovers the roots from the depths of the “unconscious” side of individuals and groups, the customs, traditions, attitudes and moods of archaic societies, the matrix that holds groups together and is expressed by strings of individual behavioral manifestations that seem to have no links between one another.

Anthropology assumes ambitions which give it its originality in treating the key issues of postmodernity. First, anthropological research seeks to reach objectivity (p. 41). Such objectivity rises above the specific values of society and the researcher’s social background, but, especially, rises above her method of thinking, in order to reach valid formulations from the point of view of any possible honest and objective observer. The second ambition of anthropology is to reflect the entirety, in the sense that social life can be conceived as a system whose elements are all organically related. In order to know a social phenomenon, one must look beyond it and identify the structure that generated it, the common form of social experiences, the invariant properties that lie at the foundation of the various types of social life. Thirdly, the anthropologist aims to find complete objectivity, at that level at which phenomena keep some meaning for individual consciousness (p. 47), a meaning in the experience lived by the subject.

Analyzing the specific problems of modernity, Lévi-Strauss argues that anthropologists will be able to bring an original contribution, starting from the fundamental distinction between two ways of social life: a traditional and archaic way of life, which is characteristic of authentic societies, and recently emerged forms (over the last centuries), which incorporate written communication, mechanization, various modern industries, but from which the first type is not missing. From this perspective, the West-

ern model of culture and civilization, forged at the beginnings of modernity and regarded as a prototype of perfection, justice and universal happiness of the people, has become “obsolete” in the twentieth century. While Western society has gained enormous power in relation to the environment and to nature, owing to the development of natural sciences and especially to increases in productivity, the very same society gave rise to perverse, harmful problems and realities, such as pollution, depletion of resources, the enormous concentration of population in urban agglomerations, which imposes an artificial and dehumanizing existence, the creation of an overwhelming bureaucracy, the genocide of wars, gulags, the explosion of “integrisms”, etc., threatening the survival of the human population. If Western society has lost its model, is anthropology able to provide compensating solutions? Lévi-Strauss is optimistic, believing that anthropologists have the mission to demonstrate that the values in which Europeans believe are not the only possible ones, “that other lifestyles, other value systems have enabled and continue to enable human communities to find happiness. Anthropology urges us to put some of our vanity aside, respect other ways of living, reexamining ourselves through the knowledge of other traditions that excite our surprise ...” (p. 64). In other words, a society or government should not believe that its institutions, habits and beliefs are the only possible ones, or that they are the best. The lesson taught by anthropology is that one can learn from others, without damage to one’s dignity and feelings of personal pride.

A good example of how anthropology can provide a foothold for solving the problems of contemporary society is the issue of finding remedies to sterility, which has become acute and obsessive in Europe, the U.S. and Australia, in the last 50 years. The development of genetics and medical science has made it possible to discover ways to assist procreation or obtain it artificially. Thus, hundreds of thousands of couples in which one or both partners are infertile can have children, using various methods such as artificial insemination, womb rental, egg donation, *in vitro* fertilization, embryo freezing etc.. With this, the problems that occur are no longer technical, but social, legal, moral, and psychological. Who will have parenting rights over children resulting from such manipulations, both *de facto* and *de jure*? Situations created by medical and genetic manipulations can be highly complex, and entail unprecedented challenges for Western civilization. For

example, a child may have a mother and two fathers, another child could have two mothers and a father, another has two mothers and two fathers, and yet another may have two fathers (when the biological father and the legal father are not one and the same man) and three mothers (egg-donating mother, carrying mother and legal mother).

Outlooks on parentage, in the case of assisted procreation, are very different, and the legislations of European countries are in disagreement on the rights and duties of social and, respectively, biological parents. Lévi-Strauss, the patriarch of social anthropology, shows that anthropology can provide support for answers to such legal, social and moral issues. For example, insemination with a donor has its equivalent in Africa, in the *Samo* people of Burkina Faso; post-mortem insemination in the *Nuer* people of Sudan; couples of two women practice assisted procreation in order to have children, in the *Yoruba* people of Nigeria etc. Such parentage relationships, present in some archaic societies, anticipate modern techniques and legal regulations. In these societies, there is no conflict between biological procreation and social parenthood. They give priority to the social side, with no clashes occurring between biological and social parenthood in the mindset of the group or individuals.

Therefore, human experiences present in archaic societies are extremely wide and varied. From their analysis, one can single out those that may be considered as constant “universals” of human nature. On this basis, anthropologists may suggest the frameworks in which some yet uncertain developments will take place.

Another problem that preoccupied Lévi-Strauss over several decades concerns relationships, affinities between mythical and scientific thinking, knowing that archaic societies are guided by myths, beliefs, and religious concepts, in different combinations, and associated with rites, resulting in a certain coherence of community life. On the other hand, modern societies no longer abound in myths. “To solve the issues raised by the human condition and natural phenomena” Claude Lévi-Strauss argues “they turn to sciences or, more precisely, for each type of problem, they turn to a specialized scientific discipline” (p. 133). But it is not always so, for both myths and science have answered a perennial human need – that of knowing the world, society and the meaning of one’s life in this world. The place and functions of myths were taken, later, by history, in its broadest sense, so that everything that myths had

done for societies without writing, history will do for modern societies: legitimize the existing social order, explain why society is as it is, justify the state of current societies through historical measures, foreshadow the future in the light of its knowledge and evaluation of the past, along with the present. But how much of it is mythical knowledge and how much of it is an objective reflection in the mind and texts of historians? The anthropological perspective invites us to think carefully and critically, as there can be no absolute interpretation of the historical past, but several relative ones. Therefore, scientific knowledge, as was designed from the beginning of modernity, generating objective and absolute truths, is a very particular and limited case. Rather, it is not the contradiction between mythical thought and scientific thought that is valid, but the emergence of the former with respect to the latter. We cannot rule out, today, Lévi-Strauss concludes, the possibility that, in the future, scientific thinking and mythical thinking would become ever closer, the development of one depending on the evolution of the other.

In conclusion, this mature work of Lévi-Strauss is a synthesis of all his works and creations, putting in a clear light the paradigm of structuralism as a theory and a method, a democratic humanism that is also a research practice, as an attraction to theoretical, explanatory models and the ambition to conquer objectivity in knowledge, as a code of social existence and heuristic approach to knowing the truth (Paul Droit 2009). As an anthropologist, Lévi-Strauss was associated with a social orientation in the field, succeeding in using structuralism both as a theory and as a method, as a model for giving personal meaning to social life and the world. The harmonious blending of conceptual and sensitivity-based knowledge, in the case of structuralism, the insight into the infinite variety of phenomena, in order to identify, in the abstract plane, the deep structure, the order that is hidden from our view, in a

kind of “subconsciousness” of the life of social groups, to find the system that interconnects the various components that we perceive as chaotic. From this point of view, the work of Lévi-Strauss, as Maurice Ollender said, “is now a laboratory of thought, open to the future” (Ollender 2011: 7), in which answers can be shaped to the crucial issues of society at the beginning of the third millennium.

The structuralism of Lévi-Strauss, synthetically expressed in this book, can be deemed an innovative work, which should be a reference frame for any new anthropological project. This is not because Lévi-Strauss was the only one who studied the structure and order behind individual phenomena in society, but because of the originality of the consequences he reached (Pauillon 1973). If so, “this is because, unlike the ‘theories’ of yesteryear, structuralism sought to reconcile, in the words of Lévi-Strauss, ‘the sensitive and the intelligible, art and logic’” (Izard, Lenclud 1999: 646). This is why the readership of Lévi-Strauss’ work is still substantial, owing especially to the renewing perspective he used in his research.

References

- Lévi-Strauss, Claude, and Didier Eribon (1990) *De près et de loin*. Paris: Éditions Odile Jacob.
- Droit, Roger – Paul (2009) Claude Lévi-Strauss, anthropologue, père du structuralisme. *Le Monde* 4 November.
- Ollender, Maurice (2011) Avant-propos. In: Claude Lévi-Strauss, *L’anthropologie face aux problèmes du monde moderne*; pp. 5-8. Paris: Éditions du Seuil.
- Izard, Michel and Gérard Lenclud (1999) Structuralisme. In: P. Bonté and M. Izard (eds.), *Dictionnaire de l’ethnologie et de l’anthropologie*; pp. 644–647. Paris: Presses Universitaires de France.