Studia Semiotyczne — English Supplement

Volume XXIX

Editor-in-Chief of *Studia Semiotyczne — English Supplement*
Tadeusz Ciecierski

Assistant Editor
Krystyna Komorowska

Warsaw • 2017

Prace redakcyjne (redakcja merytoryczna), skład elektroniczny, korekta językowa native speakera „Studiów Semiotycznych. English Supplement” są finansowane w ramach umowy 614/P-DUN/2016 ze środków Ministra Nauki i Szkolnictwa Wyższego przeznaczonych na działalność upowszechniającą naukę.

ISSN: 2451-2745

Proofreading:
Christopher Moore
Typesetting:
Michał Zawidzki

.

Studia Semiotyczne (*Semiotic Studies*) is a journal founded in 1970 by Jerzy Pelc, who was its Editor-in-Chief up until 2015. Between 1970 and 2015 *Studia Semiotyczne* was published non-periodically. In December 2015 *Studia Semiotyczne* was formally transformed into a six-monthly publication simultaneously in print and on the Internet. The journal publishers papers that fall on the borderline between philosophy and semiotics and in particular, in the fields of analytic philosophy of language, general theory of signs, applications of semantic methods in philosophy, philosophical aspects of linguistics and psycholinguistics, semiotic aspects of philosophy of mind, philosophical consequences of metalogic and the theory of formal languages, analysis of the language of philosophy and the philosophical argumentation, and history of ideas in semiotics and logic. *Studia Semiotyczne* is published by *The Polish Semiotic Society*. The guidelines for submitting manuscripts, the legal and ethical aspects of the publication and the review procedure are described in the "For Authors" section of the website:

`http://studiasemiotyczne.pts.edu.pl/english/`

Contents

Stanisław Krajewski ***Andrzej Grzegorczyk (1922-2014)*** **5**

Jerzy Pelc ***Barbara Stanosz (1935-2014)*** **39**

Cezary Cieśliński ***Paradoxes of Barbara Stanosz*** **48**

Paweł Grabarczyk ***Directival Theory of Meaning Resurrected*** **62**

Filip Kawczyński ***Two Models of Propositional Structure*** **82**

Bartosz Wcisło, Mateusz Łełyk ***Strong and Weak Truth Principles*** **107**

Olga Poller ***All The Superhero's Names*** **127**

Krzysztof Posłajko ***Semantic Deflationism, Public Language Meaning and Contextual Standards of Correctness*** **159**

Natalia Karczewska ***Evaluative Adjectives – an Attempt at a Classification*** **180**

Adrian Ziółkowski ***The Reference of Proper Names, Semantic Intuitions and Experimental Philosophy*** **201**

Joanna Odrowąż-Sypniewska ***Are There Any Subsentential Speech Acts?*** **248**

Piotr Konderak ***The Conscious Semiotic Mind*** **272**

STUDIA SEMIOTYCZNE—ENGLISH SUPPLEMENT, vol. XXIX (2017), 5–38
ISSN 2451-2745
DOI: 10.26333/stsen.xxix.01

Stanisław Krajewski[1]

Andrzej Grzegorczyk (1922-2014)

Andrzej Grzegorczyk died on 20th March 2014. Even though he was 91 years old, his passing was unexpected as he had been doing creative work almost until his last moment and his intellectual capacities seemed intact. His memoirs emphasized that he had been the previous century's last world-renowned representative of Polish logic. While this is true, it seems much more important that he was a very unique person, academically and socially active, but also a free spirit who chose his own path.

I. Life

Andrzej Grzegorczyk was born in Warsaw on 22nd August 1922, the only son of Piotr Grzegorczyk – a Polish studies specialist stemming from an intelligentsia family from the Polish Galicia, then part of Austro-Hungary – and Zofia, a doctor born in the landowner family of Zdziarski from the vicinity of Płock. Her background was strongly leftist; her brother Mirosław Zdziarski, a known communist and a member of the Communist Party of Poland, was sentenced to death in Russia in 1937. Andrzej Grzegorczyk spent his entire life, with only short breaks, in Warsaw. As a child, he attended a private Catholic school of the educational society "Przyszłość" (whose other fledgling was Władysław Bartoszewski) and from 1938 – the Władysław Reytan state secondary school. After the outbreak of war, when his previous school organized clandestine secondary school lessons, he decided to return there and passed his school-leaving exams in 1940, on the day of France's capitulation. To avoid being taken away to work in the German Reich, he

[1]University of Warsaw, Faculty of Philosophy and Sociology.
E-mail: stanislaw.krajewski@uw.edu.pl

enrolled at a chemical secondary school and after that, when the Germans allowed for vocational training, to a (intentionally) three-year chemical school located on the grounds of the Warsaw University of Technology, in which the university's Polish professors taught. He attended the school in the years 1942/1943 and 1943/1944, while simultaneously studying physics in clandestine classes at the University of Warsaw, taught (among others) by Czesław Białobrzeski and Leonard Sosnowski. Earlier, starting in autumn 1940, he had also started attending clandestine philosophy classes at the University of the Western Lands. He listened to lectures by Władysław Tatarkiewicz, who used to read chapters of his book *Analysis of happiness*, then in writing, to his students, Fr. Jan Salamucha (student of Stanisław Leśniewski), who taught logic, Fr. Piotr Chojnacki, the psychologist Fr. Mieczysław Dybowski and Mieczysław Milbrandt, who taught history of contemporary philosophy. He also visited lectures by Bogdan Suchodolski, Michał Walicki and others. He owed a lot to tutorials in logic by Henryk Hiż at the Philosophy Department, University of Warsaw. As he once said, "On the whole, the intellectual life of the capital's intelligentsia was very abundant, considering the reality of the occupation."[2] He could not, of course have ignored the reality of war: he took part in the Warsaw Uprising and escaped the Old Town through sewer channels with his colleagues from the battalion "Gustaw" (E.3 in bibliography).

He graduated after the war in Cracow, obtaining a master's title in philosophy for his thesis *Ontologia właściwości* (*Ontology of properties*), supervised by Zygmunt Zawirski, which transferred Leśniewski's ontology to a higher logical type. The ontological construction of Kotarbiński's ontological reism (propagated by Henryk Hiż) was tested by Andrzej Grzegorczyk in various contexts.

In the years 1946-48 he worked in Warsaw as an assistant to Władysław Tatarkiewicz. He was also secretary of the journal "Przegląd Filozoficzny" edited by Tatarkiewicz. After that, he obtained a doctoral scholarship in logic and mathematics. Those were the times when "the political situation favoured staying in the safety of logical and mathematical speculations." For instance, when he submitted a paper to a philosophical conference in

[2]This quotation, as all the others in this paper (except for those taken from the publication commented on in the given fragment), come from conversations with Andrzej Grzegorczyk and his notes that he made available to me (for better coherence, I sometimes reformulate them in the third person). I am also very thankful to the wife of this article's protagonist, professor Renata Grzegorczyk, for all her help. It should also be emphasized that a chief part of this text is taken from earlier articles about Andrzej Grzegorczyk authored (and co-authored) by me. They are all listed in the bibliography.

Amsterdam, he, not unlike a few other Polish philosophers, did not get a passport.

In 1950, Grzegorczyk received a doctor's title at the University of Warsaw for his dissertation *Przestrzenie topologiczne w bezpunktowych algebrach topologicznych* (*Topological spaces in pointless topological algebras*) (he passed his doctoral exam on 26th May 1950). Andrzej Mostowski was his supervisor. Grzegorczyk specialised in logic, but when taking the doctoral exam in mathematics, he took chemistry as his secondary subject. After obtaining the doctorate, he started work at the Institute of Mathematics, at the Polish Academy of Sciences (PAN). There, after three years, he successfully underwent the qualifying procedure for a senior lecturer position. The grounds for the procedure was his booklet *Some Classes of Recursive Functions* (A.2). In 1961, he received the title of professor extraordinarius and in 1972 – professor ordinarius. After March 1968, for his oppositional activity, showing in "signing all open letters that got to him," he lost his position at the University of Warsaw, where he had worked additionally for a few years, and stayed only in the Institute of Mathematics, PAN. In the 1960s, he became head of its Department of Foundations of Mathematics; the previous head, his teacher (and mine, for that matter) Andrzej Mostowski, limited himself to the chair of Foundations of Mathematics at the University. In 1973, he organised a Logical Semester in the newly opened Stefan Banach International Mathematical Centre, which brought together a few dozen well-known scholars from abroad to Warsaw. For many logicians of Poland and the countries of the block of "people's democracies", this was a rare opportunity to meet outstanding Western logicians.

In that period, Grzegorczyk's scholarly interests shifted significantly towards philosophy. As a consequence, he moved, after gaining approval of the Academy's administration, from the Institute of Mathematics to the Institute of Philosophy and Sociology, PAN, in 1974. In this institute, after its reorganisation in 1982, he chaired the Ethics Laboratory. He "did not make attempts for any higher position." In 1990, slightly early, he retired. After that – notably, already in the time of political freedom – he became a more active organiser in the Polish philosophical community. From 1995 to 1997 he directed a grant "Stulecie szkoły lwowsko-warszawskiej" (a "100 years' anniversary of the Lwow-Warsaw school"), which involved organising a number of meetings and lectures, including a large conference which took place in Lwow and Warsaw on the 100th anniversary of Kazimierz Twardowski's appointment to the Head of department at Lwow, as well as numerous publications (including Grzegorczyk's book A.15 in Ukrainian).

Cooperation with Ukraine and Russia was a vital part of Grzegorczyk's scholarly activity. In this respect, he seems quite peculiar as Polish scholars go: his academic ties with the East were as close as with the West. Even though he visited western countries (e.g. in 1965 he worked half a year in the Netherlands and in 1970 – a few months in Italy), he always felt better in Russia, where – as he claimed – the way of experiencing the world was similar to the Polish one.

Grzegorczyk actively participated in academic conferences all around the world. He took part in the famous conference *The theory of models* in Berkeley in 1963 and in most congresses of Logic, Methodology and Philosophy of Science, starting from the 1950s. He was an assessor in the Council of the Division of Logic, Methodology and Philosophy of Science, International Union of History and Philosophy of Science. Starting from 1995, he was head of the Editorial Council of "Przegląd Filozoficzny". From 1999 to 2003, at an advanced pension age, he was Head of the Committee of Philosophical Sciences of the Polish Academy of Sciences. From 1979, he was a member of an international philosophical organisation (Institut International de Philosophie), a fairly prestigious community, towards which he was, however, a bit critical, noting that that its activity encompasses mainly "keeping up its prestige." Until the very end of his life, Grzegorczyk was intellectually active. For example, he was on the editorial board of the bimonthly magazine "Bunt młodych duchem" ("Rebellion of the young at heart"), whose authors are mainly – as he himself admitted – the "old at body." He received two honorary doctorates: at the University of Clermont-Ferrand (2010) and the Jagiellonian University (2013).

In 1953, Andrzej Gregorczyk married Renata Majewska, who later became a professor at the Faculty of Polish Studies, University of Warsaw. They have two children and six grandchildren.

He took up logic thanks to the "radio lecture by Jan Łukasiewicz, popularizing logic, on consequence in the ancient Greek Stoic logic." He also liked geometry proofs and was fascinated by the so-called proofs for God's existence. Issues of formal logic and set theory became "a mania, an addiction, a drug" to him – something which never changed. However, he always considered them to be based on and mainly applicable to philosophical problems.

II. Accomplishments in logic

Andrzej Grzegorczyk may be called – as he did call himself – a philosopher, logician, methodologist and ethician. He was also a writer and – in a pretty non-standard sense of the word – a social activist. His books and papers were published not only in Polish, but also in English, French, Russian, Czech and Ukrainian. His most substantial achievements – ones recognised both in Poland and worldwide – were in mathematical logic. He believed them to be inseparable from a philosophical motivation; the formal results, in turn, motivated the worldview.

A. Computability and decidability

For Grzegorczyk, researching the computable processes, even in the form of idealised creations known as recursive functions, was as much as investigating the substantial, empirical, "palpable" aspects of the world expressed in a mathematical form.

1. Recursive functions

The notion of effectiveness became better understood in the 1930s thanks to the works by Gödel, Church, Turing and Kleene, but "in the mid-20th century it was still considered mysterious." Andrzej Grzegorczyk's contribution to the theory of recursive functions is of historical significance. In a widely cited paper *Some Classes of Recursive Functions* (published independently as A.2), he described and examined a sequence of classes of numeric functions obtained from certain source functions (which contained addition, multiplication, exponentiation, tetration and so on) through composition, limited recursion and operation of limited minimum. Limited recursion is a scheme of creating a new function f from established functions g, h and j:

$$f(0,x) = g(x), \quad f(n+1,x) = h(n,x,f(n,x)), \quad f(n,x) < j(n,x);$$

limited minimum is a scheme of creating a new function f from the data of the functions g and h through:

$$f(n) = \text{the smallest number } x \text{ which is smaller than } h(n) \text{ and for which } g(n,x) = 0$$

. Thus, a sub-recursive hierarchy (known as the Grzegorczyk hierarchy) is obtained – a strictly increasing infinite sequence of function classes whose sum is the important, and researched long before, class of primitive recursive functions. The third class of the hierarchy is identical to the class of elementary

functions, which may be defined as the smallest class of functions containing addition and subtraction, closed under composition, limited summation and limited multiplication. This class is also equal, as Ritchie later showed, to the class of predictably computable functions. It is created from the basic class F_0 of numerical functions computable by finite automata, in an infinite number of steps: the class F_{n+1} consists of functions computable by Turing machines which use in their computations for the input w an amount of tape no bigger than $g(w)$, where g is a certain function in the class F_n. Grzegorczyk's classification is thus connected with analysis of computability.

Grzegorczyk is an author of popular lectures on computability, particularly the books A.3 and A.4. Throughout his entire academic activity, he remained faithful to the issues of decidability and computable functions. His logical research mentioned below is usually closely related to this field of study.

2. Computable analysis

In the 1950s, Grzegorczyk wrote several publications examining the possibility of transferring the notion of effectiveness from the field of natural number arithmetic to the field of mathematical analysis. He offered various definitions of computable real numbers as well as methods of development of a mathematical analysis which used only such numbers and computable functions defined by such numbers (B.3, 5, 6, 8). The initiators of this field were, among others, Stefan Banach and Stanisław Mazur. The notes kept by Mazur were translated and published by Grzegorczyk and Helena Rasiowa (*Computable Analysis*, "Rozprawy matematyczne" vol. 33/1963). However, it has turned out that the notion of effectiveness has been of little use in mathematical analysis – so far at least.

In his paper B.21 Grzegorczyk examined computable functionals of higher types, which had been introduced a short time before by Gödel in order to prove non-contradiction of the axiomatic first-order arithmetic.

3. Axiomatic arithmetic

Grzegorczyk is a co-author (with Andrzej Mostowski and Czesław Ryll-Nardzewski) of a fundamental work B.9, which introduces the research on second-order arithmetic: it is a theory formalised in first-order logic, concerning both numbers and sets of numbers. The introduction of the ω-rule (which allows for inference of $(\forall x)F(x)$ from the infinite number of premises $F(0)$, $F(1)$, $F(2)$, ...) makes it possible to show that relations of the class Π^1_1 are representable.

Grzegorczyk also wrote the book A.7, which systematically describes formalised theories of various numbers.

4. Concatenation theory

His "interest in computability stems from the question, what part of the mathematics refers to the most fathomable element of the mathematical reality? Algorithmicity is no more than manipulating what is written in a way expressly given by the instructions." In the latest period of his work, Grzegorczyk embraced research on concatenation theory (i.e. the theory of putting together two texts, or strings of symbols, into one text, where the second text becomes the continuation of the first one), proposed by Tarski in the 1920s. He obtained the following results: the simple theory of this notion, even though it appears weaker than weak arithmetic, is also undecidable (C.34, B.28, B.29) and can be substituted for both metamathematics and elementary mathematics in proofs. Instead of computability, Grzegorczyk preferred to use a "more epistemological" notion of effective *recognisability* of properties of texts or relations between texts. An "empirical" relation of concatenation of two expressions is seen as a basic operation to recognise more complex properties. Entire inferences are carried out without the intermediation of arithmetic. This complies with the approach of theoretical computer science. It is also a continuation of the experience of A.2; for instance, the arithmetic relativisation of quantifiers was replaced with relativisation to subexpressions of an expression.

The interest in decidability and undecidability is present even in his earliest publications (e.g. A.2, A.3). Grzegorczyk proved the undecidability of various theories, such as elementary topological algebra, i.e. Boolean algebra with a closure in the plane (because arithmetic can be interpreted out of it – see B.1) or other weak theories (and B.16). He also delivered examples of theories without recursive models: he was the first to show that combinator calculus (a variation of λ-calculus) is such a theory. He reflected upon various proofs of undecidability starting from recursively enumerable sets that were not recursive themselves (B.7).

B. Systems of logic

For Grzegorczyk, logic is tightly bound with general methodology and formal systems – with "epistemology and scientific ontology." He believed that "propositional calculus is a way of using logical connectives in theoretical contexts. A philosophical meaning may be attributed to predicate logic. It

may be understood as the most general ontology (theory of being, theory of properties and relations)."

1. Axioms of logic

According to Grzegorczyk, axioms of logic should naturally express the basic properties of the logical notions which a given system contains. He distanced himself from "performance" metalogical research, whose aim is, for instance, to find the shortest axiom, however counter-intuitive it would be. For the metatheory of first-order logic, he formulated a theorem about no constants being marked by logic, which is a manifestation of the philosophical thesis about logic being topic-neutral.

2. Axiomatic geometry

In his doctoral dissertation, Grzegorczyk handled a representation of geometry in which, instead of points, there were only solid figures; points can be described indirectly since two solid figures may touch each other minimally, i.e. at only one point (B.12, B.13). The philosophical motivation of the work was to examine the possibility of describing phenomena in compliance with reism, advocated by Tadeusz Kotarbiński. This was also the motivation for several others of Grzegorczyk's works in methodology and semantics – including the earliest ones (B.13, C.2, C.4, C.5, C.8, C.18, C.19, C.30). The language of reism is, in his opinion, the most natural language for a fundamental, empirical description of the world. "The reistic ontological interpretation of the full propositional calculus is a simple continuation of Aristotle's ontology." On the other hand, Grzegorczyk noted that restrictive reism renders the pursuit of mathematics very difficult because, for example, one cannot talk about infinite sets (C.8).

3. Non-classical logics

In B.4, Grzegorczyk showed that Leśniewski's systems of ontology and mereology are formally equal to the Boolean algebra with the zero (corresponding to the empty set in algebra of sets) removed, which in turn is virtually equivalent to the ordinary Boolean algebra. This would mean that Leśniewski's systems are not a meaningful contribution, which, however, does not close the discussion on their philosophical sense.

Grzegorczyk dedicated a number of works to intuitionistic logic and its various interpretations. particularly those using topological notions, as well as the connections with modal logics (B.20, B.23, C.15, C.16). In B.20, he provided a formal interpretation of forced assertion of statements in

scientific research. It turns out that a formula is provable in the intuitionistic logic if and only if everyone, with any information given, must assert it when conducting every such research. Thus, Grzegorczyk has obtained a semantics of intuitionistic propositional calculus. His description was similar to Kripke's semantics, which was created at the same time and instantly became important and influential.

4. Interpretations of logic: the defence of psychologism

In C.6, Grzegorczyk defends the ontological interpretation of the laws of logic: the laws of logic are about the world. He also delves into the history of logic; for instance, (in C.7) he follows the process of emergence of the very important notion of quantifiers, which had been used by mathematicians but became a distinct logical term as late as in the 19$^{\text{th}}$ century. It was the time when a psychologist interpretation of logic was dominant; it was questioned by Frege and Husserl, who said that logical relations were objective, regardless of what people perceived or thought. In the 20$^{\text{th}}$ century, anti-psychologism dominated logical thought. From the very beginning of his academic activity, Grzegorczyk defended psychologism understood as the belief that the relation of signification depends on the human and its description must relate to human behaviours. The description is in language and the language is someone's language, and for someone. We use logic to describe the world. Grzegorczyk tries to precisely describe the way the world can be described. This topic is present in his works starting from C.2, C.4 and C.5 up to C.27 and C.28, as well as in his book A.17, which carries the telling title *Logika – sprawa ludzka* (*Logic – a human affair*) and the paper C.29, whose title is no less telling: *Is antipsychologism still tenable*?

From this approach follows the reinterpretation of semantic antinomies. What they indicate is not so much the contradiction of language as the limitations of the notions we have created. For instance, Grelling's antinomy can be understood as the foundation of a proof of the statement that a certain correctly defined set of expressions cannot be precisely named. Similarly, the liar antinomy allows us to prove that there exists a correctly worded problem on which no methodologically educated person can think in a non-contradictory, honest and fully aware (i.e. with an awareness of the sentences he does and does not assert) manner.

Andrzej Grzegorczyk takes a distinguished place in the history of mathematical logic. He lends his name to the mentioned hierarchy of primitive recursive functions. Another sign of his influence is the fact that, in B.20 and B.23, he handled the modal logic connected to the pattern

$\Box(\Box(A \Rightarrow \Box A) \Rightarrow A) \Rightarrow A$, whose addition to the system S4 creates a system which George Boolos, in his monograph *The unprovability of consistency*, called the system S4Grz – from Grzegorczyk's name. His work with Andrzej Mostowski and Czesław Ryll-Nardzewski remains the starting point of the research on axiomatic second-order arithmetic and arithmetic with an infinite inference rule.

His logic textbook, reprinted multiple times (A.6 in the Polish version, A.8 in English), played a substantial role in the field. His book A.4 on recursive functions was published in French upon request of the publisher and was used in France as a textbook. Grzegorczyk was the first one to popularise logical calculus and the issues of decidability in Poland, in his books A.1 and A.3. The first of these was also published in Czech and Russian.

A sign of the recognition he had as a logician is the fact that after the death of the famous Dutch logician Evert W. Beth, it was Andrzej Grzegorczyk who was asked to become his successor in Amsterdam. Grzegorczyk went there – as it later turned out – only for a few months, as he was not able to settle there. As he explained, he was too attached to Warsaw.

Andrzej Grzegorczyk was a tall man, slim, thin even, with sharp features. His untidy hair and lack of concern with clothing well fitted the stereotypical image of a philosopher or scholar. Notably, he did not care about the impression he made or whether somebody would like him or not; all he was concerned about was the truth. This is why he easily entered relations of distance, often at the discomfort to others, though he never sought conflict. One of the reasons was his "logicality": he preferred things to be said directly, without understatements; he was at odds with allusions or subtle associations. Hence, he always stayed on the margin, even though he was both academically and socially active; he mingled with mathematicians, philosophers, Catholic intellectuals, artists as well as international political and social activists supporting the idea of non-violence.

Andrzej Grzegorczyk always worked in a highly independent manner and, though respected for his knowledge and acute mind, he avoided entering the sort of cooperation that would enable the emergence of a school or at least of his students. It is hard to say whether he had any students or continuators in the narrow sense. This was, to an extent, an effect of his personality traits as well as his high demands with respect to the ability of formal reasoning. As he auto-ironically claimed: "I scared people off." He rarely did research together with others, although he cooperated with various people – in the beginning, mostly with his teacher Andrzej Mostowski

(B.9, B.11 and the review work *The present state of investigations on the foundations of mathematics*, i.e. "Rozprawy matematyczne" vol. 9/1955, written by Mostowski and six of his students) and quite recently, in 2004, with his younger colleagues in a seminar on concatenation theory run together with Andrzej Salwicki and Marian Srebrny at the Institute of Mathematics, University of Warsaw. Being an employee of the Polish Academy of Sciences throughout his career, Grzegorczyk spent little time teaching regular student classes. He was the supervisor of two doctoral dissertations: in logic in 1975 (Stanisław Krajewski, *Niestandardowe klasy spełniania i ich zastosowania do badania niektórych rozszerzeń teorii aksjomatycznych* (*Non-standard satisfaction classes and their applications for the research on some expansions of axiomatic theories*)) and in ethics in 1992 (Bohdan Misiuna, *Analiza filozoficzna zjawiska oburzenia i jej konsekwencje aksjologiczne* (*Philosophical analysis of the phenomenon of indignation and its axiological consequences*)). I must add that I myself, despite having quite a tight bond and always regarding him as a reference point, do not feel his student in any distinct sense.

III. Views: logic and anthropology, ethics and religion

Andrzej Grzegorczyk is logic incarnate, more so even than most of the great logicians. According to him, regardless of the motivation for our reflections, logic is the criterion of their value: whether the reasoning is logical, systematic and self-aware.

Grzegorczyk's fundamental approach to philosophy is largely a continuation of his youthful discovery of logic, suggesting that "everything in this world can be justified in a precise and certain manner." He surely began to take this statement less seriously later, but the firm belief remained that everything can, and should, be formulated in a logical and precise way. Apart from the logical topics, Grzegorczyk took up issues of ethics and philosophical anthropology, always with a distinct methodological self-awareness, within the philosophy he called "rationalism open to values." Logic is supposed to help overcome particularisms. Hence, as he wrote, "a worldview requires logical culture and analytical philosophical insight." He was concerned about this in his work, whether he handled development of the formal construction of the Universal Syntax or examined the human condition – the essence of humanity, the ability to create new notions, theories, classifications, that is, the "megatools of the mind."

I think that Andrzej Grzegorczyk may be regarded as a prototypical logician. I must note here, however, that this is not only meant as praise. Surely, we all have heard people whose near ones were logicians talk – with resignation or even disgust – about their rigidity, insensitivity to the ephemeral, not noticing fuzzy notions, impractical attitude to the world. What is more, many philosophers see the role of logic differently from Grzegorczyk's views. For example, he says: "We might add an evaluation system to the system of world description, but the evaluations must be clearly distinguished, indicated and ordered, so that no one feels cheated. Only formal logic can secure the language from this threat." This is a radical opinion, unacceptable to the greater part of contemporary philosophers. They considered an approach like this to be the source of threats itself because each language must be embedded in Husserl's "Lifeworld"; moreover, a total domination of Pascal's *esprit de géométrie* over *esprit de finesse* poses the threat of missing the reality. All the more so, one might add, that on his reflection upon the need for logic Grzegorczyk added a peculiar thought: he was inclined to believe that a worldwide enforcement of it should be necessary. This thread could be called platonic; it shows why a person invited to symposia and discussions, valued for his erudition and ability to conduct a wide-ranged reasoning in a non-emotional way, hardly anyone identified with him.

1. Views in the form of a philosophical system

Several issues were taken up by Grzegorczyk time and time again, with hope to develop a better, more precise approach. These were mainly issues of ethics and philosophical anthropology, always handled with a distinct methodological self-awareness. The Author himself distinguished the following fields among them: epistemology and ontology, human condition, general human axiology. He found them to be mutually connected. In A.14, he called reflection upon them "rationalism open to values."

As far as epistemology and ontology are concerned (A.13–A.17), Grzegorczyk believed that the structure of the world may be contained in a certain recognised formal structure of notions. The motivation is, to an extent, practical: "a formal system at the foundation of a worldview seems indispensable nowadays due to the linguistic diversity of humankind and the necessity to communicate in an increasingly precise way, the need for unambiguous, objective communication separated from emotions in the increasingly complex matters of coexistence. Logic at the foundation of ontology and metaphysics may ensure freedom from biased or emotionally

marked notions, which from the very beginning contain some selfish pressure and which are easily born in regional or national cultures, where group interest distorts the objectivism of thought."

A.17, besides the issues that were further handled in his other books (particularly A.18 and A.19), contains a formal construction of the Universal Syntax, kept in Tarski's style, which leads to the following statement:

> To say that a sentence A is true is equivalent (within our system) to asserting the sentence A (relativised to the field for which A is being applied).

This may be called a "trivialisation of the notion of truth," though – as the Author emphasised – "the proof for this trivialisation is not trivial" (A.17: 147). While reflecting upon the liar paradox, which lays the foundation for the aforementioned construction, Grzegorczyk transformed the antinomy into a statement about the human, or rather the human condition: as said before, this makes it possible to prove that there is a problem about which a human cannot think in a non-contradictory, honest and conscious way. The anti-psychologist interpretation of meaning is inspired, the Author wrote, by an idealistic vision of the world (C.29: 109).

Human condition is a topic researched and examined by the broadly defined philosophical anthropology. The issues handled in the book *Psychiczna osobliwość człowieka* (*Mental peculiarity of the human*) (A.19) were earlier discussed in A.10, A.12, A.14 and – to some extent – even as early as in A.5 and A.9. The fundamental problem is to detect what constitutes the essence of humanity, what distinguishes the human from other creations. "Such a representation itself is philosophical in its nature although it refers to the knowledge of natural science. However, the natural sciences rarely afford a perspective that is general enough"

The human condition is free existence, even though it is limited by various factors. The human as an "animal" has specific features: it constantly enriches the standards for the quality of life (A.19: 34) and, to a large extent, creates its own habitat (A.19: 37). The individual nature of human, however, is only visible to an approach that goes beyond biology. In principle, the human is distinguished by sensitivity to values and the spiritual sphere, but the more perceptible part is the ability to use language and symbolic thinking, thanks to which an individual can gain control over their emotions. Humans recognise the dimension of sanctity and transcendence. They are also capable of creativity, which gave rise to a civilisation far superior to the "civilisations" of other animals. The most important part, however,

is our ability to create new mental tools such as new notions, theories, classifications etc. The systems of thoughts are the "megatools of the mind" (A.19: 104).

Grzegorczyk approached even the formal issues from a philosophical point of view, "combined with the will to simplify the entire vision and the desire to reveal the humanistic (axiological) overtone of the problem's solution." It is visible, for example, in the analysis of antinomy in A.17. The conclusions of the antinomy concern the intellectual condition of human. What is more, "his entire interest in computability is also humanistic and concerns the human condition. This is a field of thought where there is unwavering certainty. Distinguishing the domain of effectiveness (computability) shows a limitation to our intellectual capacities, that is, a limitation to our cognition. The truly certain and obvious is the very tangibly provable."

With regard to methodological reflection, Grzegorczyk believed that "the way of arriving at the certainty of knowledge is in itself a crucial part of experiencing the value of the gained knowledge. The only way to experience the deeper truths is through linguistic formulas built in compliance with the rules of a language code. Divine knowledge, direct and transcending language, is unavailable and unimaginable to us."

2. Ethics

Apart from reason, the human is characterised by ethics. "A mind purged from egoism and subjected to the discipline of logic (both these things may be very difficult to attain)" should reveal "elements of universal human axiology." Their appearance in the experience of an individual, though, often requires a deeper ethical shock, one's own experience or an encounter with someone's powerful testimony. Grzegorczyk added that "it is quite striking how most people worshipped as saints in the Christian religion are former sinners who went through the stage, or 'cultural device', of metanoia, a great internal conversion, some kind of a fundamental 'turn' in their personal code of conduct. At some point, attempts were made to transfer the device of metanoia to the newly created lay culture of the communist society but, as it seems, without any deeper results."

Formal logic research, however fascinating, remains "child's play" when compared to the real problems humanity is struggling with. Grzegorczyk expressed this belief in dramatic manner: "Sometimes, when looking at the power of the mathematical minds concentrated around abstract problems, I was under the impression that there was some satanic power at work there,

causing the most talented people to be paid for work without meaning for the good of the people. Not one of them works towards the deletion of the real source of human misery. Scientists great and small are employed in an intellectual circus lest they as much as try to think about what is really worth accomplishing in this world." Hence, the "mathematical 'play' – which is quite well-paid, incidentally – may be considered a waste of energy that should instead be used to devise real actions with a distinctly good purpose." Therefore, detached intellectuals should start feeling guilty and a desire for a more dedicated contribution to solving the socially important problems of the country of the world.

Humans "create their own mental tools which let them exceed their earlier standards. It is not enough anymore to feel and suggestively express these humanistic intuitions, as, for instance, the phenomenologists did. A philosopher nowadays must present a clear and consistent system of notions." Grzegorczyk added: "natural scientists present a worldview that is cognitively sloppy, although they gain great authority with respect to general philosophical views because of their scientific authority and referring to certain scientific research. However, they spread imprecise ways of thinking in the process. Of course, most philosophers also contribute to spreading the lack of precision in thinking because they skip the rules of building a logical reasoning for better effect or unconsciously neglect them."

In his social activity as a philosopher, Grzegorczyk was interested in the ethical attitude and method of conflict resolution known as non-violence, whose widely known propagators were Mahatma Gandhi and Martin Luther King (e.g. D.9, D.49). He co-organised visits of well-known activists Jean and Hildegard Goss in Poland and in 1991, he provided substantial help with the organisation of a symposium in Moscow with the participation of the Goss's as well as leaders of the movement such as Jean Vanier from France or Gene Sharp from the USA. Grzegorczyk was a radical: keeping the non-violence ideal, he supported dialogue with everyone, including – as it logically follows – terrorists.

Readiness to coerce others to logic and to dialogue with literally everyone are not the only examples of Grzegorczyk's radicalism of thought. An even more interesting statement of his was that it is harmful to strive to defend one's dignity, "what is one's own, including beliefs and good opinion" (F.2) in every situation, as the right thing to do is to turn the other cheek.

3. Religion

Grzegorczyk combined reistic inspirations with his own sort of naturalism with a convinced religious participation. He was always interested in religion, particularly its moral dimension. He wrote about it from a Christian perspective. He handled religious issues in a literary form in his book of short stories and essays *Moralitety* (*Morality plays*) (A.11). As the title suggests, it is always morality that is the crucial problem to the Author. In the book's "pseudostories", "pseudosermons" and "pseudotreatises", he wrote about Prometheus and about Arjuna, but mainly focuses on the Christian motives that may be seen as a radical commandment of "testimony of selfless care," also towards the opponent. The radicalism he preaches is uncompromised: "And it is not being destroyed that is important, but not allowing for the internal diminishment of one's own intentions."

Grzegorczyk approaches Christianity in a more systematic way in his book *Europa – odkrywanie sensu istnienia* (*Europe – discovering the meaning of existence*) (A.18), in which he also indicates the role of logical thinking as the foundation of the achievements of civilisation. The book is an attempt at an "axiological look at history." According to the author, "the Revelation, that is a certain special divine intervention in the development of human cultures, was adjusted to the evolutionary development of the *homo sapiens*." He advocated the value of such attitudes as altruism and serving. Europe had seen the creation of logically ordered scientific theories. They are founded, according to the author, in the "deductive logic, rules of empiricism and a search of the essence of phenomena" (A.18: 50). Grzegorczyk regarded the meaningfulness of the world as the way of seeing the world like a "text that can be understood." Everything is potentially comprehensible. The history of religion is the pursuit of meaning. Abraham is the beginning of a new era of monotheism. Of course, biblical thinking is metaphorical. A more philosophical version is obtained by a combination with the Greco-Roman European intellectualism. Jesus calls to a consistent individual testimony, while passing over the problems of normal life and handling "almost exclusively extreme situations" (A.19: 152). Christianity calls to "fulfilling the spiritual values, not vital ones" (A.19: 159). Grzegorczyk read Christ's words "let your word 'yes' be 'yes,' and your 'no' be 'no'" as "a kind of approval for the European logic" (A.19: 191). It is quite clear that such an approach may be easily criticised. However, the author strived not to approach the issues in a naive manner, noting that the intermediaries of the Revelation were using notions appropriate to their time and place. I

shall add that one should remember metaphors are used not only in religion, but also in sciences.

In Grzegorczyk's regard, logic in a broad understanding was a pillar of European rationalism, with which he identified (A.18). According to this rationalism, knowledge must be logically and empirically consolidated and reach what is substantial. He said a number of times that only the statements which are intersubjectively communicable and provable should be accepted, a belief he shared with Kazimierz Ajdukiewicz. Grzegorczyk was a religious person, which, of course, led to the problem of agreement between faith and logic. He basically adopted two solutions: the way of ethics and the acceptance of what cannot be expressed by words. First of all, he emphasised ethics and its universality. In the introduction to his analysis of the Decalogue (D.3), Grzegorczyk emphasised that "the commandments of religion contain the same intuitions that are a part of all human instincts."

The other way of overcoming the conflict of reason and religion was to accept – in spite of the radical, narrowly understood rationalism – the entire realm of what is impossible to express. This can be best illustrated by fragments of his poem from 1974 (E.1):

> I do not contradict those who say You are not.
> I agree with my friends who say the notion of You is contradictory.
> [. . .]
> only the one has a pure idea of God who does not have it at all.
> [. . .]
> the one has You to whom even the idea of You is strange
> and who lives a hope
> that [. . .]
> there is something that will remain.

Rational theology is not appreciated under such an approach. Religious worship is a fundamental thing, by which he meant mostly, but not exclusively, Christianity and within Christianity not only Catholicism, in which he was raised, but also Orthodoxy. The biblical tradition was important to Grzegorczyk but the most important thing is the moral testimony of Jesus. Humanity is constituted by both reason and openness to values.

4. Social situation

According to Grzegorczyk, an axiological understanding of human history is "the greatest intellectual challenge" we should feel nowadays. "Particularly

from a religious point of view, axiological experiences are the foundation of human condition enforced on us by the Creator. People are trying to get away from the axiological exam of life in this world and, instead of justice, strive for skill, mainly in the self-centred pursuit of riches. This, however, is an ostrich policy. God shows us repeatedly that it is justice we are being asked for, but the activity of God is limited to the relentless offering of options. Humans remain free and hardly ever choose the best way, except for some moments of heroic decisions that, when made, change the shape of the world – though just for a moment, until the next trial. Nothing important happens automatically, the good cannot be programmed. It has to be constantly created with a new effort."

Grzegorczyk understood long ago what has only recently become obvious to all commentators. Our times are witnessing the growth of "a great civilisational conflict. The wealth of some (privileged classes that exist in all countries of the world) is driving an increasing contrast to the poverty of the rest, who throughout world have become marginalised, excluded from the system, which revolves around the things important mainly for the rich, powerful and clever. Intellectual divisions perpetuate the conflicts. Understanding or agreement on a global scale requires a common language, one common look at the entirety of human affairs. Without a common language of notions, world peace seems impossible." The agreement of notions in question should be carried out in the following way: because of the increasing data flow, decision making requires appropriate justifications, which in turn require a suitable theoretical system. One cannot fulfil the desires of everyone so we have to bear the limitations in solidarity; global regulations have to be introduced and this requires convincing argumentation. Moreover, we need a synthesis of scientific knowledge, which also should "serve the just and peaceful coexistence of people."

Earlier than almost everyone under communist rule in our part of Europe, Grzegorczyk understood that the problems our civilisation is facing are global and that they therefore require worldwide cooperation. He made an appeal to the UN, suggesting the introduction of a rule declaring a new fundamental human right: "every person has the right to help any other person in a worse position than himself or herself in whatever country that person may reside." Solidarity of all humans is a fundamental message of Andrzej Grzegorczyk's writings. Moved by a report by the Club of Rome, he was one of the first people in Poland to advocate limiting consumption and combating waste, an idea that sounded abstract, if not absurd, in Poland forty years ago.

It is much harder to evaluate his work in ethics, anthropology etc. than his accomplishments in mathematical logic and related fields. Grzegorczyk reasoned on his own, with few references to literature (the bibliographies in his papers and books were very limited) and did not contribute to cycles of works created by the specialist circles in universities. For instance, he does not refer to the classic author in philosophical anthropology and theory of values, Max Scheler. Therefore, it might not come as a big surprise that no reviews of his books were published. It may be expected that they will have a limited reach and influence. One of the reasons for that may not be their content but the specific attitude of Andrzej Grzegorczyk. As he utterly ignored all sorts of political and social connections, no circle fully treated him as "one of them". This attitude, focusing exclusively on the reflection on notions – including values – regardless of their sources and context, may have been regarded as obliviousness to the reality, or at least to its parts that were vital for others. For example, Grzegorczyk criticised the "Solidarność" movement at the time when Poland was divided into its supporters and opponents, with little place for a middle option. Even the environment of the Catholic individuals around the weekly magazine "Tygodnik Powszechny" broke cooperation with him at one point. I must add that I myself was frequently shocked by his insensitivity to matters important to me, like the wound left by the Holocaust. However, no one ever accused him of dishonesty.

When evaluating Grzegorczyk's attitude, one might want to note a letter published in the "New York Review of Books" on the 4th August 1977, regarding the mentioned appeal to the UN. In the letter, Grzegorczyk was introduced by Noam Chomsky (the controversial character of his later views is of no importance here) as a "logician and philosopher, a man of great scholarly distinction and courage, whose views have been suppressed in Poland."

Although his non-logical publications generally remained unnoticed, there are exceptions: in 1987, he was awarded a literary prize for his book *Moralitety*. Another exception is his "Decalogue of reason" (D.42), which is still being referred to in discussions about how one should argue in a responsible way. It is worth citing here: 1. You shall not clap. 2. You shall not catcall. 3. Listen to the content, not the tone. 4. Fight an argument not a person. 5. Do not flatter another or yourself. 6. Do not blindly believe another or yourself. 7. Seek the essential. 8. Try to build something better, do not look for scapegoats. 9. Do not generalise too readily. 10. Do not use proverbs, as they are usually the folly of nations.

Grzegorczyk himself followed these rules quite strictly, more than an average philosopher. This meant, among other things "never to care about praise from the audience." He believed that by using only substantive criteria, avoiding patterns we "might deprive our life of the appeal of a boxing ring but will contribute to the better understanding between people." Understanding of people, solidarity of all humans, logic, precision of notions – these are the messages he left us. Simple but always up to date.

Publications by Andrzej Grzegorczyk

A. Books:

1. *Logika popularna*, PWN, Warszawa 1955, 1958^2, 1961^3, 130 p.

   1a. *Popularni logika* (in Czech), Praha 1957.

   1b. ***Популярная логика*** (in Russian), Moskva 1965.

2. *Some Classes of Recursive Functions*, Rozprawy Matematyczne No. IV, Instytut Matematyczny PAN, Warszawa 1953, 45 p.

3. *Zagadnienia rozstrzygalności*, PWN, Warszawa 1957, 142 p.

4. *Fonctions Recursives*, Gauthier-Villars, Paris 1961, 100 p.

5. *Schematy i człowiek*, Biblioteka Więzi VIII, Znak, Krakow 1963, 220 p.

6. *Zarys logiki matematycznej*, Biblioteka Matematyczna vol. 20, PWN, Warszawa 1961, five revised editions: 1961^2, 1973^3, 1975^4, 1981^5, 1984^6.

7. *Zarys arytmetyki teoretycznej*, Biblioteka Matematyczna vol. 37, PWN, Warszawa 1971, 19832, 314 p.

8. *An Outline of Mathematical Logic*, Reidel-Holland 1974, 596 p. (English translation of a revised text of A.6).

9. *Filozofia czasu próby*, Edition du Dialogue, Paris 1979, 223 p.; 2nd ed. PAX 1984, 244 p.

10. *Próba treściowego opisu świata wartości i jej etyczne konsekwencje*, Zakład Psychologii PAN, Ossolineum 1983.

11. *Moralitety*, PAX, Warszawa 1986, 176 p.

12. *Etyka w doświadczeniu wewnętrznym*, PAX, Warszawa 1989, 474 p.

13. *Mała propedeutyka filozofii naukowej*, PAX, Warszawa 1989, 116 p.

14. *Życie jako wyzwanie. Wprowadzenie w filozofię racjonalistyczną*, IFiS PAN, Warszawa 1993, 19952, 288 p.

15. ***Життя як виклик*** (Ukrainian translation of a revised Polish text), (translation and commentaries Борис Домбровский, Олег Гірний), Scholar, Warszawa – Lviv 1997.

16. ***Жизнь как вызов*** (Russian translation of a revised Polish text) (translation and revision Виктор Макаренко), Вузовская Книга, Moskva 2000.

17. *Logic – a Human Affair*, Scholar, Warszawa 1997, 147 p.

18. *Europa – odkrywanie sensu istnienia*, Fundacja Akademii Teologii Katolickiej dla Wspierania Rozwoju Chrześcijańskich Nauk Społecznych i Społecznej Gospodarki Rynkowej, Warszawa 2001, 250 p.

19. *Psychiczna osobliwość człowieka*, Scholar, Warszawa 2003, 174 p.

Papers (more important)

B. Formal logic

1. "Undecidability of some topological theories," *Fundamenta Mathematicae* 38/1951, 137–152.

2. (with Kazimierz Kuratowski) "On Janiszewski's property of topological spaces," *Annales de la Societe Polonaise de Mathematique* 25/1952, 69–82.

3. "Elementary definable analysis," *Fundamenta Mathematicae* 41/1954, 311–338.

4. "The systems of Leśniewski in relation to contemporary logical research," *Studia Logica* 3/1955, 77–97.

5. "Computable functionals," *Fundamenta Mathematicae* 42/1955, 168–202.

6. "On the definition of computable functionals," *Fundamenta Mathematicae* 42/1955, 232–239.

7. "Some proofs of undecidability of arithmetic," *Fundamenta Mathematicae* 43/1956, 166–177.

8. "On the definitions of computable real continuous functions," *Fundamenta Mathematicae* 44/1957, 61–71.

9. (with Andrzej Mostowski and Czesław Ryll-Nardzewski) "The classical and the omega-complete arithmetic," *Journal of Symbolic Logic* 23/1958, 188–206.

10. "Some approaches to constructive analysis," in *Constructivity in Mathematics*, North Holland, Amsterdam 1959, 43–61.

11. (with Andrzej Mostowski and Czesław Ryll-Nardzewski) "Definability of sets in models of axiomatic theorems," *Bulletin de l'Academie Polonaise des Sciences, Serie Math.* 9/1961, 163–167.

12. "Axiomatizability of geometry without points," in *The concept and the role of the model in mathematics and natural and social sciences*, Proceedings of the colloquium in Utrecht, January 1960, Dordrecht 1961, 104–111.

13. "Axiomatizability of geometry without points," *Synthese* XII (2/3)/1960, 228–235.

14. "Le traitement axiomatique de la notion de prolongement temporel," *Studia Logica* 11/1961, 31–35.

15. "An example of two weak essentially undecidable theories F and F*," *Bulletin of Academie Polonaise des Sciences, Serie Math.* 10/1962, 5–9.

16. "A theory without recursive models," *Bulletin of Academie Polonaise des Sciences, Serie Math.* 10/1962, 63–69.

17. "A kind of categoricity," *Colloquium Math.* 9/1962, 183–187.

18. "On the concept of categoricity," *Studia Logica* 13/1962, 39–66.

19. "A note on the theory of propositional types," *Fundamenta Mathematicae* 54/1964, 27–29.

20. "A philosophically plausible formal interpretation of intuitionistic logic," *Indagationes Mathematicae*, XXVI/1964, 596–601.

21. "Recursive objects in all finite types," *Fundamenta Mathematicae* 54/1964, 73–93.

22. "Some relational systems and the associated topological spaces," *Fundamenta Mathematicae* 60/1967, 223–231.

23. "Assertions depending on time and corresponding logical calculi," *Compositio Mathematica* 20/1968, 83–87.

24. "Logical uniformity by decomposition and categoricity in $_0\aleph$," *Bulletin of Academie Polonaise des Sciences, Serie Math.* 1968, 687–692.

25. "Decision procedures for theories categorical in $_0\aleph$," *Proceedings* of a conference on automatic proving.

26. "Unfinitizability proof by restricted ultrapower," *Fundamenta Mathematicae* 73 (1971–72), 37–49.

27. "Axiomatic theory of enumeration," in J. E. Fenstad, P. G. Hinman (eds.) *Generalized Recursion Theory*, North Holland, Amsterdam 1974.

28. "Undecidability without Arithmetization," *Studia Logica* 79 (2005, issue 2), 163–230.

29. (with Konrad Zdanowski) "Undecidability and Concatenation," *Andrzej Mostowski and Foundational Studies*, Andrzej Ehrenfeucht, V. Wiktor Marek, Marian Srebrny (eds.), IOS Press 2008, 72–91.

C. Philosophical and epistemological logic (excluding reviews and other short texts)

1. (with Ludwik Kasiński) "Bibliografia filozofii w Polsce: 1945–46," *Przegląd Filozoficzny* 1–4/1947.

2. "Un essai d'etablir la semantique du langage descriptif," *Proceedings of the X Congress of Philosophy*, Amsterdam 1948, vol. I, 419–421.

3. *Prace o polskiej filozofii współczesnej (bibliografia)*, Warszawa 1948.

4. "Próba ugruntowania semantyki języka opisowego," *Przegląd Filozoficzny* 44/1949, 330–371.

5. "The pragmatic foundations of semantics," *Synthese* 8/1950–51, 300–324.

6. "Uwagi o rozumieniu praw logiki," *Myśl Filozoficzna* 15/1955, 206–221.

7. "Uwagi z historii logiki," *Myśl Filozoficzna* 27/1957, 164–176.

8. "O pewnych formalnych konsekwencjach reizmu," *Fragmenty Filozoficzne*, seria II, PWN, Warszawa 1959, 7–14.

9. "La comprehension des theories abstraites dans les sciences et dans la philosophie." Conference papers, 195–196.

10. "Uzasadnianie aksjomatów teorii matematycznych," *Studia Logica* 13/1962, 197–202.

11. . "Zastosowanie logicznej metody wyodrębniania formalnej dziedziny rozważań w nauce, technice i gospodarce," *Studia Filozoficzne* 1963, 63–75.

12. "Sprawdzalność empiryczna a matematyczna," in *Księga pamiątkowa ku czci Kazimierza Ajdukiewicza*, PWN, Warszawa 1963, 73–76.

13. "Konsekwencje teoriopoznawcze dwóch twierdzeń metamatematyki," *Studia Filozoficzne* 3 (1965), 115–118.

14. "Vers une synthese methodologique de la connaissance," *Logique et analyse* 8/1965, 223–237.

15. "Klasyczne, relatywistyczne i konstruktywistyczne sposoby uznawania twierdzeń matematycznych," *Studia Logica* XXVII (1971), 151–161.

16. "Individualistic formal approach to deontic logic," *Studia Logica* XL, 2 Wrocław 1981, 99–102.

17. "La position des valeurs cognitives dans la structure generale des valeurs," *Epistemologia* VI (1983), 35–44.

18. "Reizm konsekwentny," *Studia Filozoficzne* 5 (222)/1984, 23–29.

19. "On certain formal consequences of reism," *Dialectics and Humanism* 1977, 73–80.

20. "Raison pratique, raison theorique," *Epistemologia* IX/1986, 25–30.

21. "Dyskusja z Kazimierzem Ajdukiewiczem," *Przegląd Filozoficzny* II (2)/1993, 115–116.

22. "Zmienne języka naturalnego i klasyczne definicje pojęć semantycznych," in *Znaczenie i prawda*, J. Pelc (ed.), Biblioteka Myśli Semiotycznej, PWN i Polskie Towarzystwo Semiotyczne t. 26, 451–460.

23. "Zasada racji dostatecznej," *Przegląd Filozoficzny* III (3)/1994, 99–102.

24. "Classical, relativistic and constructivist ways of asserting theorems," in *Philosophical Logic in Poland* (ed. J. Woleński), Kluwer 1994, 19–31.

25. "Uwagi o Szkole Lwowsko-Warszawskiej oraz przemianach i różnicach w jej obrębie," *Przegląd Filozoficzny* IV (1)/1995, 83–89.

26. "Tarski's conception of truth. Application to Natural Language," *Dialogue and Universalism* VI, 1–2/1996, 73–90.

27. "'Kłamca' i błąd antypsychologizmu," in *Między prawdą i normą a błędem*, E. Żarnecka-Biały (ed.), Wyd. UJ, Krakow 1997, 30–40.

28. "The paradox of Grelling and Nelson presented as a veridical observation concerning naming," in *The Lvov-Warsaw School and Contemporary Philosophy*, K. Kijania-Placek (ed.), Kluwer 1998, 183–190.

29. "Is antipsychologism still tenable?" In *Alfred Tarski and the Vienna Circle*, J. Woleński and E. Kohler (eds.), Kluwer 1999, 109–114.

30. "Reizm i analiza logiczna (Spojrzenie w epoce postmodernizmu)," in *Polska Filozofia Analityczna*, Toruń 1999, 31–41.

31. "Czy istnieje ścisłość poza wymaganiami logiki?," in *Język współczesnej humanistyki*, Jerzy Pelc (ed.), Biblioteka Myśli Semiotycznej vol. 46, Polskie Towarzystwo Semiotyczne, Warszawa 2000, 89–96. (Cf. *Studia Semiotyczne* 21 (1998) 47–55.)

32. "Ostrzeżenie przed błędem w podręczniku logiki," *Przegląd Filozoficzny* IX 1(33)/2000, 184–187.

33. "Jeszcze krótki komentarz do „Ostatniego słowa skazańca" prof. L. Gumańskiego," *Przegląd Filozoficzny* IX 1(33)/2000, 192–193.

34. "Tarskiego rachunek tekstow jako właściwy teren dla definicji obliczalności," in *Alfred Tarski: dedukcja i semantyka*, J. Jadacki (ed.), Semper, Warszawa 2004, 28–36.

35. "Czy i jak formalizować filozofię?" *Przegląd Filozoficzny* XI, 1(41)/2002, 202–207.

36. "Rekontra (odpowiedzi na odpowiedzi Jadackiego i Koja)," *Przegląd Filozoficzny* XI, 2(42)/2002, 257–258.

37. "Wiedza dedukcyjna w 20-tym wieku," in *Ratione et Studio*, Białystok 2005, 289–308.

38. "Testability, Feasibility and General Recursiveness," in *Filosofia, scienza e bioetica nel dibattio contemporaneo, Studi internazionali in onore Evandro Agazzi*, Roma 2007, 839–850.

D. Ethics, social aspects and other fields of philosophy (selection)

1. "Jeszcze o podstawach etyki naturalnej," *Studia Filozoficzne* 6(9)/1958, 40–55.

2. "Między dyskursywnym a kontemplacyjnym myśleniem," *Znak* 43/1958, 36–57.

3. "Dekalog – po świecku przeczytany," *Więź* 8 (1958), 20–26; also in A.5, 155–162.

4. "Analiza filozoficzna, kontemplacja, wartościowanie," *Studia Filozoficzne* 5/1959, 161–172.

5. "Przerosty cywilizacyjne a wartości twórcze," in *Charisteria. Rozprawy filozoficzne złożone w darze Władysławowi Tatarkiewiczowi w siedemdziesiątą rocznicę urodzin*, Tadeusz Czeżowski (ed.), PWN, Warszawa 1960, 97–107.

6. "O nie najlepszym rodzaju apologetyki," *Znak* 76/1960, 1343–1344.

7. "Metafizyka bez spekulacji," *Znak* 77/1960, 1417–1421.

8. "Metafizyka rzeczy żywych," *Znak* 80/1961, 154–162.

9. "U podstaw prakseologii bezgwałtu," *Fragmenty Filozoficzne*, seria III, PWN, Warszawa 1967, 317–322.

10. "W poszukiwaniu moralnej podstawy pokoju," in *Filozofia i pokój*, Warszawa 1971.

11. "Dialektyka szacunku i przemocy," *Studia Filozoficzne* 5(78)/1972, 52–55.

12. "Przeżycie transcendencji a mity kultury," *Znak* 236/1974, 224–235.

13. "The moral basis for Peace: the absolute value of the human individual," *Dialectics and Humanism* 1, No. 1/1974, 19–28.

14. "Kiedy życie nabiera sensu?" *Więź* 7–8/1974, 64–70.

15. 15 "Prawo niesienia pomocy," *Więź* 6/1975, 4–7.

16. 16 "Problemy moralności międzyludzkiej dotyczące ludzi w Polsce," Studia Filozoficzne 3(112)/1975, 105–116.

17. "Etyczne problemy demografii," *W drodze* 11/1977, 85–99.

18. "Problemy integracji nauk przyrodniczych i społecznych," in *Model wykształconego Polaka*, Ossolineum 1980, 85–92.

19. "My version of the Christian vision of sense," *Dialectics and Humanism* 1981, 51–53.

20. "Wartości moralne, a postawy grup społecznych," in *Społeczeństwo wychowujące*, Ossolineum 1983, 163–173.

21. "Antropologiczna wizja kondycji ludzkiej," *Roczniki Filozoficzne* XXXI (2)/1983, 59–81.

22. "Pojęcie godności jako element poznawczej regulacji ludzkiego zachowania," *Studia Filozoficzne* 213–214/1983, 57–76.

23. "Moralność, a mity współczesności," in *Nauka – technika – społeczeństwo*, Ossolineum 1983, 169–179.

24. "Moralność intelektualna koleżeńsko kontrolowana," in *Nauka w kulturze ogólnej*, Ossolineum 1985.

25. "Wolność jednostki ludzkiej," *Studia Filozoficzne* 9/1986, 87–101.

26. "Wierność i świadectwo," *Studia Filozoficzne* 5/1987, 33–39.

27. "The philosophy of fate as the basis of education for peace," *Dialectics and Humanism* 1987, 167–172.

28. "New social self-steering," *Dialogue and Humanism* 3–4/1988, 251–254.

29. "Działania pokojowe, a postawy etyczne," *Studia Philosophiae Christianae* ATK 25/1/1989, 141–159.

30. "Dźwiganie utrapień życia," *Ethos* 6–7/1989, 77–80.

31. "Filozofia człowieka a pedagogika," in *Humanizm, prakseologia...*, Ossolineum 1989, 199–208.

32. "Solidarność: ethos, czy przyjęcie losu," *Ethos* 11–12/1990, 114–118.

33. "Struktura psychiczna człowieka," *Kosmos* 39(1)/1990, 169–171.

34. "Moralistyczna wizja dziejów," in *Sens polskiej historii*, Warszawa 1990, 48–60.

35. "The Principle of Transcendence and the foundation of Axiology," in *Logic and Ethics*, P. Geach (ed.), Kluwer 1991, 71–78.

36. "Konstytucja jako wyraz etosu," in *Nowa Konstytucja RP*, Karol B. Janowski (ed.), Toruń 1991, 12–19.

37. "Социльные мотивации и воздействия," in ***Этика ненасилия*** (conference papers), Moskva 1991, 44–53.

38. "Religions and the intellectual challenge of our times," *Dialogue and Humanism* 1992, 59–64.

39. "Інтелектуальна моральність у науковому співтоваристві," in ***Філософська і соціологічна думка*** 9/1992, 78–85.

40. (with A. Góralski) "Uwagi o stylach tekstów filozoficznych," *Przegląd Filozoficzny* I, 2/1992, 107–126.

41. "God's action in the human world, our intellectual humility and the dialogue between religions," *Dialogue and Humanism* No. 3/1993.

42. "Dekalog rozumu," *Wiedza i życie* 3, 1993, 18–20; also in *Nauka i język*, Biblioteka Myśli Semiotycznej, Jerzy Pelc (ed.) vol. 32, Wydział Filozofii i Socjologii UW, Warszawa 1994, 81–85.

43. "Wiara dzisiejszych oświeconych," *Ethos* 27/1994, 45–63.

44. "Racjonalizm kultury europejskiej," *Przegląd Filozoficzny* III (2)/1994, 107–136.

45. "Tolerancja – pokonanie własnej pychy," *Problemy opiekuńczo-wychowawcze* 7/1995.

46. "Z Józefem Marią Bocheńskim po koleżeńsku," *Przegląd Filozoficzny* IV (2)/1995, 17–20.

47. "Czasy małej wiary," *Przegląd Filozoficzny* IV (2)/1995, 103–108.

48. "Samodyscyplina i życie intelektualne," *Protosłownik uniwersalizmu*, Warszawa 1996, 4–8.

49. "Non-violence: wychowanie do negocjacji, demokracji i współistnienia," in *Edukacja wobec wyzwań XXI wieku*, Komitet Prognoz Polska 2000 PAN, Warszawa 1996, 57–92.

50. "Stosunki polsko-ukraińskie a filozofia," *Przegląd Filozoficzny* V 4(20)/1996, 135–146.

51. "Postmodernizm przeciwko prawdzie," *Ethos* 33–34/1996, 150–159.

52. "Argument – nie autorytet," *Ethos* 37/1997, 186–187.

53. "Kroniki lwowskie," *Przegląd Filozoficzny* VI 4(24)/1997, 179–182.

54. "Kapłan i błazen – w nowej rzeczywistości," *Przegląd Filozoficzny* VI 4(24)/1997, 49–66.

55. (with Z. Morokhoyeva and S. Zapaśnik) "Universalistic Social Education," *Dialogue and Universalism* vol. VIII 5–6/1998, 159–163.

56. "Subsydiarność w filozoficznej wizji działań społecznych," *in Subsydiarność*, Monografie i studia. Centrum Europejskie Uniwersytetu Warszawskiego, D. Milczarek (ed.), 2nd edition revised, Warszawa 1998, 42–50.

57. "Encyklika w obronie sensu istnienia," *Przegląd Filozoficzny* VII 4(28)/1998, 22–31.

58. "Logika rewolucji i polski sprzeciw," *Przegląd Filozoficzny* VII 1(25)/1998, 193–203.

59. "The rationalism of European culture," Proceedings of the Conference „The Lvov-Warsaw Philosophical School and the Contemporary Philosophy" K. Kijania-Placek (ed.), Kluwer 1998.

60. "Creativity and freedom," in *Freedom in Contemporary Culture*, Lublin 1999, vol. II, 97–106.

61. "Czasy i wyzwania," in *Pedagogika czasu przemian*, J. Łaszczyk (ed.), ZMWSPS, Warszawa 1999, 9–21.

62. "The vocation of Europe," *Dialogue and Universalism*, vol. IX 5–6/1999, 11–41.

63. "Wizja kondycji ludzkiej we współczesnym świecie," *Przegląd Filozoficzny* VIII 4(32)/1999, 129–135.

64. "Przemoc naszych czasów," in *Wyzwania moralne XXI wieku* (conference papers 10–11 IX 1999). Seria Konferencje i seminaria PWSBiA. Warszawa.

65. "Odpowiedź dyskutantom," *Przegląd Filozoficzny* VIII 4(32)/1999, 151–153.

66. "Ogólna wizja nowoczesnych studiów filozoficznych," *Przegląd Filozoficzny* VIII 4(32)/1999, 201–209.

67. "Antropologiczne podstawy edukacji globalnej," *Forum oświatowe* 1–2 (20–21)/1999, Toruń 1999, 5–13.

68. "Moralne jądro przemian," in *Problemy polityki społecznej* I/1999 IFiS PAN, 115–117.

69. "Dysonans poznawczy. Jak o nim mówić i kto jest jego odkrywcą?" *Przegląd Filozoficzny* VIII 3(31)/1999, 153–166.

70. "Analityczne uzasadnienie etyki," in *The peculiarity of man* vol. 5, Tradycyjne i współczesne systemy wartości, Warszawa-Kielce 2000, 9–22.

71. "Alienacja według Schaffa," *Przegląd Filozoficzny* IX 4(36)/2000, 201–205.

72. "Nie ma kryzysów w pracy twórczej!" *Forum akademickie* No. 7–8/ 2000, 58–59.

73. "The praise of mortality," *Dialogue and Universalism* vol. X 11/2000, 93–97.

74. "Polemika z markizem de Sade," in *The peculiarity of man* vol. 5, Tradycyjne i współczesne systemy wartości, Warszawa-Kielce 2000, 143–6.

75. "Moralistyczna wizja dziejów," Ethos 3(51)/2000, 103–112.

76. "Racjonalizm europejski jako sposób myślenia," in *Wspólnotowość i postawa uniwersalistyczna* No. 2/2000–2001, 5–8.

77. "W poszukiwaniu zbiorowego sensu życia," in *Dylematy etyczne dnia dzisiejszego*, Warszawa 2001.

78. "Wyznaczniki zbiorowej moralności," *Ethos* 55/2001, 213–223.

79. "Etyczne problemy i powinności pracownika nauki," *Nauka* 1/2001, 45–62.

80. "For an intellectual dimension of the dialogue of conciliation," *Dialogue and Universalism* vol. XII, No. 6–7/2002, 87–101.

81. "Europe: Discovering the meaning of existence," *Dialogue and Universalism* vol. XII, No. 6–7/2002, 111–126.

82. "Terroryzm i religia," *Etyka* 35/2002, 177–181.

83. "Насилие как плачевный опыт человечества," in ***Социокультурные трансформации второй половины xx века***, Moskva 2002, 69–78.

84. "Насилие как плачебный опыт человечества," in *Synergia zniewolenia*, Andrzej Góralski (ed.), PTU, Warszawa 2002, 154–166.

85. "Czasy i wyzwania," in *Wspólnotowość i postawa uniwersalistyczna* 3/2002–2003, 5–20.

86. "Idee kierownicze zachowań społecznych i uwarunkowania epoki," in *Kondycja moralna społeczeństwa polskiego*, PAN i WAM, Kraków 2002, 515–531.

87. "Wspólnota," in *Indywidualizm, wspólnotowość, polityka*, Toruń 2002, 13–30.

88. "Sytuacja świata i wyzwania intelektualistów," *Bunt młodych duchem* No. 9–10/2002, 5–6.

89. "Filozofia społeczno-polityczna a badanie kondycji ludzkiej," in *Rozum a porządek społeczny*, Kraków 2003, 11–26.

90. "Vocation of intellectuals and religion in our epoch," *Dialogue and Universalism* vol. XIII, No. 5/2003, 27–32.

91. "Sprawiedliwość a przebaczenie," *Etyka* 36/2004.

92. "Używanie „rozumu" a stan aktualny ludzkości," in *Bogdan Suchodolski. W stulecie urodzin – trwałość inspiracji*, Komitet Prognoz PAN, Warszawa 2004, 257–274.

93. "Niezbywalność *Sacrum*," Księga VII Zjazdu Polskiej Filozofii w Szczecinie, Szczecin 2005, 80–85.

94. "Sercem i rozumem. Refleksje o 'filozofii życia' Karola Wojtyły i jego 'filozofii wypełniania misji Kościoła'," *Kwartalnik Filozoficzny* 33/2005, 4, 225–246.

95. "Empiryczne wyróżnienie duchowości," in *Fenomen duchowości*, Anna Grzegorczyk, Jacek Sojka, Rafał Koschany (eds.), Poznań 2006, 29–34.

96. "Kondycja moralna homo sapiens oraz sens ludzkiego istnienia jako możliwy temat czegoś w rodzaju nowoczesnej teodycei," *Kwartalnik Filozoficzny* 34 (2006), 4, 153–180.

97. [Addition to] *Wojna. Doświadczenie i zapis*, Universitas, Kraków 2006.

98. "Tadeusz Kotarbiński – Humanista i analityczny filozof," in *Myśl Tadeusza Kotarbińskiego i jej współczesna recepcja*, PAN i TNP, Warszawa 2006.

99. "Władysława Tatarkiewicza 'O bezwzględności dobra'," in *Sens, prawda, wartość: filozofia języka i nauki w dziełach: Kazimierza Ajdukiewicza, Witolda Doroszewskiego, Tadeusza i Janiny Kotarbińskich, Romana Suszki, Władysława Tatarkiewicza.* Biblioteka Myśli Semiotycznej, Jerzy Pelc (ed.). Warszawa 2006, 33–38.

100. "O języku ludzkim jako istotnej cesze człowieka," in *Swoistość człowieka? JĘZYK*, Jacek Tomczyk, Grzegorz Bugajak (eds.), Wyd. UKSW, Warszawa 2008, 47–66.

101. "Człowiek zwierzę rozumne." in *Swoistość człowieka? ROZUMNOŚĆ*, Grzegorz Bugajak, Jacek Tomczyk (eds.), Wyd. UKSW, Warszawa 2009, 31–35.

102. "Wizja wartości i dramat wyboru wartości w myśli europejskiej," in *Wspólnotowość i postawa uniwersalistyczna*, Roczniki PTU, No. 8/2012–13, 5–20.

103. "Une dialectique des valeurs (de nos jours): valeurs spirituelles primaires versus liberte," *La philosophie et l'etat du monde*, VRIN 2013, 225–230.

E. Other

1. "Epifania," in *Antologia polskiej modlitwy poetyckiej*, A. Jastrzębski, A. Podsiad (eds.), Wyd. PAX, Warszawa 1974, tom 2, 336.

2. "Osoby, które wywarły wpływ na moje życie," *Bunt młodych duchem*, styczeń-luty 2013, 12–13.

3. Conversation with Andrzej Grzegorczyk 29.03.2006, Archiwum Historii Mówionej, `http://ahm.1944.pl/Andrzej_Grzegorczyk/1` – `/6`

4. Conversation with Professor Grzegorczyk 31.10.2012, series of meetings „Nestorzy polskiej filozofii," `https://www.youtube.com/watch?v=W3cUe8Hv2w8`

Selected publications about Andrzej Grzegorczyk

Krajewski, Stanisław, "Andrzej Grzegorczyk," *Edukacja Filozoficzna* 37 (2004), 185–204; the same in *Polska filozofia powojenna III*, W. Mackiewicz (ed.), Agencja Wydawnicza Witmark, Warszawa 2005, 99–118.

Krajewski, Stanisław and Jan Woleński, "Andrzej Grzegorczyk: Logic and Philosophy," *Fundamenta Informaticae* 81, 1–3 (2007), 1–10; the same in *Topics in Logic, Philosophy and Foundations of Mathematics and Computer Science. In Recognition to Professor Andrzej Grzegorczyk*, S. Krajewski, W. Marek, G. Mirkowska, A. Salwicki, J. Woleński (eds.), IOS Press, Amsterdam etc. 2007.

K r a j e w s k i , S t a n i s ł a w , "Andrzej Grzegorczyk – logika i religia, samotność i solidarność," Roczniki Polskiego Towarzystwa Matematycznego, Seria II: *Wiadomości Matematyczne* 44, 2008, 53–59.

K r a j e w s k i , S t a n i s ł a w , "Andrzej Grzegorczyk," in Encyklopedia Filozofii Polskiej KUL, 456–458.

Introduction to *Studies in Logic, Grammar and Rhetoric* 27 (40), 2012, Papers on Logic and Rationality: Festschrift in Honour of Andrzej Grzegorczyk, Kazimierz Trzęsicki, Stanisław Krajewski, Jan Woleński (eds.).

K r a j e w s k i , S t a n i s ł a w , "Andrzej Grzegorczyk," NEWSLETTER OF IMPAN, Spring 2014, 9–10.

T u r n a u , J a n , "Andrzej Grzegorczyk – Pożegnanie," *Gazeta Wyborcza – Stołeczna*, 4.04.2014, 13.

T r e l a , G r z e g o r z , "Logika – sprawa ludzka Wspomnienie o profesorze Andrzeju Grzegorczyku (1922–2014)," *Argument 4* (2/2014), 491–498.

J a n k o w s k a , M a ł g o r z a t a , "Filozoficzne dekalogi – tekst dedykowany pamięci profesora Andrzeja Grzegorczyka (1922–2014)," *Zeszyty naukowe Centrum Badań im. Edyty Stein*, Poznań, 12/2014, 251–265.

Andrzej Grzegorczyk. Człowiek i dzieło, Andrzej Góralski (ed.), Warszawa: Akademia Pedagogiki Specjalnej 2015, selected papers by Grzegorczyk and his biography with a *laudatio* by Jan Woleński on the occasion of awarding Grzegorczyk with doctorate *honoris causa* and reviews by Witold Marciszewski and Roman Murawski, as well as papers by Urszula Wybraniec-Skardowska, Józef Wilewski, Barbara Pilipczuk, Daniel Kantor, Joanna Łukasiewicz-Wieleba and Alicja Baum, Jan Łaszczyk, Krystyna Najder-Stefaniak, Małgorzata Jabłonowska i Justyna Wiśniewska, Oleg Hirny.

Originally published as "Andrzej Grzegorczyk (1922-2014)" *Studia Semiotyczne* 28–29 (2015), 63–88. Translated by Agnieszka Przybyła-Wilkin.

STUDIA SEMIOTYCZNE—ENGLISH SUPPLEMENT, vol. XXIX (2017), 39–47
ISSN 2451-2745
DOI: 10.26333/stsen.xxix.02

Jerzy Pelc[1]

Barbara Stanosz (born in Warsaw, 8th January 1935 died in Warsaw, 7th June 2014)

If we developed a habit of placing keywords in biographies, her person and life would be best suited by "rebellious independence", "daring", "courage", "rationalised radicalism" and "succinctness". Klemens Szaniawski would jokingly remark that Barbara wrote only brief summaries of her own works. She treated it as a compliment, and rightly so. Her satisfaction while reminiscing these words was apparent. The conciseness of her utterances sometimes had a nearly hypnotic effect.

She was an exceptionally stubborn soul, striving towards her goals against all odds. Very Polish in her nature, she loved and practiced freedom as the core value associated with democracy rather than the Polish nobility. She regarded the no smoking policy as an assault on her own freedom and citizen rights. Therefore, she cured her chronic bronchitis with packages upon packages of cigarettes smoked daily, nourishing her future fatal disease day by day; from her lungs, it spread to her brain, the organ that had been working so well, and soundly, within her.

She entered the philosophical faculty in Warsaw during the worst period, probably before Commissary Wyszyński proclaimed mathematics, and logic as part of it, not to be the "leading force" in the fight for spreading Marxism-Stalinism all over the world. Only after this partial acquittal could these disciplines, freed from the tasks of battle forces, become a refuge – in exceptional cases, also for the politically independent researchers. Still, the "production meetings" of the Polish Youth Association, socialist work discipline and competition of workers "fighting" to exceed the 100% of the norm persisted all around; posters warned of the "spitting dwarves of the

Home Army"; deeds to celebrate the anniversary of the October Revolution were proclaimed; on the 1st May and 22nd July people were sent to compulsory marches and demonstrations; before elections and the "Three Times Yes" referendum, all-night vigils were organised in university halls to make sure that the "class enemy", made up to boost the socialist alertness, did not attack undercover; state collective farms were aided in their harvesting actions; students underwent their exams in groups of three: let there be witnesses that a failing grade was not a "reactionary" assault of the class enemy, waged from "you know what positions" on "the youth activists", who tried to defend their apparent ignorance with the unquestionable argument: "we were revising as a collective!".

In such a situation, I do not know how the student Barbara Zatryb fared. On one hand, she had excelled at her school leaving examination. On the other hand, she had an individuality well characterised by her novel-worthy surname ("Zatryb" is the imperative form of the colloquial Polish word for "understand"), additionally tainted by her suspicious background: a daughter of a professional soldier, lieutenant or captain of the bourgeoisie-originated Home Army, who did not return after September 1939, leaving behind the four-year-old Basia (Barbara) and her mother. After Barbara graduated with flying colours, gaining her master's degree in philosophy, she managed – despite some opposition from the Party – to get a job at the Institute of Philosophy, University of Warsaw, which was an important part of the "ideological front". The decision to employ her was soon justified by the rapid development of the young academic, her publications and the position she had gained with her superiors, colleagues and students. Thanks to these achievements in the very beginning of her career, she deserved the introduction of "a rising star of our philosophy of language" during one international symposium. However, her professional success was not accompanied by a financial one: the scant salary of an academic assistant could not satisfy the material needs of the family of (at first) two. She and her mother struggled with poverty, as they had done since Barbara's early childhood. A few years later, Barbara changed her surname to her husband's, Stanosz, equally novel-worthy and emphasizing the tendency to be independent ("stanąć" being Polish for "to stand"). On their own, Barbara and her mother, for as long as she lived, soon started raising their respective son and grandson as well as another beloved and important family member, their dog. After her mother's death, Barbara, as the only provider for the three, had to carry the entire burden on her own shoulders. Her material situation forced her to accept one extra job after another. These

were not "odd jobs", though, but always new and valuable works which gave our literature splendid translations of the world philosophy, complete with masterpiece introductions, analyses and commentaries. But this was not all. Barbara wanted to dress fashionably but did not have enough money. Hence, she sewed herself a fur or sheepskin coat – which of the two, I do not remember. It turned out that she could also live up to the role of a furrier.

Her strong personality, ambition and pride, as well as a decisiveness which was visible even in her handwriting, were leading her towards the role of a soloist rather than a team worker. She agreed to join the Collegium Invisibile (a Polish association providing tutelage to academically talented students) but only as an ordinary tutor, leaving the honourable places to less deserving ones. It required a lot of effort to convince her to allow for her nomination to the Warsaw Scientific Society, which she left after some time, in 2008, on her own request; for what reason, I know not. She would not be convinced, however, to join the Polish Academy of Arts and Sciences or the Institut International de Philosophie, which she undoubtedly deserved in both cases and to which she could have been certain of being chosen.

As far as confrontation was concerned, she surely preferred duels to round table panel discussions. She also valued attack more than defence. When certain of being in the right, she was unwilling to show generous understanding to her opponent. "Let him humble himself" – she once said when asked to "spare shame" to a defeated adversary. At the same time, she was always ready to defend another's endangered or violated rights and stormed against those who limited the freedom of speech, conscience or thought. She would not look at her own safety or the price she was about to pay. And pay she did, dearly and – considering her situation as a lonely mother – very painfully. For one act of opposition against a violation of civil rights, she lost her job at the University of Warsaw and for months she had to commute to Częstochowa to earn a living by holding commissioned classes in a higher education establishment there. For some reason unknown to me, she retired early, a great loss for students and yet another material loss for her.

She was a talented and passionate teacher, working with real mastery. We taught and examined the same groups of students for years. One time, when she thought I graded a student too severely, she instructed me: 'one need not have very good knowledge to be graded "very good"'. I remembered this remark well. Thus, in 1971, when I was giving Barbara a lift in my first East German-made two-door Trabant car to her home at Polna street, near Plac Unii Lubelskiej (having made a huge effort, as every beginner driver,

to plan the route so that I had to turn left as little as possible), and heard from the courageous passenger that I was a "very good driver", I understood the praise adequately as it reminded me of that earlier teaching hint. Her contact with the young generation was facilitated, and the bonds tightened, by the fact that she clung to her youth. The students felt this and thanks to this attitude, saw an ally in her. The way she taught and the lot of work she put into it – also by publishing her lecture and tutorial materials – and even the high demands she put on them were rewarded by respect, appreciation, attachment, gratitude and sympathy.

Besides her teaching, her temperament brought her towards journalism as a form of civic activity. She published in a periodical co-initiated and co-founded by her. The title of the bimonthly adequately illustrated its main thought and program: anti-irrationalism and criticism, driven by reason and governed by logic – "Bez Dogmatu" ("Without Dogma"). Without any dogma whatsoever, be it religious or party-political, calling to internationalism (but not cosmopolitanism! – as the Polish United Workers' Party emphasised) that strove towards the worldwide rule of the peoples' masses. Without the superstitions born of emotions and imagination: the fear of annihilation and the rebellion against the inevitability of one's own non-existence; and of desires and dreams to protect the humanity from what Tadeusz Kotarbiński called "elementary catastrophes". She made the case for the freedom of conscience. She decidedly opposed any violation or limitation of minorities' rights. She was making way for progressive thought.

Not without reason was Tadeusz Kotarbiński mentioned here. She still had a chance to listen to his lectures. If some of their ethical ideas transpired to her mind, these were rather not the "quaker" elements or the rejection of "tone, expression and sarcasm" postulated and recommended by him but foreign to the personality of the combat-ready oppositionist that Barbara was. More likely, she learned what Kotarbiński himself had characterised with the sentence "I have always chewed on the bit and chew on the bit I will". Her world view and the judgements she uttered loudly brought attacks from both sides. So many compatriots wanted to tell her directly what they thought about her in a sincere "Pole-to-Pole" talk that she was forced to change her telephone number. Before she clandestinely gave me the new one, a pause occurred in our phone calls, infrequent anyway, as besides the classes, which neither of us lacked, she would rest between 3 in the morning and 1-2 in the afternoon and work in the afternoons, evenings and partly nights.

Whose fledgling was she? Barbara's PhD supervisor was Roman Suszko; the thesis was titled *Funkcje znaczeniowe wyrażeń w ujęciu logiki formalnej* [*Functions of expressions: a formal logic approach*]. The choice of the dissertation topic heralded her future academic focus: logic of natural language. In the paper *Roman Suszko na Uniwersytecie Warszawskim* [*Roman Suszko at the University of Warsaw*] in *Sens, prawda, wartość – filozofia języka i nauki...* [*Sense, truth, value – philosophy of language and science...*] (Warsaw 2006, Biblioteka Myśli Semiotycznej, vol. 50), she recollects an occurrence from the time of her writing the thesis. I remember her being impatient and nervous when her doctoral defence date was delayed; this could have ended in it passing her thirtieth birthday on 8th January 1965 after she had made it a matter of honour to gain the academic title before that date – she managed. When I once asked her whom she regards as her academic master, she said "Roman Suszko". She may have also gained a lot from the academic contact with Janina Kotarbińska, which gave her a chance to learn inquisitiveness, diligence, critical thoroughness, unbiased forming of opinions and determination in keeping her ground on the essentials. It was from her own personality, though, that she took the unwavering courage. In Barbara's writings there are no expressions of uncertainty such as "it seems" or "presumably". Her concentrated works decree on the things as they are in a truly manly manner. She succinctly brought up the quintessence of the described thoughts of others, showing what was their most important part, naming the major advantages, disadvantages and controversies and presenting her own stance. She did not have to compromise any of her concision for the sake of didactic persuasion, repetitions according to the rule "*repetitio est mater studiorum*" or diversification of form – her style of writing was far from all those anyway. She wrote only as much as was needed to present the issue adequately, without a single redundant word. I sometimes thought students should receive a bottle of thinner with each of Barbara's books to be able to swallow such a concentrated concoction.

Barbara's academic face was undoubtedly strongly influenced by Kazimierz Ajdukiewicz. For some time, she was charmed by his mind and overpowered by his brilliant intellect. However, it seems that she was not convinced by Ajdukiewicz's radical conventionalism, just as the semantic and ontological thought of Kotarbiński's reism had not broken through the robust wall of her individuality.

I am under the impression that the stages of her academic development were marked by changes of theoretical predilections: Carnap, Ajdukiewicz, Chomsky, Quine, Davidson. She drew the borders of her privacy and in-

dependence wide; no one unauthorised was allowed inside. Once, at the beginning of her academic pilgrimage, I jokingly remarked that we might have been witnessing a change in her taste, swaying from Carnap towards Ajdukiewicz; she reacted in a surprisingly abrupt way, saying she did not wish her personal academic sympathies to be interpreted. I had apparently dared to trespass upon the carefully guarded territory of her intimacy.

The axis of her researcher's mind stretched between language and cognition, with issues of philosophy of language and logic of language wrapped around it. Language was a window offering a view on cognitive processes. She regarded the logicised philosophy of language as a gateway to the entire philosophy, which allowed her to perceive its wide landscapes. She did her research in the spirit of analytical philosophy, rather American than Oxford style. From the university offer, she always chose lectures and tutorials for psychologists, which were an extension of her theoretical interests and research. She believed that every student of arts and social sciences needed a "good course in formal logic". We did not agree on that point: why disappoint the youngsters who sought to flee mathematics under the wings of the liberal arts? In my opinion, a general logic course after the fashion of *Elements* by Tadeusz Kotarbiński and *Pragmatic logic* by Kazimierz Ajdukiewicz, tailored to the needs of the given field of study, would be more useful.

She never used the name "semiotics", abiding by the terms "philosophy of language" and "logic of language". She was also distrustful towards the field of semiotic research at first. At the beginning, back in the 1960s, she sarcastically asked me, "when is this happening of yours taking place?" as she was going to attend it. Even though the name did not convince her, from the very beginning she actively participated, as a speaker, in worldwide and countrywide semiotic events in Europe and the USA and she entrusted several of her works to semiotics-oriented publishers, Studia Semiotyczne and Biblioteka Myśli Semiotycznej.

The volume *Logiczne podstawy języka* [*Logical Foundations of Language*], Wrocław 1976, Ossolineum, co-authored by Adam Nowaczyk, is one of the books published as a part of the semiotic research program Znak – Język – Rzeczywistość [Sign – Language – Reality], which has been going since the 1960s until the present day. Other parts of the program are: the journal Studia Semiotyczne, published since 1970, in whose volume XXVIII–XXIX I am writing these words, as well as the series of books called Biblioteka Myśli Semiotycznej [Library of Semiotic Thought], started in 1990, whose 54th volume – and hopefully not the last one – was published in 2013. In the *Logical Foundations of Language*, the two authors consistently use the plural,

taking common responsibility for all the thoughts. It is surely possible to guess which parts of the book can be attributed to Barbara Stanosz and which ones to Adam Nowaczyk; that said, a future historian of science should rather ask this vital piece of information of the other author, remembering that they certainly consulted the entire content with each other, than make assumptions. Other works by Barbara Stanosz in the series Biblioteka Myśli Semiotycznej are *10 wykładów z filozofii języka* [*10 lectures on the philosophy of language*] (BMS vol. 19, Warsaw 1991) and *Logika języka naturalnego* [*Logic of natural language*] (BMS vol. 43, Warsaw 1999); besides these, BMS vol. 46, Warsaw 2000, with the title *Język współczesnej humanistyki* [*The language of the contemporary humanities*], included her article *Bełkot i przesąd* [*Gibberish and superstition*] and the already mentioned vol. 50 *Sens, prawda, wartość: filozofia języka i nauki...* – another article with the title *Kazimierza Ajdukiewicza pojęcie racjonalności* [*Kazimierz Ajdukiewicz's notion of rationality*]. Studia Semiotyczne published her following papers: *O pojęciu języka prelogicznego* [*On the notion of a prelogical language*] (in vol. I, 1970), *Kodeks języka naturalnego* [*Code of conduct for natural language*] (in vol. II, 1971), *Status pozanwczy semantyki* [*The cognitive status of semantics*] (in vol. V, 1974), *O ustalaniu znaczeń nieznanego języka* [*On Establishing the Meanings of Expressions of an Unknown Language*] (in vol. VI, 1975), *Teorie, modele i dane empiryczne w lingwistyce* [*Theories, Models and Empirical Data in Linguistics*] (in vol. X, 1980), *Uwagi do artykułu Renaty Grzegorczykowej "Opis lingwistyczny a opis języka"* [*Comments on Mrs. Renata Grzegorczyk's Paper "Opis lingwistyczny a opis logiczny języka"* ("*Linguistic description versus the logical description of language*")] (in vol. XIX–XX, 1994), *Rozwiązywanie paradoksów* [*Paradox resolution*] (in vol. XXV, 2004). As a part of the semiotic research program Znak – Język – Rzeczywistość, she took active part in the following congresses, symposia and academic meetings as a keynote speaker, presenter or discussion participant, with lectures: *Semantyka Rudolfa Carnapa* [*Semantics of Rudolf Carnap*] (Department meeting, 1st April 1966), *O pojęciu języka prelogicznego* [*On the notion of a prelogical language*] (Department meeting, 11th May 1969), *Język a komunikacja* [*Language and communication*] (academic meeting of the Department of Logical Semiotics, University of Warsaw and the Polish Semiotic Society, 23rd November 1973), *Methodological status of semantics* (in the Polish Semiotic Seminar of the North American Semiotics Society Colloquium and 1975 Linguistic Institute Tampa, Florida, July 1975), *Theories, models and empirical data in linguistics* (International Semiotic Symposium, Radziejowice, 22nd–27th May 1978), *On a mysterious principle of mod-*

ern linguistics (International Symposium on Theoretical Semiotics: Verbal Signs – Visual Signs, Warsaw, 23rd–24th September 1980), *Deduction and the behavioristic concept of assertion* (Poland-wide logical conference Uzasadnianie w matematyce i filozofii [Justification in Mathematics and Philosophy] co-organised by the Polish Semiotic Society and the Polish-Bulgarian symposium Types of Logical Systems and Problems of Truth, Jabłonna, 27th–31st October 1983), *Przekład i znaczenie* [*Translation and meaning*] (academic meeting of the Department of Logical Semiotics, University of Warsaw and the Polish Semiotic Society, 27th March 1983), *Logicy i językoznawcy o języku* [*Logicians and linguists on language*] – discussion by Renata Grzegorczyk, Barbara Stanosz, Wacław Mejbaum and Jan Woleński (academic meeting of the research program Znak – Język – Rzeczywistość and the Polish Semiotic Society, 11th March 1995), *Język współczesnej humanistyki – o niejasności naukowych tekstów humanistycznych* [*Language of the contemporary humanities – on the lack of clarity of the academic writings in humanities*] – discussion by Barbara Stanosz, Henryk Hiż, Jacek Jadacki, Leon Koj, Jerzy Pelc and Bogusław Wolniewicz (academic session of the research program Zank – Język – Rzeczywistość and the Polish Semiotic Society, 12th April 1996), *Roman Suszko na Uniwersytecie Warszawskim* [*Roman Suszko at the University of Warsaw*] (academic session co-organised by the Warsaw Scientific Society, the research program Znak – Język – Rzeczywistość and the Polish Semiotic Society, 18th January 2002).

Some of these were later published in Studia Semiotyczne or in Biblioteka Myśli Semiotycznej. The above list does not include Barbara Stanosz's writings, lectures and presentations that were not part of the research program Znak – Język – Rzeczywistość. A proper place for those in the future would be a full bibliography of her works, prepared by her audience and readers as a sign of gratefulness for what they could learn from her. The academic achievements of Barbara Stanosz place her among the greatest Polish philosophers of language of the last fifty years.

* *

*

"It is dreadful how from a subject, you become an object", she whispered as the coffin of Janina Kotarbińska was being placed in the grave.

The penultimate e-mail from Barbara came on the 19th April 2014. It was an answer to my question about how she was faring. I was about to ask her for an article for the Studia Semiotyczne. She wrote – as always – concisely and matter-of-factly. She started with a remark: "I presume our both situations might isomorphically map each other"; she informed me about her health and about having declined an operation. "Instead," she wrote, "I decided to write a book (a popular-philosophical one, on the dualism of the human vision of the world – cognitive and world-view). I was doing quite well-" she went on explaining why her work was interrupted. The letter ended with the words: "I am sending you a warm embrace but let us not say goodbye yet!" I offered her technical help, like having the further part of the mentioned book dictated and recorded on a tape, and asked how we should stay in touch, by the telephone or through letters. On the 23rd April she wrote back: "For now I prefer the e-mail because things tend to fall from my right hand and I have not become accustomed to being one-handed yet. B."

These were the last words from Barbara.

Originally published as "Barbara Stanosz (ur. w Warszawie 8 stycznia 1935 r., zm. w Warszawie 7 czerwca 2014 r.)" *Studia Semiotyczne* 28–29 (2015), 33–40. Translated by Agnieszka Przybyła-Wilkin.

STUDIA SEMIOTYCZNE—ENGLISH SUPPLEMENT, vol. XXIX (2017), 48–61
ISSN 2451-2745
DOI: 10.26333/stsen.xxix.03

Cezary Cieśliński[1]

Paradoxes of Barbara Stanosz

Professor Barbara Stanosz was a years-long lecturer at the Institute of Philosophy, University of Warsaw. In her work she mainly – but not exclusively – focused on the theory of language, particularly semantics and the issues of logical description of phrases in language (the problem of logical form). She was an author of renowned textbooks, including the famous *Ćwiczenia z logiki* [*Exercises in logic*], a vastly popular exercise book helping students to acquire the material on propositional logic, predicate logic and set theory. It is worth mentioning that besides her academic and teaching activities, she was also a social activist and a great supporter of the state's worldview neutrality. She died on 7$^{\text{th}}$ June 2014.

In the late 1980s, as a student at the Institute of Philosophy, University of Warsaw, I had the opportunity to attend the Professor's seminars. I first went there drawn by the seminar's topic... and whoever came there, usually stayed. The great combination of the Professor's rigor of thought with her casual style and spot-on ripostes impressed us greatly; one could feel (which is not a common thing at seminars) that she genuinely *cared* about the topics we were discussing. In some mysterious way, she was able to solve the classical problem probably known to all lecturers: how to make the course participants notice *their own questions* within difficult, often technical issues they were tackling – fascinating problems which they would later like to handle themselves.

This is exactly what happened to me. As a result, I not only became a seminar participant for years but also wrote my MA thesis supervised by the Professor. If I were asked about the source of my interest in the theories of truth, which are the main subject of my study nowadays, I would

[1]University of Warsaw, Faculty of Philosophy and Sociology.
E-mail: c.cieslinski@uw.edu.pl

indicate the conversations with Professor Barbara as the key factor. Here is the main thread of our seminar discussions: according to Barbara, the central issue of semantics was "to explain the phenomenon of understanding any sentence of a given language based on a limited number of sentences that were understood before" (Stanosz 1999: 103). In other words: when we learn a language, we (necessarily) encounter a limited number of sentences that are actually uttered by other people. How do we acquire the ability to understand new utterances (ones never heard before) on this basis? That is the question. When answering it, Barbara always emphasized the statement that understanding a sentence is nothing else but knowing its truth conditions. This is where the notion of truth comes into the foreground. Barbara tried to convince us that the definition of truth by Tarski allows us to describe the recursive procedure of establishing truth conditions, which is why it can serve to create a model description of language acquisition. The main thought here is that by learning a language, we master a procedure, or an algorithm, for establishing the truth conditions of sentences. The description of this algorithm can be drawn from the works of logicians working on truth theory (particularly Alfred Tarski).

One of the last works by Barbara Stanosz is the paper *Rozwiązywanie paradoksów* [*Paradox resolution*] published in "Semiotic Studies" in 2004. My impressions from reading this? Well, I must admit that the clarity of this work and its care for detail is something natural and obvious to me. The Professor had spoiled her students: she had got us too accustomed to some things! An understandable piece of writing with attention to detail? What else can a reader of a philosophical paper expect? It is an obvious thing, isn't it? Isn't it indeed?

A much greater surprise to me was the scepticism of the last paragraphs of the paper. They radiate a deeply rooted doubt in the perspectives of truth theory for natural language. This doubt of Barbara Stanosz is – at least for me – something new: I remember her from the seminar times as a supporter and propagator of the formal study of natural language, not caring too much about such obstacles as semantic paradoxes. Where had the change come from?

Never mind, I would soon know everything as shortly I was going to have a seminar talk on paradox resolution! I had been preparing all week long and I wanted to discuss, intended to convince, all the participants that I was right! I ran in, panting, but the only person I could see was a peculiarly aged course colleague who says, "You are too late, colleague. The Professor is no longer with us".

* *

*

The starting point of Barbara's paper is the definition of a paradox as "apparently valid inferences that lead from acceptable premises to unacceptable conclusions" (Stanosz, 2004/2015, p. 5). Let us add that a conclusion may be unacceptable for different reasons. For example, it can be obviously non-compliant with our experience – here we can categorize the famous ancient paradoxes by Zeno of Elea, arguing for the impossibility of movement. However, logicians are particularly interested in the special kind of paradoxes, for which we will here reserve the notion of "antinomy": an antinomy is a paradox that results in a contradiction.

The key condition here is that paradoxical reasoning uses premises and inferences that are *accepted* by us – and often even obvious. Not every reasoning that concludes in a contradiction is an antinomy! The exceptional role of paradoxes stems from the fact that they reveal loopholes and weaknesses in the system of our basic conceptions. According to Barbara Stanosz:

> we tend to regard paradoxes as painful blows to human reason [w]e feel [. . .] must be parried or eliminated by means of ironclad solutions. (Stanosz, 2004/2015, p. 5)

Indeed. Still, I have a strange impression that the only exception to this rule is a logician's mind. A logician is not a typical kind of person: they love paradoxes and can never get enough of talking about them.

The paper by Barbara describes the strategy of handling paradoxes. The author distinguishes four methods of solving paradoxes:

(A) to justify the thesis that the conclusion merely appears to be unacceptable when in fact it is quite natural and harmless;

(B) to show that at least one of the steps in the inference is logically invalid;

(C) to prove that at least one of the premises is false

(D) to show that a premise is nonsensical (Stanosz, 2004/2015, p. 5). I consider this description of possible strategies very fitting. All the mentioned methods are then illustrated by various examples (Solution (A) – Eubulides' paradox, (B) – Zeno's paradox of movement, (C) – Russell's paradox, (D) – the liar paradox).

In this text, I will try to illustrate all four methods by using a single example: the liar paradox. This classic antinomy turned out to be a hard nut to crack and the very existence of many unequal solutions inspires thought: wouldn't one really good solution be enough?

Let us quickly remind ourselves of the paradox here. Let (L) mark the sentence:

(L) is false.

We then ask whether (L) is true or false. We consider all the cases. If (L) is true, then it is as (L) says, so (L) is false – contradiction. On the other hand, if (L) is false, then it is not as (L) says, which means that (L) is not false – another contradiction. Thus, we get a contradiction regardless of the case considered. This is in an antinomy.

I would like to emphasize that the version of the liar paradox presented above is intuitive and non-formal. Hence, it has one feature specific for intuitive reasoning: it is not entirely clear what premises and rules are being used in it. In such situations, a logician's first task is to write down the reasoning without any loopholes or shortcuts. That said there is no guarantee that a given intuitive reasoning will correspond to one and only one full, formalized version.

I will now present the liar reasoning in a more precise shape (not forgetting the fact that this is still just one of many possible versions – this will later be important!). In the formalisations below, I assume that "obtaining a contradiction" means proving a sentence in the form "$\varphi \wedge \neg\varphi$" (where φ may be selected freely).

Liar – version 1

Let Tr be our predicate of truth. We recreate the reasoning by means of a theory T, which we assume to fulfil the following conditions:

1. T contains all substitutions of the expression $Tr(\varphi) \equiv \varphi$
2. There exists a sentence (L) such that $T \vdash \mathrm{L} \equiv \neg Tr(\mathrm{L})$
3. For any formula A, if $T \vdash (\varphi \equiv \psi)$ and $T \vdash A(\varphi)$, then $T \vdash A(\psi)$
4. For any sentence φ, if $T \vdash \varphi \equiv \neg\varphi$, then $T \vdash \varphi \wedge \neg\varphi$

Here is a short commentary. The first condition is an equivalent of the following intuitive claim: a sentence (any sentence, let us add, with or without a truth predicate) is true when it is as the sentence says. The second condition introduces a liar sentence L, understood as follows: (L) is identical (provable on the grounds of T) to its own falseness. It is worth mentioning that this condition will be fulfilled by every theory T containing a big enough fragment of first-order arithmetic, so it turns out not only possible to fulfil but even fairly natural[2]. Conditions 3 and 4 in turn characterize the fragment of the logical apparatus of our theory.

Now we can prove that the theory T defined that way is contradictory.

Observation 1. There exists a sentence φ such that $T \vdash \varphi \wedge \neg\varphi$.

Proof. Based on 2, let us take a sentence (L) such as:

$T \vdash L \equiv \neg Tr(L)$

Then we obtain:

$T \vdash Tr(L) \equiv L$ (based on 1)

$T \vdash Tr(L) \equiv \neg Tr(L)$ (based on the two previous steps and 3)

$T \vdash Tr(L) \wedge \neg Tr(L)$ (based on 4)

□

The paradox is created because the conditions for T seem natural: one *would like* our theory of the world to be exactly like T! However, it turns out that every such theory is contradictory. What shall we do?

Strategy (A) is the solution used by dialetheists[3]. Do the natural assumptions 1 to 4 allow us to obtain a contradiction? Then these assumptions are not controversial, a dialetheist will say. No doubt arises from any steps toward the proof within T, and yet they lead to a contradiction. The conclusion is simply harmless – and here is the solution! Of course, it results in a contradiction but why should we be concerned about a contradiction?

This is when a classical logician enters the stage. "We should be concerned about a contradiction because", he will say, "everything logically follows

[2]To be more precise, condition 2 will be fulfilled if within T we have arithmetic means to prove what is known as the Diagonalisation Lemma. It is fully sufficient if T contains Robinson's arithmetic.

[3]The term was introduced by Graham Priest and Richard Routley (Priest, *et al.* 1989). In Polish and English alike, the terminology is not homogenous. The more common version in English these days is "dialetheism" but one can also come across the spelling "dialethism".

from the contradiction. This is the essence of the principle *ex contradictione quodlibet*! If one accepts the contradiction, they must subsequently accept any sentence, which is not something we would wish for." However, it is exactly *this opinion* of a classical logician – let us stress, this opinion and not the liar reasoning! – that a dialetheist would deem invalid. He notes that the non-classical paraconsistent logic used by him blocks the possibility to infer any sentence from a contradiction. Obviously, a modification is introduced at this point. However, this is the type (A) solution because the liar reasoning itself is left intact by our dialetheist. Only the conclusion is made harmless.

About the strategy (B), Barbara Stanosz writes as follows:

> [It] is difficult to apply because the authors of well-known paradoxes had usually taken great care to make their inferences logically valid. The only exception I know of is an analysis of Zeno's paradox of the arrow (Stanosz, 2004/2015, p. 6).

If only for this reason, a different illustration is worth introducing. Again, I will use the liar paradox.

First, however, let us consider what exactly a solution utilizing strategy (B) entails. Every reasoning requires the use of some rules of inference. The rules should not be confused with premises: they are dynamic elements of the deductive system; they are what allow us to *move* from assumptions to conclusions. To solve a paradox using strategy (B) is to question the correctness of some rules of inference used in a paradoxical reasoning. Then we say: this rule, which we thought correct, is invalid after all.

As mentioned before, it is not absolutely clear what means the intuitive liar reasoning employs. They are only exposed by a more accurate, formal description. It has already been emphasised that the liar reasoning can be recreated in various formal systems. Let us now consider another version of it.

Liar – version 2.

Let us assume that a theory S fulfils the following conditions:

(a) S contains all substitutions of the expression $Tr(\varphi) \to \varphi$.

(b) There exists a sentence (L) such that $S \vdash L \equiv \neg Tr(L)$.

(c) The laws of classical logic apply to S.

(d) For any sentence φ, if $S \vdash \varphi$, then $S \vdash Tr(\varphi)$.

It should be emphasised that such a theory S does not need to contain all substitutions of the equality scheme "$Tr(\varphi) \equiv \varphi$". Condition (a) exclusively specifies implications, not equalities. Despite that fact, it turns out that:

Observation 2. Every theory S that fulfils conditions (a) – (d) is contradictory.

Proof. Based on (b), let us take a sentence (L) such that $S \vdash L \equiv \neg Tr(L)$. We obtain:

$S \vdash Tr(L) \to L$ (condition (a))

$S \vdash \neg L \equiv Tr(L)$ (based on the choice of L and condition (c))

$S \vdash Tr(L) \to \neg L$ (rules of classical logic applied to step (2))

$S \vdash \neg Tr(L)$ (rules of classical logic applied to steps (1) and (3))

$S \vdash L$ (based on (4), rules of classical logic and the choice of the sentence L)

$S \vdash Tr(L)$ (based on (d))

$S \vdash Tr(L) \wedge \neg Tr(L)$ (rules of classical logic applied to steps (4) and (6))

□

However, the conditions for the theory S again seem convincing and desirable. And thus, there is a paradox again; and again, we are facing the question of how to avoid a disaster.

One of the possibilities is to reject the condition (d). At this point, let us notice that condition (d) corresponds to a *rule of inference* known in the literature as "NEC"[4]. Based on this rule, we are allowed to add the expression $Tr(\varphi)$ to the proof if we have earlier proved φ[5]. By rejecting this rule, we use the (B) type strategy: what we question is the validity of one of the steps in the reasoning. This is when we say: this rule is incorrect!

(It is worth adding that some logicians have *indeed* followed this path and so invalidated the given reasoning: all steps in the above proof are

[4]From *necessitation*. This type of rule is a part of modal logics: if we have obtained a proof of a sentence φ in modal logic, we can add "it is necessary that φ" to the proof. In our rule, necessity is replaced by truth.

[5]The NEC rule should not be confused with the implication "$\varphi \to Tr(\varphi)$". Examples are known of non-contradictory theories with an unlimited (i.e. applicable to any sentences) NEC rule in which not all such implications will be theorems in that theory.

re-created in their theories except for the transition from (5) to (6). This is not an *ad-hoc* example![6])

Strategy (C) is – let us remind ourselves – to prove the falseness of one of the premises. In the case of the liar paradox, a popular move is to question some premise in the form of "$Tr(\varphi) \equiv \varphi$" (from the first version of the paradox). For example, one can claim that the truth predicate is *stratified*. The supporters of this conception argue that in fact we are not dealing with one language containing the truth predicate but a family of languages containing predicates of increasing levels (Tr_0, Tr_1, Tr_2,...), which express the truthfulness of sentences in the languages that are one level lower in the hierarchy. For instance, let J_0 be the language of the arithmetic of addition and multiplication without any predicates except for the symbol of identity. A language J_{n+1}, in turn, will be defined as the extension of Jn by a new, one argument predicate symbol Tr_n. Now we can also consider a family of T_n theories which fulfil the following conditions:

1. T_n contains all substitutions of the expression $Tr_n(\varphi) \equiv \varphi$ for the sentences φ of the language J_n,
2. T_n contains arithmetic,
3. The laws of classical logic apply to T_n.

Can we recreate the liar paradox within the theories T_n? It turns out we cannot. For instance, let us consider the theory T_0. If T_0 contains arithmetic, a liar sentence for the truth predicate Tr_0 being a part of this theory's language will exist, i.e. there will exist a sentence such that:

$$T_0 \vdash L \equiv \neg Tr_0(L).$$

However, the construction analysis of the sentence L shows that L is not a sentence of the language J_0 – in fact, L itself contains the truth predicate Tr_0 and hence belongs to the language J_1, not J_0. Other than in the classical liar reasoning, the first condition does not allow us to obtain the equivalence:

$$T_0 \vdash Tr_0(L) \equiv L.$$

Yet this is exactly the key equivalence for the inference of contradiction. Thus, we block the paradox.

[6]The entire reasoning except for the transition from (5) to (6) is recreated in the axiomatized Kripke-Feferman truth theory (known in literature as KF).

Let us emphasise the fact that stratification indeed leads to a type (C) solution. So far, we have only said that we will not obtain the equivalence $Tr_0(L) \equiv L$ within T_0. This is not a drawback of this theory; on the contrary: this equivalence being false in the intended interpretation of J_1 (i.e., the language of the theory T_0) is exactly the point. The mentioned intended interpretation is the model $(\mathbb{N}, T)$ where $\mathbb{N}$ is the standard arithmetic model and T is a subset of $\mathbb{N}$ consisting of sentence codes of J_0 (i.e., arithmetic sentences). It is then easy to notice that:

- $(\mathbb{N}, T) \models \neg Tr_0(L)$, because L does not belong to J_0, so L does not belong to T,
- $(\mathbb{N}, T) \models L$, because L is equal to the sentence $\neg Tr_0(L)$, which is true in $(\mathbb{N}, T)$.

Thus, the equation $Tr_0(L) \equiv L$ is false in $(\mathbb{N}, T)$ and it is as such that we reject it! Let us stress: this is a type (C) solution.

(As a pre-emptive remark, it is worth stressing that sentences in a form $Tr_0(\varphi)$ where φ *contains* the predicate "Tr_0" are grammatically valid. No syntactic rule forbids their construction. They are valid but false, sharing the sorry fate of sentences such as "$0 + 0 = 1$").

The last strategy described – type (D) – is questioning the meaningfulness of a premise. Only in this case (type (D)) did Barbara Stanosz illustrate the liar antinomy. What is the illustration? Barbara Stanosz writes:

> The common feature of most (variously formulated) solutions to the liar paradox is that they treat semantic notions as systematically syntactically ambiguous. What we actually have, instead of two notions "true" and "false," are infinite families of notions: "true0," "true1," "true2,"..., "false0," "false1," "false2,"..., and, furthermore, when you have a sentence predicating truth or falsity about a sentence that itself features "true" or "false" with the subscript x, syntactic coherence demands that it contain the appropriate term with the superscript $x + 1$. In light of this requirement, what we have marked as S above [L in this paper] is not a well-formed sentence of any language. (Stanosz, 2004/2015, p. 8)

This is one of the few fragments of Barbara Stanosz' article that I am forced to disagree with. By my assessment, it is fairly uncommon in today's

source literature to make such a condition of syntactic coherence. The standard approach is different: for *any* one-argument predicate P and *any* term t, the expression $P(t)$ is usually deemed grammatically valid[7]. This applies particularly to predicates such as "$truth_0$" or "$truth_{500}$". It also applies to the terms which, interpreted naturally, refer to sentences containing predicates with even higher indices.

I am guessing, of course, that Barbara Stanosz wanted to characterize a language similar to that of Russell's theory of types. I also agree that such languages can be formally described[8]. The problem is, it is rarely done nowadays. Why? Well, probably because such a complication of syntax theory is simply *not viable*. We can use hierarchical truth predicates *without* complicating the syntax; we then consider the problematic "mixed type" expressions grammatical but false[9]. It is much easier this way. As a result, the syntactic rules of the theory of types are no longer "The common feature of most (variously formulated) solutions to the liar paradox" – in fact, they rarely appear there at all.

Much more often, the (D) type solution questions the *differently* (not syntactically) *understood* meaningfulness of one of the premises of antinomic reasoning. Barbara Stanosz notices this direction of thinking when mentioning the attempts to prove that:

> S [the liar sentence] is either ungrammatical or does not constitute a complete, autonomous unit of natural language and, as such, cannot be true or false; in a sense then the meaningfulness of S is being questioned here along with the role S plays in the liar paradox. (Stanosz, 2004/2015, pp. 8–9)

Let us note that, in the cited fragment, the ungrammaticality of the liar sentence is just one operand of a disjunction! Let us focus on the other operand now. Indeed, in many attempts at a solution which are popular nowadays, the liar sentence is denied logical value. If we identify meaningful

[7]In other words, the expression $P(t)$ is considered a *well-formed formula*, not a sequence of symbols from outside the set of well-formed formulas of the given language.

[8]Not only the truth predicates but also all terms of the described language, starting with simple variables, would have to have type indices. The basic restriction here would be for an atomic formula $\mathrm{truth}_i(t^k)$ to be a well-formed formula only under the condition that $i = k + 1$.

[9]Which does not mean all "mixed type" expressions are considered false. For example, the sentence "It is not true_0 that $truth_0$("$0 = 0$")" mixes types but is true with the intended interpretation: indeed, the sentence "$truth_0$("$0 = 0$")" contains the $truth_0$ predicate, so it is not itself subject to the given predicate.

sentences with sentences that say something particular about the world – something true or false – then the solution would be indeed to question the meaningfulness of the liar sentence.

In this case, the intuitive version of the liar reasoning does not lead to a contradiction. We have earlier considered two cases – the truthfulness of L and the falseness of L – and proved that none of them is possible. This is a good thing: that way, we know that L is neither true nor false! On a formal level, Saul Kripke (1975) described this strategy in detail in his paper *Outline of a theory of truth*. Importantly, the formal description presented in the paper does not contain any type indices: it features an established language distinguishing one predicate "Tr", which gains better interpretations at each stage until the final step, where we obtain the desired effect: it turns out that for any sentence φ, the logical value of φ is identical to the logical value of $Tr(\varphi)$. Except that... besides true and false, there is another logical value: undetermined. It might happen that both φ and $Tr(\varphi)$ are undetermined; this is what happens to the liar sentence. Moreover, a good intuitive interpretation of the "undetermined logical value" is *lack of logical value*, very much in the spirit of the strategy (D). In any case, Kripke's work shows that one can build a formally rigorous interpretation of a language containing its own truth predicate.

The finishing fragments of Barbara Stanosz' article contain a number of very sceptical remarks on the possibility of applying the solutions of the liar paradox proposed by logicians for semantic analysis of natural language. The author notes that the suggested solutions are, "[o]f course, [...] not a description of the actual use of semantic concepts in any of the previously existing languages"; they should rather be "a prescription of how to use semantic concepts in order to avoid contradiction." This remark is, by all means, justified, though the ambitions are bigger in some cases. For instance, Kripke writes:

> I do hope that the model given here has two virtues: first, that it provides an area rich in formal structure and mathematical properties; second, that to a reasonable extent these properties capture important intuitions. [...] It need not capture every intuition, but it is hoped that it will capture many. (Kripke, 1975, p. 699)

In particular, Kripke's model divorces the idea of stratification: the theory of one non-stratifiable predicate indeed seems closer to natural language than the hierarchical approach. This resonates with Barbara Stanosz' comment:

> a grammar is not an adequate description of language if it excludes from the set of sentences (as nonsensical or non-autonomous) many expressions used in communication as independent sentences. (Stanosz, 2004/2015, p. 9)

This is where one could add: in acts of language communication we indeed use one truth predicate, also towards sentences containing this predicate. A grammar that excludes such utterances does seem to be an inadequate description. Still, is that in itself a reason to be sceptical? Logicians have created tools that allow them to cope with more than one such construction!

Questioning of the meaningfulness of the liar sentence makes the author uneasy. She asks: "How can one secure such a claim? The task seems hopelessly difficult" (Stanosz, 2004/2015, p. 9).

According to Barbara Stanosz, one encounters the following problem:

> if [such a claim] is to escape the charge of being ad hoc, such a defense of the ordinary notion of truth must cast doubt on the meaningfulness of [the liar sentence] S along with a whole class of expressions with a similar structure. Yet [...] there are a multitude of expressions that bear close structural resemblances to S but which [...] raise no suspicions. More specifically, one should not dismiss as senseless all self-referring statements. (Stanosz, 2004/2015, p. 9)

The last remark is undoubtedly well aimed and applies not only to natural language. It is known that higher-order arithmetic theories have sufficient means to – in a sense – refer to the expressions of their own language (the Gödel numbers of the expressions of the language of arithmetic). I agree without reservation that to deny these abilities to natural language is not a sensible course. That said, do we have to deny meaningfulness (i.e., logical value) to sentences that "bear close structural resemblances" to the liar sentence? Having denied the meaningfulness of A, should we, as a general rule, deny the meaningfulness to all sentences with the same structure as A? This is, after all, a very dubious claim. The lack of meaningfulness might be a result of the semantic characteristics, not structural ones! It is the case in Kripke's construction. It can even happen that two simple atomic sentences – say, $Tr(t)$ and $Tr(s)$ – with an identical term-predicate structure are classified in two different ways: one determined (true of false), the other undetermined. The deciding factor is the *semantic characteristics* of the sentences, not their syntactic structure. What is wrong with it?

However, there is no pretending: there are some serious issues. One of these issues is the so-called "reinforced liar paradox", which results from the consideration of the following sentence L':

$L'$is false or L' is neither true nor false.

This time, considering three possibilities (not two, as in the previous case), we once more arrive at a contradiction. The constatation that L' is neither true nor false does not help this time, for if L' is indeed devoid of logical value, it seems that L' is true after all!

This reinforced paradox is where Kripke's theory does poorly. Things are even worse: various semantic concepts from the literature encounter their own versions of the reinforced liar. I do not know a theory free from the revenge problem[10].

* *

*

The research on the applicability of new formal truth theories to natural language is in its infancy. Additionally, it must be admitted that the proposed formal theories have their own serious issues. "Nonetheless, at the bottom of existence, at its very foundations, sticks some hellish nonsense, and it is a boring nonsense too" – wrote Stanisław Ignacy Witkiewicz in his *Farewell to autumn*. I think that, having faced the unyielding matter of natural language, Barbara Stanosz would have agreed with the first part of his opinion. I could never, ever, believe that she would have agreed with the second part.

Bibliography

Kripke, Saul (1975). Outline of a theory of truth. *The Journal of Philosophy*, *72*(19), 690–716.

Graham, Priest, Richard, Routley, Norman, Jean (Eds.) (1989). *Paraconsistent Logic: Essays on the Inconsistent*. Munich: Philosophia Verlag.

[10]In literature, the name *revenge problem* encompasses a whole wide family of such problematic phenomena. We believe, to our self-satisfaction, to have solved the liar paradox when suddenly... the liar takes a revenge and comes back in the reinforced, vicious form!

Stanosz, Barbara (2004). Rozwiązywanie paradoksów. *Studia semiotyczne*, *XXV*, 27–31; English version: Paradox resolution. *Studia Semiotyczne — English Supplement*, *XXV* (2015), 5–9, translated by Witold Hensel.

Stanosz, Barbara (1999). Problemy definicji prawdy dla języka naturalnego. In *Logika języka naturalnego*. Warszawa: Polskie Towarzystwo Semiotyczne.

Originally published as "Paradoksy Barbary Stanosz" *Studia Semiotyczne* 28–29 (2015), 51–62. Translated by Agnieszka Przybyła-Wilkin.

STUDIA SEMIOTYCZNE—ENGLISH SUPPLEMENT, vol. XXIX (2017), 62–81
ISSN 2451-2745
DOI: 10.26333/stsen.xxix.04

Paweł Grabarczyk[1]

Directival Theory of Meaning Resurrected[2]

Abstract The first aim of this paper is to remind the reader of a very original theory of meaning which in many aspects has not been surpassed by subsequent theories. The theory in question is Kazimierz Ajdukiewicz's Directival Theory of Meaning. In the first section I present a version of this theory which, I trust, retains the gist of the original but loses its outdated language. In the second section I analyze some problematic consequences of the directival theory (specifically Tarski's counterexample) and show how they can be addressed.

The second aim of this paper is exploiting some of the similarities between the directival theory and later theories of meaning. In the third section I argue that using the directival theory as an interpretative tool enables us to create explications of some of the notoriously vague notions which contemporary theories of meaning employ.

Keywords Kazimierz Ajdukiewicz, directival theory of meaning, semantics, indeterminacy of translation, Wilfrid Sellars

There are two aims of this paper. The first aim is to remind the reader of a very original theory of meaning which in many aspects has not been surpassed by subsequent theories. The theory in question is Kazimierz Ajdukiewicz's Directival Theory of Meaning (henceforth DTM). It was the world's first foray into functional role semantics, predating Wittgensteinian intuitions of "meaning as use" (Wittgenstein, 1967) by almost 20 years. Despite this it has never been widely recognized or analyzed outside of Poland (apart

[1]University of Lodz, Faculty of Philosophy and History.
E-mail: pagrab@gmail.com

[2]This paper is part of the project financed by the Polish National Science Centre, Grant number 2014/15/B/HS1/01928.

from passing remarks by Carnap (Carnap, 1959) and Quine (Quine, 2013, p. 59)). There are two reasons for this. The first reason is that its original presentation leaves a lot to be desired when it comes to accessibility. In section 1 I present a version of the DTM which, I trust, will retain the gist of the original version but lose its outdated language and will simplify it as much as possible. The second reason is that it was quickly abandoned by Ajdukiewicz because of its counterintuitive consequences. In section 2 I will analyze these consequences (specifically Tarski's counterexample) and show how they can be addressed.

The second aim of this paper is to exploit some of the similarities between the DTM and later theories of meaning, specifically the theories of Wilfrid Sellars, Ned Block, Jerry Fodor and Willard Van Orman Quine. In section 3 I will show that apart from being a theory of meaning DTM can also be used as a pretty robust interpretative tool. I argue that using DTM in this manner not only helps us to understand these theories better but also enables us to create explications of some of the notoriously vague notions these theories employ.

1. Directival theory of meaning explained

The directival theory of meaning was developed by Kazimierz Ajdukiewicz over two papers: *O znaczeniu wyrażeń* (*On The Meaning Of Expressions*)[3] and *Sprache und Sinn* (*Language and Meaning*)[4]. Although only the latter paper presents the full-blown version of the theory, it is important to remember about the former as it contains some preliminary considerations that have shed much needed light on assumptions which are crucial for understanding the theory[5].

It is worth starting with the central intuition that motivated DTM. It is so ubiquitous and common that it could be summed up in a popular slogan: "People do not argue over semantics". What this means is that sometimes the argument between two sides reaches a point where the sides start to suspect that the disagreement is merely verbal.

[3]The original Polish version has been published in (Ajdukiewicz, 1985b) and can also be found in (Ajdukiewicz, 1985b), the English translation can be found in (Ajdukiewicz, 1978b).

[4]The original paper can be found in (Ajdukiewicz, 1934), the Polish translation can be found in (Ajdukiewicz, 1985a), the English translation can be found in (Ajdukiewicz, 1978a).

[5]Some researchers consider both papers to be two different versions of the theory (Hanusek, 2013).

What happens next is interesting because of two reasons. The first interesting thing is that (for most of the time) people know how to test their suspicion. Contrary to what might seem to be the obvious solution they do not expect their interlocutors to provide a full definition of the problematic expression. Instead, they try to detect the suspected verbal difference by asking a few key questions about the expression. So, for example, if I was to discover that my interlocutor uses the term "idea" the same way I do, I may start by asking if "ideas" are mental entities. If the answer indicates a difference in usage, it might be enough to decide that the dispute was only verbal, that she meant something different – e.g. platonic ideas.

The second interesting thing is that the moment the two sides discover that the difference was only verbal the disagreement disappears[6]. Most of the time people do not have the motivation to fight with conventions because there is no right or wrong there and some of the conventions are mandatory: either you accept the convention and stay with the community that supports it, or you do not and you are automatically excluded from that community. Starting with these common sense observations Ajdukiewicz presumed that for every noncompound expression there are mandatory conventions and that they are adhered to in the act of confirming certain sentences. When someone knows the meaning of a given expression, and are then asked about it, they have to confirm certain sentences that this expression figures in. And if they refuse to do so, they are excluded from the community of users of this particular expression. Naturally, the model examples of these mandatory conventions are analytic sentences. For example, if you refuse to confirm a sentence "A circle is a figure" then you will be denied the knowledge of the meaning of the term "a circle"[7] and once it is revealed that there is a (admittedly unspecified) number of expressions you do not know the meaning of, you will not be treated as an English speaking person.

The novel idea Ajdukiewicz adds to these observations is his insistence that it should work both ways – if you accept a certain set of sentences which contain a given expression, you can be said to know its meaning. There is nothing more to it – to know the meaning of a word is to have a disposition to confirm its meaning directives (as the specified set of obligatory sentences

[6]Or is vastly diminished. The point here is that it is significantly easier to achieve agreement, even if we have different views on which of the available dictionaries is to be treated as obligatory.

[7]Of course you might as well be denied the knowledge of the meaning of the word "a figure" but it will be tested the same way – you will be asked to accept some other sentences the term "a figure" figures in.

Directive type	Axiomatic	Inferential	Empirical
Character of S	Anything	Set of sentences	Physical stimulus
Example of S	"p implies q"	"p obtains"	The touching of a nerve
Sentence to confirm	"A is identical to A"	"q obtains"	"It hurts!"

Table 1. Types of meaning directives

are to be called). So, what do these meaning directives look like? In general a directive can be presented as a sentence in the form:

> If u is a user of a language L and u is in a situation S then u confirms a sentence p.

It is easy to see that the normativity of meaning is built into the directives from the start. Using a simple rule of contraposition we can derive the following consequence: if someone does not confirm sentence p than either they are not in the situation S or they are not a speaker of the language L. It means that if the user is allowed to disregard language directives they are automatically excluded from a given speaking community[8]. One thing to keep in mind is that what we talk about is the act of confirmation of a sentence and not the act of utterance. It is worth pointing this difference out because ignoring it may easily lead to a significant misinterpretation. The theory does not require the user to produce utterances automatically whenever they are in a given situation but only to react accordingly whenever they are asked to confirm the sentence p in a proper situation. Again, analytic sentences are a good example here. We are not expected to walk around and whisper them to ourselves all the time. What is expected of us instead is a constant, enduring disposition to confirm them when asked to.

We can now group the meaning directives into three sets depending on the type of situation S.

Now let me characterize the types of directives indicated above.

[8]Needless to say it is an idealization. The forbidden behavior would have to be somewhat systematic for her to be really excluded. The important part is that the behavior would be treated as an error and not as an expression of their (even very peculiar) point of view.

In the case of axiomatic directives there are no requirements specified as to what situation S has to be. It can be any possible stimulus – verbal, physical or a combination of both. There can be no stimulus at all. The point here is that in each and every situation the user is expected to confirm some of the sentences of her language (such as the identity statement used in the table).

The inferential directives seem to be another intuitive example of the idea of obligatory rules: after all, this is how most of us learn logic – we are told that whenever we confirm a given sentence we have to confirm another, subsequent sentence. If we do not follow those instructions we will not master logic because it is exactly what mastering logic boils down to. This normative aspect of logic works exactly the same way as it is supposed to work in the DTM.

Last but not least, we have empirical directives. It is important to note that the way I explain them here presents the most significant departure from Ajdukiewicz's version. As can be seen in Table 1, I have described the situation S which precedes the confirmation of the sentence p as a physical stimuli. Contrary to this, Ajdukiewicz referred to mental states rather than to their physical causes. But despite the psychological language that he was using most of his examples of empirical directives adhere to physical stimuli and not their mental correlates. Case in point: in the example I have used in the table above Ajdukiewicz talks about the expected confirmation of the sentence "It hurts!" when a dentist touches the nerve of a patient's tooth and not about the feeling of pain[9].

There is an additional difficulty that most of the examples of empirical stimuli lead to. If I am presented with an object and asked to confirm the sentence "This object is red", I may refrain from doing so because I believe that the lighting in the room is so different from normal lighting that I am no longer sure of the object's color. It complicates matters because we have to expand the directive by a requirement that the user has a belief that the situation (understood as a state of the environment and the perception apparatus) is typical or normal. The addition of beliefs introduces a hybrid category of directives, a mix between the empirical and the inferential ones, one part of the situation S being a sentence expressing the belief and the other being a stimulus. Ajdukiewicz mentions this complication but does not elaborate on it (Ajdukiewicz, 1934). I too am going to skip it in the present exposition of the DTM.

[9]The other important reason for preferring physical stimuli over mental states is that it will make our task in section 3 much easier.

So, how is the notion of meaning to be derived from these three types of directives? Let us assume that we created a list of directives for every noncompound expression of the language. Once we have it, the next step would be to get the notion of synonymy. The intuitive formulation of the relation between the meaning directives and synonymy is this: expressions are synonymous when the meaning directives describe them identically. To present the notion of synonymy in a less metaphoric fashion we have to use an example of a very simple language. Let us say that it contains only the following axiomatic directives[10]:

$$P(a),\ S(c),\ R(c),\ R(d),\ Q(b),\ P(b),\ P(c),\ Q(a)$$

Now, focus on terms a and b. The interesting thing about them is that if you switch their places – replace every instance of a with b and vice versa you will end up with the same list of directives – the only difference being the order of the directives:

$$P(b),\ S(c),\ R(c),\ R(d),\ Q(a),\ P(a),\ P(c),\ Q(b)$$

Using this observation Ajdukiewicz proposed to use this operation of systematic simultaneous replacement of terms to define the notion of synonymy:

> Expressions a and b are s y n o n y m o u s iff they can be simultaneously replaced in all respective meaning directives without changing the sum of all the meaning directives of the language.

The obvious next step is to use abstraction to obtain the definition of meaning:

> The m e a n i n g o f a n e x p r e s s i o n is the set of all the expressions which are synonymous with it.

It is easy to see that in most cases this definition yields rather disappointing results: in the case of expressions which are not synonymous with any other expression their meaning turns out to be a singleton consisting only of themselves. To counter this, Ajdukiewicz introduces a new (and at the time rather novel) idea: he proposed to define meaning by appealing to the notion of translation. To present it, we will use another example of a simple language, let us call it L. Let L contain the following terms: two

[10]That these are axiomatic directives can be easily deduced from their syntactic structure. Only axiomatic directives can be presented as a single sentence.

one-place predicates: – $P(x)$, $Q(x)$; three constants – a, b, c; one zero-place predicate (a sentential constant) Z. Additionally we introduce three symbols which signify physical stimuli: α, β, γ. It is important to stress that these symbols are not parts of L. They symbolize the extra-linguistic element in empirical directives. Now assume that L contains the following directives:

Axiomatic directives:

1. $P(a)$[11]
2. $P(a)\,\&\,Q(b)$

Inferential directives:

1. $P(a) \models Q(b)$[12]
2. $P(a)\,\&\,Q(b) \models Q(c)$[13]
3. $Q(b) \models Z$

Empirical directives:

1. α; Z[14]
2. β; $Q(b)$
3. γ; Z

Having all this we are ready to build something Ajdukiewicz called a language matrix[15]. A language matrix is divided into three sections

[11]Understood as: "in every situation confirm the sentence P(a)" and so on.

[12]Understood as: "If you confirm the sentence $P(a)$ you have to confirm the sentence $Q(b)$" and so on.

[13]You might be surprised that, given the existence of the axiomatic directive 2 and the inferential directive 2 the sentence $Q(c)$ is not an axiomatic directive as well. After all, it is a consequence of these directives taken together. The point of this example is to show that some of the consequences of the language rules are not by itself language rules and can be overlooked by the language user. This characteristic of the DTM will be used later in section 3.

[14]Understood as "In this situation (when the situation is α) confirm the sentence Z" and so on (I use an indexical term to stress the extra-linguistic aspect of α).

[15]This part is a substantially modified version of the original example. First of all, I use a modern predicate logic notation and secondly, I present the matrices in a more visual way which I believe makes the whole idea much easier to grasp.

corresponding to three types of directives. In our example they are designated by numerals 1, 2 and 3 in the first column of the table. Horizontally the table is divided into two parts indicating two parts of a directive: the first part (designated by the Roman numeral I) contains the situation specified by the directive (or the lack of a specified situation in the case of axiomatic directives), the second part (designated by the Roman numeral II) contains the sentence which the directive requires to be confirmed. Every sentence put into a language matrix is divided into its constituent parts using the following procedure: the first cell contains the sentence itself, the next cell contains its main connective or a predicate (in the case of an atomic sentence), the next cell contains the first argument of the connective (or an argument of the predicate). Then the same procedure applies to the first argument – we put its main connective first, then its first argument and so on. When we achieve the level of atomic parts we move on to the second argument of the main connective of the sentence we started with. The pattern is repeated for as long as there is nothing more to decompose. If we applied this procedure to our simple language we would end up with the following table (note the extra-linguistic part in the left bottom corner).

	I								II							
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
1.1									$P(a)$	P	a					
1.2									$P(a)\,\&, Q(b)$	&	$P(a)$	P	a	$Q(b)$	Q	b
2.1	$P(a)$	P	a						$Q(b)$	Q	b					
2.2	$P(a)\,\&\,Q(b)$	&	$P(a)$	P	a	$Q(b)$	Q	b	$Q(c)$	Q	c					
2.3	$Q(b)$	Q	b						Z							
3.1	A								Z							
3.2	B								$Q(b)$	Q	b					
3.3	Γ								Z							

Table 2. The language matrix of L

The main point about a language matrix is that it enables us to extract the structure of the language and abstract away from the actual expressions it uses. We could do that in a variety of ways but I find it the easiest to simply use some sort of visual indication. To extract the structure we are

interested in we simply replace the symbols with graphical patterns; let us call it an expressionless language matrix.

	I								II							
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
1.1																
1.2																
2.1																
2.2																
2.3																
3.1	*A*															
3.2	*B*															
3.3	*Γ*															

Table 3. The semantic structure of L
(expressionless language matrix)

Now you could fill this table anew using the following rules:

1. You do not change the α, β, γ records as they are extra-linguistic elements of the table.

2. You do not fill the white records.

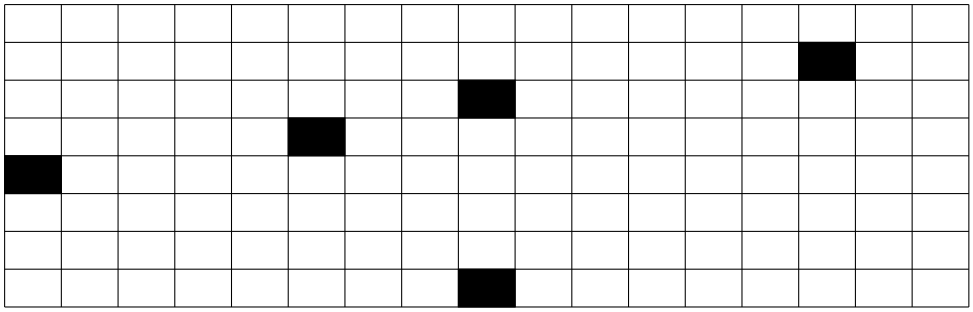
3. Whenever you put something in the record you have to repeat the same symbol in every record with the same pattern.

Every table obtained this way represents a language, which is translatable to the language we started with. Finally, the idea of a language matrix gives us the possibility to define meaning:

> The meaning of a noncompound expression t in the language L is an ordered pair $\langle SL, P\rangle$ consisting of the structure of L (SL) and the set of places t occupies in this structure (P)[16].

[16]The relation of synonymy can still be defined using the notion of mutual exchangeability in meaning directives, just like we did on page 67.

As you have seen, the structure can be presented in the form of a language matrix and the set of places a given expression occupies can as well be shown visually. So, for example, the meaning of the expression $Q(b)$ from our table can be presented via the following diagram[17].



It shows that the DTM realizes the noble goal of a reductive, syntactic definition of meaning – the meaning can be literally represented as a shape, which makes it easy to handle mechanically. The fact that what we started with are the acts of confirmation of sentences just adds a dash of pragmatics to the definition. Because of this, the DTM could not be called a purely syntactic theory. The fact remains, though, that it is a theory in which no part uses any semantic notion. It is an idea entertained by many, but I guess that it is summed up most eloquently by Chomsky:

> It is possible that natural language has only syntax and pragmatics; it has a "semantics" only in the sense of "the study of how this instrument, whose formal structure and potentialities of expression are the subject of syntactic investigation is actually put to use in a speech community" (Chomsky, 1995).

It is important to realize that even though the language matrix contains an extra-linguistic part, the theory does not stipulate that any of the expressions present in the matrix refer to these extra-linguistic elements. Moreover, even if the theory deals with the confirmation of sentences, in no part does it assume the sentences to be true. You might assume that they are held to be true by the users but it would be an additional assumption the theory does not depend on.

[17] To stress the possibility of representing the meaning of the expression visually I omitted the extra-linguistic parts of the table. It is possible whenever a language matrix is fixed.

2. Directival theory of meaning challenged

Ironically, this attractive feature of the DTM (its independence from reference) is exactly what killed it. It happened because of a very simple example that Alfred Tarski confronted Ajdukiewicz with (he did so in a conversation and it was reported many years later in (Ajdukiewicz, 1978c)). Consider a very simple language of predicate logic (with identity) and add to it two new axiomatic directives:

$$A \neq B$$

$$B \neq A$$

A and B are extra-logical constants which appear only in these very directives. The problem is that the two expressions are mutually interchangeable in all the meaning directives of the language (because there are only two such directives and you can mutually replace them). On the other hand, we have to assume that both expressions do not refer to the same object, because it is precisely how we normally interpret the negation of the identity sign. It means that the DTM allows two expressions to have the same meaning but a different reference and it seems that we do not have any means within the theory to block this unintuitive result because the theory does not say anything about the reference of the expressions[18].

It turns out that in spite of deliberately ignoring all the semantic notions Ajdukiewicz still wanted his theory to be Fregean – the meaning of the expression was supposed to determine its reference. It was so obvious to him that he did not even try to argue for it and remarked only that such a consequence was unacceptable (Ajdukiewicz, 1978c). Fortunately, it is a sentiment we do not have to share today as there are at least three ways out of the trouble Tarski's example puts us in – ways which do not force us to abandon the reductive, non-semantic aspect of the DTM.

First of all, we can say that the objection works only because the example language does not contain any empirical directives. If it did, they would have differentiated the terms A and B. And in the case of uninterpreted languages there is no problem of reference anyway. This is the solution suggested by

[18]It is worth noting that Tarski's example is very similar in spirit to Fodor and Lepore's objection against functional role semantics. As Fodor and Lepore rightly argue (Fodor & Lepore, 1992, p. 170) the price hybrid theories pay for their flexibility is that there is nothing that prevents a given sentence having the inferential role of "4 is a prime number" but the truth conditions of "water is greenish" (as there is no necessary connection between inferential role and truth conditions).

Ajdukiewicz himself (Ajdukiewicz, 1978c). The question of whether this solution is effective is highly debatable though (see section 3).

The second thing we can modify is the simultaneous interchangeability requirement of the synonymy relation. It has been shown that we can modify this requirement and demand only when the terms A and B can be considered synonymous if and only if it is possible to replace A with B and then B with A (but not simultaneously) without changing the character of the directive we applied this procedure to. This means that if something has been an axiomatic directive, it remains an axiomatic directive after the replacement of the term (similarly for the other two types of directives). This solution has some disadvantages, but they will not be discussed here[19].

The third, and perhaps most interesting option, is that we could simply accept and embrace this surprising consequence of the theory – especially that it is not so surprising anymore. After all, this is what Putnam's Twin Earth thought experiment was set to do – it showed us that we do not have to hold to Fregean intuitions about the relation between meaning and reference (Putnam, 1975). Could not we simply decide that a sensible strategy for a theory of meaning is to contain two parallel theories – a theory of reference and a separate theory of meaning which answers the questions about synonymy, translatability and meaningfulness of expressions?

Unfortunately the DTM has more issues than that. Specifically, there are two problematic theses it holds (one of them being an assumption, the other a consequence) which we have to analyze if the theory is to be useful for contemporary philosophers. We will refer to them later, so it might be convenient to label them:

(T1) The meaning of every word in the language changes whenever a new word is added to the vocabulary.

(T2) Syntaxes of all translatable languages have to be perfectly compatible.

(T1) is a direct consequence of definition (D2) presented above. If the meaning of a particular expression is the ordered pair of a language matrix and a set of places the expression figures in, then the meaning changes whenever the matrix changes, and the matrix changes whenever a new expression is added. It is so because the new expression has to have a set of new directives which regulate its usage and these directives have to be added to the language matrix.

[19]The results in question has been published only recently by (Nowaczyk, 2006) and (Buszkowski, 2010). Unfortunately both articles are only available in Polish.

(T2) follows on from the way the matrices are built and from the introduced notion of translatability. Whenever a given expression A is to be a translation of some term B, both expressions have to figure in the same places in identical language matrices. Such a strict notion of translatability does not allow the translatable expression to differ syntactically. To see why it is so let us consider the opposite situation – let us say that we found two expressions which figure in exactly the same places of their respective language matrices, but one of them is atomic and the other is not. There would have to be a place in the second matrix where the second expression was decomposed into its atomic constituents but there would be no such place in the first matrix (because there was nothing to decompose there). But if the matrices are different then the expressions are by definition not translatable.

It is important to stress that neither of these claims present a serious challenge to the theory – they are simply counterintuitive. Nonetheless, I believe that it is worth pointing them out and analyzing ways of dealing with them because, as I hope to show, even small modifications to these claims produce interesting and useful variants of the theory.

In order to understand how we could deal with the thesis T1 we have to introduce an important requirement that Ajdukiewicz added to the theory. As he points out the directival theory can only be formulated for languages which are coherent and closed.

> A given language is c o h e r e n t if every expression it contains is connected to every other expression (directly or indirectly) *via* meaning directives[20].

In other words – if the language in question is coherent, we should be able to pick any expression and "reach" any other expression by "jumping" from a meaning directive to a meaning directive.

> A language is c l o s e d if for every new expression, which is to be introduced to it, it already contains an expression synonymous with it.

In other words – a closed language is a language that already contains all meanings which can be added to this specific language (as further enrichment would have produced either synonyms or an incoherent language).

[20]Two expressions are directly connected if they figure together in a single meaning directive. Expressions A and B are indirectly connected if they are not directly connected but there exists an expression C such that A and B are directly connected to C.

The bad news is that Ajdukiewicz's requirement creates bigger problems than the problem we wanted to solve with its help (thesis T1). The second requirement is simply much too strong – there are no existing closed languages and, what is worse, we could not create a closed language even if we wanted to (see Buszkowski, 2010).

3. Directival theory of meaning resurrected

In the remaining part of this paper I am going to show how we can utilize DTM as an interpretative tool for other theories of meaning – theories which often lack the precision of Ajdukiewicz's account[21] and which can be seen as sketches DTM fleshes out. What Ajdukiewicz's theory can provide here is showing something which other theories only hint at.

Let us start with a suggestion, which, I hope, will be rather obvious for the reader – the possibility of treating DTM as a theory of narrow content. Let us use the example of Sellars-Block's account because the similarity between it and the DTM is striking. Sellars introduced four types of language rules, depending on whether the character of the stimulus provided for the user and her response is linguistic or not (Sellars, 1963). There are three obvious possibilities:

1. Extra-linguistic stimulus – linguistic response.
2. Linguistic stimulus – linguistic response.
3. Linguistic stimulus – extra-linguistic response.

There is also a fourth, less obvious option:

4. Any stimulus – linguistic response[22].

It is not hard to see that 1. can be understood as empirical directives, 2. as inferential directives and 4. as axiomatic directives. There is nothing similar to 3. in the DTM but what prevents us from adding a new type of directive to the theory[23]? This new category of directives could be called imperative directives – they instruct the speaker to perform a certain action whenever she acknowledges a certain sentence by confirming it.

[21] This account is neatly summarized in (Putnam, 1991).

[22] Sellar calls this type of rule a "free rule".

[23] In fact adding new directive types is a very natural way of extending the theory and deserves further inquiry.

Now, the idea Block adds to the mix is that language described this way can be understood as a network of inputs and outputs which in turn enables us to define the narrow content of an expression (or its "conceptual role", as Block prefers to call it) as a role the expression plays in this computational structure (Block, 1986). The problem with this account is that, while attractive, it does not show us how exactly a set of user actions (sentence confirmations) translates into a network of interrelated expressions of the language. Is the network just a set of beliefs connected by their inferential roles? If so, which ones – all of them? Maybe they should be decomposed somehow or perhaps even translated into language of thought? It is precisely what language matrices can help us with. They start with a set of pragmatic phenomena and then break it down into syntactic constituents of expressions enabling us to see the mechanism that underlies the phenomenon of narrow content.

Speaking of the language of thought – arguably the biggest flaw of this hypothesis is the elusiveness of the language it postulates. What does it look like? What is the ontological status of its expressions? What exactly are its meanings and how can they determine the meanings of natural languages? To see how the DTM could help here let us modify the idea of closed languages and introduce a more liberal (and realistic) notion of s e m a n t i c a l l y p r e d e t e r m i n e d languages.

> A language is s e m a n t i c a l l y p r e d e t e r m i n e d if every new expression introduced to the language is synonymous with a compound expression built from the expressions the language already contains.

What we mean by that is that even if the language does not contain a proper synonym for the new expression, its meaning can be construed out of the language's existing expressions and this is exactly what Fodor assumes (Fodor, 1975). The other thing we have to change is (T2) – we have to decide which syntaxes of translatable languages do not have to be identical. Instead, we assume only that the syntaxes are compatible in a sense that the differences they demonstrate are only superficial and what is important is the identity of deep syntactic structures of both languages[24]. It is possible that it is a solution Ajdukiewicz tacitly assumed anyway. Consider the way

[24]One notable complication is that the relation between a given language and the language we use to show its deep structure could not be explained by the same notion of translation we use in the DTM, but it is a small price to pay.

we build language matrices. What we look for in sentences are connectives, their arguments, their ordering and nothing else. Ajdukiewicz was a pioneer of categorial grammars so it is possible that he assumed that a working theory of meaning presumes a developed theory of universal grammar. In other words – categorial grammar could be thought of as a description of the surface grammar of two languages that is general enough so it abstracts away from unimportant details and enables us to represent two superficially syntactically different expressions as expressions of the same type. What we end up with, then, is a theory which fits the language of thought hypothesis quite well because it gives us the answer it lacked – it shows how the semantic structure of the language can be construed out of its non-semantic aspects. Moreover, it gives us the much needed model of linguistic structure which contains no actual labels or sentences but is still compatible with many different sets of such labels and predetermines the relations between them. The result is a detailed functional model for LOT. We can postpone the question of what the expressions of this language actually are. Instead we point at an expressionless language matrix (similar to the one presented in Table 3) and say only that LOT is anything that works "like that". As a functional semantics the DTM is compatible with different answers to the question about actual expressions. They can turn out to be patterns of firing neurons or parts of the brain or whatever else.

Another theory that could benefit from the DTM is Quine's behavioral theory of meaning. For the DTM to be useful here we would have to modify the requirement of coherency a bit. Let me digress for a second and say a few things about the notion of coherency I introduced earlier, because it proves to be even more useful than Ajdukiewicz had assumed. One disappointing aspect of the DTM I did not talk about is that although it provides the notions of translatability and synonymy, it does not give us any clue as to what it is for a given expression to simply "have a meaning" (as opposed to nonsense words). Does it suffice for an expression to simply be a part of a language matrix? This is where the notion of coherency can help: we can simply assume that an expression is meaningful if it is a part of a coherent language (which means that it is somehow connected to all the other expressions of the language). The problem with this idea is that it renders all the expressions meaningful. Consider the axiomatic directive of identity. The directive instructs the user to confirm every substitution of the formula $x = x$ regardless of the circumstances. What it means is that for every expression of the language there exists a meaning directive of the form $x = x$ where the expression is substituted for x. It follows that every

expression is directly connected to the identity sign and the identity sign is directly connected to every other expression in the language. But it means that every expression is indirectly connected to any other expression.

Now, getting back to Quine's theory – instead of assuming, as we did above, that the meaningfulness of an expression depends on the number of connections to all the other expressions, we should focus only on selected connections, namely on the connections with the non-linguistic, empirical parts of the table (that is the part specified in empirical directives, denoted by Greek letters in matrices). This way we could easily provide an explication for Quine's s t i m u l u s meaning. Note that by doing that we do not have to give up the non-semantic aspect of the theory because Quine's behavioral account does not imply that the expressions refer to stimuli.

If we allow for this modification of the DTM what we get in return is a theory which can be very well understood as a description of the manual constructed by Quine's radical translator. Remember that what the translator was supposed to do was to collect data on sentence confirmation. He collected the sentences which were confirmed in every situation, sentences which were confirmed after certain different sentences were confirmed and sentences which were confirmed whenever the empirical situation was such-and-such (Quine, 2013). It is not hard to see that these three sets of data can be treated as our axiomatic, inferential and empirical meaning directives. Once again – the point here is that this convergence of theories goes far beyond a mere analogy. The DTM can be used to explain the idea which was originally rather vague – the idea of a translation manual (as created by a radical translator). The translation manual is a mapping of two different sets of expressions into a common language matrix.

Compatibility between the DTM and Quine's account is so great that we can easily recreate the infamous consequence of the latter theory, namely the indeterminacy of translation thesis. Let us get back to Tarski's counterexample. One way of looking at the problem it poses is that DTM allows for two expressions to be synonymous contrary to the beliefs of language users. If two expressions play the same role in a language (that is: figure in the same directives in the same places) they are synonymous no matter what. They are synonymous even if no one knows about it. They are synonymous even if the language users believe they are not synonymous! The latter happens when two expressions function the same way but one of the directives specifies that they are not identical (in these very words). A very well-known example of this is Putnam's elm/oak distinction (Putnam, 1975, p. 226). If a given

language[25] does not have tools to differentiate between two meanings, than an empty claim that they are different will not change anything. After all – the very claim still says the same thing about each expression. Let us call this peculiar type of synonymy t a c i t s y n o n y m y.

To see how it generates the indeterminacy thesis consider two languages: L_1 and L_2. Let us say that they are translatable (in the specified sense) and that they both contain tacit synonyms: In L_1, A_1 and B_1 are synonymous and correspondingly in L_2, A_2 and B_2 are synonymous. Now the problem is that if you assume that A1 is a translation of $A2$, then, because $B2$ is a synonym of A2, B2 is just as good a translation of $A1$ as $A2$ is. Of course we might rightfully say that it does not matter whether we translate $A1$ to $A2$ or to $B2$ – all of these expressions are synonymous. The only thing we have to remember is that we keep the structure intact (so both languages still contain a pair of synonymous expressions). There is no "fact of the matter" as to which translation is better – both are just as good (unless you treat homonymy as a sufficient reason). But this is exactly how some of the researchers interpret the indeterminacy thesis (Field, 2001, p. 282). Indeterminacy of translation does not present any genuine skeptical worry. The only thing it does is to subvert our expectations towards meaning.

Additionally, we have to realize that that the language we analyze might contain synonyms on various levels. It can have one-on-one mappings that we have just discussed but it could also contain surprising mappings where a single expression figures in the same set of places as a different, syntactically complex expression. Let us call it s y n t a c t i c a l t a c i t s y n o n y m y. This is something Ajdukiewicz did not foresee but there are no reasons as to why such a situation could not happen. For example, we could realize that for every sentence which figures in language directives and contains the word "a rabbit" there is an analogous sentence which contains a complex expression "an organized set of rabbit parts". The important difference between the current situation and the case of simple tacit synonymy we discussed in the preceding paragraph is that a syntactical tacit synonymy is much more bewildering to us than a normal tacit synonymy. We believe that there is a difference between rabbits and organized rabbit parts but try as we might, we cannot find directives to support this assumption. Tacit synonymy

[25]In Putnam's example they are, of course, idiolects. Although DTM has been originally construed as a theory of language and we retained this aspect of the theory in this paper, it is worth pointing out that the theory can be easily tailored to function as a theory of idiolects.

of syntactically different expressions can then be understood as the main mechanism of indeterminacy of translation.

What I wanted to show in this paper (apart from presenting a version of the DTM better suited for contemporary readers) is that the main reason why DTM had been abandoned by its creator and largely forgotten, namely because of the Tarski objection, is not very serious anymore, because we learned a few lessons along the way and no longer expect the theory of meaning to be also a theory of reference. The other problematic aspects of the DTM can be interpreted as challenges to be met by different, enhanced versions of the theory. Such enhanced versions can then be used as explications of existing ideas, like the notion of narrow content (understood as a product of language rules), or as a framework for the behavioral theory of meaning or as a model for the language of thought hypothesis. The Directival Theory of Meaning is worth resurrecting because it can provide much needed details for propositions which are notoriously presented as sketches or outlines of possible future theories rather than as full-blown accounts.

Bibliography

Ajdukiewicz, K. (1934). Sprache und Sinn. *Erkenntnis*, *4*(1), 100–138.

Ajdukiewicz, K. (1978a). Language and Meaning. In J. Giedymin (ed.), *The Scientific World-Perspective and Other Essays, 1931–1963* (pp. 35–66). Dordrecht: Springer.

Ajdukiewicz, K. (1978b). On The Meaning of Expressions. In J. Giedymin (ed.), *The Scientific World-Perspective and Other Essays, 1931–1963* (pp. 1–34). Dordrecht: Springer.

Ajdukiewicz, K. (1978c). The Problem of Empiricism and the Concept of Meaning. In J. Giedymin (ed.), *The Scientific World-Perspective and Other Essays, 1931–1963* (pp. 306–319). Dordrecht: Springer.

Ajdukiewicz, K. (1985a). Język i znaczenie. In *Język i znaczenie*, vol. 1 (pp. 145–173). Warszawa: PWN.

Ajdukiewicz, K. (1985b). O znaczeniu wyrażeń. In *Język i znaczenie*, vol. 2 (pp. 31–77). Warszawa: PWN.

Block, N. (1986). Advertisement for a semantics for psychology. *Midwest Studies in Philosophy*, *10*(1), 615–678.

Buszkowski, W. (2010). O równoznaczności wyrażeń w ujęciu Ajdukiewicza. In J. Grad, J. Sójka, A. Zaporowski (eds.), *Nauka – Kultura – Społeczeństwo. Księga jubileuszowa dedykowana Profesor Krystynie Zamiarze*. Poznań: Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza.

Carnap, R. (1959). *Introduction to Semantics, and Formalization of Logic*. Cambridge: Harvard University Press.

Chomsky, N. (1995). Language and Nature. *Mind*, *104* (413), 1–61.

Field, H. (2001). *Truth and the Absence of Fact*. New York: Oxford University Press.

Fodor, J.A. (1975). *The Language of Thought*. New York: Crowell.

Fodor, J.A., Lepore, E. (1992). *Holism: A Shopper's Guide*. Cambridge, Mass.: Blackwell.

Hanusek, J. (2013). On a Non-Referential Theory of Meaning for Simple Names Based on Ajdukiewicz's Theory of Meaning. *Logic and Logical Philosophy*, *21* (3), 253–269.

Nowaczyk, A. (2006). Dyrektywalna teoria znaczenia, czyli dramat Filozofa. In J. Pelc (ed.), *Sens, prawda, wartość: filozofia języka i nauki w dziełach Kazimierza Ajdukiewicza, Witolda Doroszewskiego, Tadeusza i Janiny Kotarbińskich, Romana Suszki, Władysława Tatarkiewicza*. Warszawa: Biblioteka Myśli Semiotycznej.

Plato (1998). *Cratylus*. Indianapolis: Hackett Pub. Co.

Putnam, H. (1975). The Meaning of "Meaning". In *Mind, Language and Reality: Philosophical Papers*, vol. 2 (pp. 215–271). Cambridge, Mass.: Cambridge University Press.

Putnam, H. (1991). *Representation and reality*. Cambridge, Mass.: MIT Press.

Quine, W.V.O. (1969). *Ontological Relativity & Other Essays*. New York: Columbia University Press.

Quine, W.V.O. (2013). *Word and Object*. Cambridge, Mass.: MIT Press.

Sellars, W. (1963). *Science, Perception and Reality*. London; New York: Routledge and Kegan Paul; Humanities Press.

Wittgenstein, L. (1967). *Philosophical Investigations*. Oxford: Blackwell.

Originally published in *Studia Semiotyczne* 31/1 (2017), 23–44.

STUDIA SEMIOTYCZNE—ENGLISH SUPPLEMENT, vol. XXIX (2017), 82–106
ISSN 2451-2745
DOI: 10.26333/stsen.xxix.05

Filip Kawczyński[1]

Two Models of Propositional Structure

Abstract This paper is a comparison of two structural theories of propositions: the theory proposed by Kazimierz Ajdukiewicz in the 1960s and the theory developed by Jeffrey King at the beginning of the 21st century. The first section of the paper is an overview of these theories. The second part is a detailed discussion of significant similarities shared by them. In this section, I also identify and analyze ways in which these theories differ and attempt to determine if these differences are substantial or apparent. The last part is an attempt to determine if the discussed theories are capable of coping with the Benacerraf Problem.

Keywords Kazimierz Ajdukiewicz, Jeffrey King, proposition, propositional structure, structural theories of propositions, the Benacerraf Problem, truth conditions

---

Realists regarding logical propositions (cf. Loux 2003) disagree as to the set of properties attributable to propositions.[2] One of the most contentious issues in the debate regarding the nature of propositions is their structure. Generally speaking, the locus of controversy is the question of the divisibility of propositions into constituents. Philosophers holding that propositions are divisible are referred to as the proponents of the structural theory of propositions. According to this theory, propositions are structured objects, and the structure of a proposition corresponds, to a lesser or greater degree, to the structure of the sentence expressing it. The alternative approach to the structural theory is the functional theory, according to which propositions are functions relating possible worlds to truth values.

[1]Department of Philosophy and Sociology, University of Warsaw.
Email: f.kawczynski@uw.edu.pl

[2]Perhaps the only property that does not arouse controversy is abstractness. See Kirkham (2001, p. 57), Loux (2003, p. 137), or Lycan (2002, p. 80).

The functional theory is traditionally taken to derive from Rudolf Carnap's conception of extension and intension (Carnap 1947/2007). In this conception, propositions are identified with intensions of sentences, and truth values with their extensions. It is required that the intension of a sentence unequivocally determine its extension on the one hand, and on the other, that it be possible that two sentences sharing the same extension possess different intensions. Over time, an entire tradition emerged of identifying intension with the function $I : W \to \{0, 1\}$, that is, one relating the set of possible worlds (W) to truth or falsity.[3]

The main group among the proponents of the structural theory of propositions are philosophers developing Bertrand Russell's position presented in *The Principles of Mathematics* (Russell 1903/2008). According to this position, a proposition is, roughly speaking, a complex of objects, properties and relations spoken of in the sentence expressing this proposition insofar as[4] these objects possess these properties and remain in these relations. Russell's position is sometimes referred to as direct realism because, according to it, a proposition expressed by the sentence xyz is in principle[5] constituted by objects x, y and z, not by their representations, concepts, or the meanings of expressions designating them. In other words, one constituent of the proposition expressed by the sentence *Russell is British* is Russell himself – the flesh and blood human being. Propositions of this kind are referred to as singular propositions, and expressions introducing their designates (not their meanings) into these propositions, as directly referential.

A theory inspired by this position and at the same time markedly different from other Russellian conceptions[6], has recently been proposed by Jeffrey King (2007)[7]. One of the distinctive features of this conception is

[3]The most important proponents of the functional theory have been Lewis and Montague. A significant extension of the functional theory is so-called two-dimensional semantics developed by Stalnaker, Kaplan, and Chalmers. A solid Polish language discussion of issues related to the functional theory can be found in (Ciecierski 2003), and those related to two-dimensional semantics, in Odrowąż-Sypniewska (2006, 330–336).

[4]The expression "insofar as" should be understood in a sense independent of the cognitive act performed by the subject since Russell considered propositions to be objective entities, independent of the mind. Compare Russell (1902/2008, p. 33) and Makin (2000, p. 11).

[5]The infamous denoting concepts being the exception. Compare Russell (1903/2008, p. 5).

[6]Soames, Salmon, Richard and sometimes Kripke and Kaplan are considered to be the continuators of those other conceptions. See Deutsch (2008).

[7]King later proposed a revised version of his conception (King 2014). It differs from the 2007 formulation in that much more attention is given to the impact of context on the proposition expressed by a sentence. However, since the fundamental ideas have

its refusal, *contra* many theoreticians of singular propositions, to identify propositions with any kind of formal constructs. King brings propositions back to earth, so to speak, by identifying them instead with a special kind of facts. In this paper, I compare King's famous conception to a less discussed, including in Poland, theory of propositions as functions proposed by Kazimierz Ajdukiewicz (this theory is not a functional theory in the sense indicated above). I am of the opinion that they have enough in common for their juxtaposition to be interesting not only for the historian of philosophy but also for the contemporary philosopher of language. The reason is that their comparison can help shed light on certain nontrivial issues related to the problem of propositional structure and render salient some of the consequences of choosing a particular model thereof.

The first part of the paper is a detailed discussion of the two conceptions. The second part is dedicated to their comparison; here, I indicate similarities and apparent, as I am going to argue, differences between them. In the final part of the paper, I test both conceptions in light of the Benacerraf Problem. I argue that neither of them passes the test since the proponents of both conceptions are forced to introduce *ad hoc* solutions, or to accept difficult consequences, in order to tackle this problem.[8]

1. Conceptions from King and Ajdukiewicz

1.1. King's conception of propositions as facts

According to King's conception, the proposition expressed by the sentence *Rebecca swims* is a fact, although not the fact that might come to mind as corresponding to this sentence at first glance, that is, not the fact that Rebecca has the property of swimming. The fact of Rebecca's having the property of swimming is a truth-maker of the fact-proposition expressed by this sentence[9], but the two facts are not the same. Significantly, as King points out, if Rebecca did not in fact swim, the fact of her having the property of swimming would not obtain (would not exist); the proposition

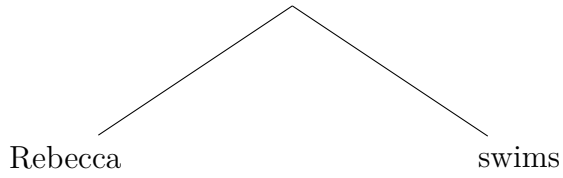
remained unchanged, and King's conception found its most comprehensive expression in his 2007 book, I rely mainly on the latter.

[8] I thank Tadeusz Ciecierski for first pointing out that Ajdukiewicz's and King's conceptions share similarities. He mentions this in passing in Ciecierski (2012).

[9] In other words, the proposition-fact [that Rebecca swims] is true if and only if Rebecca instantiates the property of swimming.

under consideration, on the other hand, would exist, although in these circumstances it would of course be false (cf. King 2007, p. 26).

The methodological background of King's theory is a syntactic sentence analysis based on Chomsky's categorial grammar, and more precisely, on a version of the latter called the minimalist program. In particular, King uses the method of representing the real (deep) syntax of sentences using so-called trees—a method well entrenched in the tradition of syntactic investigations. For example, the syntax of a simple subject-predicate sentence such as *Rebecca swims* can be represented in the following way:



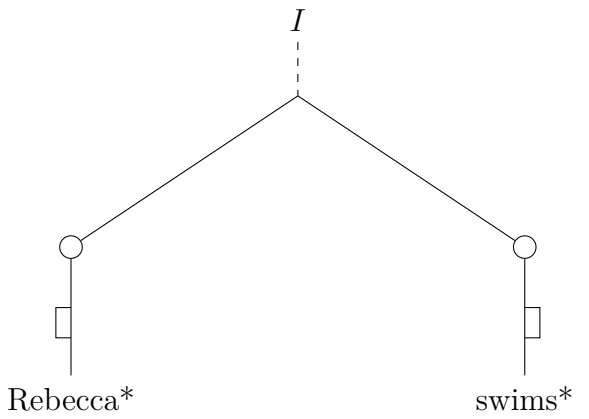
Tree 1. (the sentential relation)

King refers to the relation responsible for binding simple expressions into a complex whole, that is, a sentence, here graphically represented by the branches of the tree, as the sentential relation (King 2007, p. 29). There are two options concerning the nature of sentential relations, according to King. We can assume either that the sentential relation is a nondefinable primitive concept or that its nature is currently impossible to explain, although it might be explained in the future by means of cognitive and neurological concepts (King 2007, p. 47–50). King admits that he is inclined toward the latter option, but the first one does not diminish the value of the proposed description of propositional structure, in his opinion.

According to King, objects constituting fact-propositions include, first of all, properties and relations (such as the property of swimming), and secondly, individual objects, including macroscopic physical objects (such as Rebecca or Mount Everest). A complete tree representing the fact-proposition expressed by the sentence *Rebecca swims* looks like this:

The two branches converging at the root of the tree (its highest point) and shaping the entire structure represent the sentential relation (in short, R) binding the relevant simple expressions into a sentence, as featured in Tree 1. "Rebecca*" represents the physical individual, that is, Rebecca, and "swims*", the property of swimming conceived as an abstract object.[10] The

[10]The asterisks serve to emphasize the fact that it is Rebecca as such and swimming as such that make up the syntax of the proposition, not meanings or concepts etc.



Tree 2

lines linking "Rebecca*" and "swims*" to the small circles represent the semantic relations obtaining between the expression "Rebecca" and Rebecca (or Rebecca*) and the expression "swims" and the property of swimming (or swimming*). These semantic relations are relations of reference, or designation, that is, relations obtaining between expressions and objects constituting their reference.

The aforementioned circles represent a relation referred to by King as joint instantiation. It holds between two properties of the expressions making up the analyzed sentence: the property of referring to a designate forming a part of the fact-proposition and the property of constituting a particular node in the relation R characteristic of the discussed sentence. For example, the circle located on the left branch of the tree represents the fact that two properties of the expression "Rebecca" are jointly instantiated: first, that the expression refers to Rebecca, and second, that the expression constitutes the left node in the relation R, its right node being the expression referring to the property of swimming. Graphically speaking, joint instantiation is the point at which the syntactic relation meets the semantic relations to jointly make up the proposition (this can be seen from the suggestive location of the symbols for joint instantiation in the structure of the tree).

The rectangles located on the branches of the tree representing the relation of reference represent the relativization of the arguments of this relation to the context in which the sentence was used. In the example under consideration, this is meant to signal that "Rebecca" designates Rebecca in this context and "to swim" designates swimming in this context. King refers to the complex relation obtaining between the terms of the

proposition – the relation constituted by the syntactic relation and the (context relative) semantic relations described above – as the propositional relation (in short, P).

The last part of the discussed model is the broken line and the relation I located at its terminal node. The letter "I" represents an instruction, as King calls it, and the broken line linking R and I, the fact that I is encoded by R (one could say that the relation R is the carrier of the instruction I) ((King 2007, pp. 34–38)). An instruction indicates the most general rules for the determination of the truth conditions of the proposition it is an element of. In the case of the proposition discussed here and expressed by *Rebecca swims*[11], the instruction reveals that the proposition can be deemed true if and only if Rebecca has, or instantiates, the property of swimming. Put more generally, the instruction determines the configuration in which the objects and the properties constituting the reference of the expressions located on the individual branches of the relation R must remain in order for the analyzed proposition to be true. In the case of propositions expressed by simple subject-predicate sentences, the instruction usually determines the proposition to be true if and only if the object constituting the reference of the expression located on one branch of R instantiates the property constituting the designate of the expression located on its other branch.[12]

The idea of introducing the instruction I into the structure of the proposition should become clearer once we imagine a natural language in which the structure of the proposition expressed by the sentence *Rebecca swims* belonging to this language looks exactly the same as the structure illustrated in Tree 2 except for one difference: instead of I, the structure features $\tilde{\text{I}}$, determining that the proposition is true if and only if Rebecca *does not* have the property of swimming. It is obvious that this hypothetical language is very different from Polish (and, most likely, the great majority of natural languages). However, its examination points to an important idea in King's conception – namely, that the relation P constituting the structure of the

[11] In fact, the instruction I can be seen as an element of the proposition responsible for determining which constituent of its truth conditions relates to the subject of the proposition and which to the predicate.

[12] In King's approach, propositions can also be represented in an abbreviated linear form (omitting joint instantiation, which must be taken to feature implicitly), where the shape of the relation R is represented by a sequence of square brackets, and C (relativization to context) and I (instruction) are represented by letters located at the beginning of the notation. For example, the proposition [that Rebecca swims] can be represented as {C, I, [[Rebecca*] [swimming*]]}, and the proposition [that Mont Blanc is shorter than Mount Everest] as {C, I, [[Mont Blanc*] [being shorter than* [Mount Everest*]]]}.

proposition is insufficient, in and of itself, to determine even the most general truth conditions of this proposition; the proposition, as a carrier of truth, must also feature an element providing hints regarding its truth conditions. According to King, the syntactic relation R encoding the instruction I is such an element. At first glance, it might seem controversial to link truth conditions – even most generally construed, as is the case in I – to the level of syntax. The controversy subsides, in my opinion, once the role of syntax is taken to be such that the syntactic characteristics of expressions somehow determine their semantic characteristics; more specifically, the syntax determines the order of the semantic values of the particular expressions making up the sentence (that is, their designates). Knowledge of the syntax allows one to recognize this order – the order is in a sense encoded in the syntax.

Having explained what is represented by the particular elements of the fact-proposition model, King gives the following definition of a proposition:

> The proposition expressed by a sentence of the form "$xyz\ldots$" is the following fact: there exist a context C and expressions x, y, z of a language L whose semantic values in X are objects X*, Y*, Z*... and these expressions occur in a particular order determined by the sentential relation R encoding the Instruction I (King 2007, pp. 39, 42).

This formulation is surprising since a proposition is herein identified with the existence of expressions bearing certain syntactic and semantic characteristics, whereas the fact-proposition model presented earlier did not feature expressions at all.[13] Moreover, in another part of his book King gives another characterization of a proposition, corresponding to what is presented using trees:

> [...] the facts that are propositions came into existence in part as a result of lexical items acquiring semantic values and syntactic relations coming to encode certain functions (King 2007, p. 65).

The two explications of the notion of a proposition given by King are not mutually exclusive but they certainly differ, the difference being more than verbal. According to the first explication, a proposition is the existence of expressions etc.; according to the second one, it is the fact that the designates of the relevant expressions remain in a certain order. In the remainder of this

[13] I thank an anonymous reviewer of this article for drawing my attention to this difficulty in King's conception.

paper, I refer to the latter interpretation, that is, I assume that, according to King's theory, the proposition expressed by a sentence of the form "ψ is φ" is the fact that the designate of ψ and the designate of φ remain in a particular order because they are the designates of these expressions. This decision is occasioned, in the first place, by the fact that the second interpretation predominates in King's book, while the first only occurs in the initial parts of the text—it might thus be read as a not too fortunate initial statement. Secondly, there is no doubt that King presents his conception as a structural theory, and identifying a proposition with the existence of expressions bearing certain characteristics does not befit this strategy. King simply does not analyze the structure of the existence of such expressions.

Having established that, we can conclude that, according to King's standpoint, a fact-proposition is something else than the fact intuitively assumed to be the proposition's truth maker. The fact that Rebecca and swimming stand in the relation P – comprising the relation of designation, the sentential relation, and the relation of joint instantiation – is certainly different from the fact of Rebecca's having the property of swimming. Both facts feature Rebecca and the property of swimming, but these stand in different relations in the first and the second fact.

The proposition [that Rebecca swims] identified with an appropriate fact is thus true if and only if Rebecca has the property of swimming or if Rebecca belongs to the extension of this property. In light of this, general truth conditions, as construed in King's conception, can be characterized in the following way (assuming standard I):

> The proposition expressed by a sentence of the form $\varphi(a_1, a_2, \ldots, a_n)$ is true if and only if the objects $a_1, a_2, \ldots, a_n$ belong to the extension of φ.

1.2. Ajdukiewicz's structural-functional conception of propositions

A conception similar in its general outline to the one developed by King had been presented several decades earlier by Kazimierz Ajdukiewicz (Ajdukiewicz 1967/1971). Ajdukiewicz's approach is based on his own syntactic analysis of sentences. The method in question consists in assigning to each expression in the syntax of a sentence an unequivocal description of its syntactic position in this sentence (compare Ajdukiewicz 1960/1985).

According to Ajdukiewicz, the proposition expressed by a sentence can be characterized as a function relating each syntactic position to exactly one

object constituting the reference of the expression occupying this position in this sentence. For example, the proposition expressed by the sentence:

Mont Blanc	*is shorter than*	*Mount Everest*
$(1,1)$	$(1,0)$	$(1,2)$

is a function relating position (1,1) to Mont Blanc, position (1,0) to the relation of being shorter than, and position (1,2) to Mount Everest. Since every function is identical with an appropriate set of ordered pairs, the function constituting the proposition expressed by the above sentence is a set of the following form (cf. Ajdukiewicz 1967/1971, pp. 122–123):

$$\{\langle(1,1), \text{M. Blanc*}\rangle, \langle(1,0), \text{being shorter than*}\rangle, \langle(1,2), \text{M. Everest*}\rangle\}$$

The following formulation can thus be used to explicate Ajdukiewicz's concept of proposition:

> The proposition expressed by a correctly constructed sentence S is a function $\alpha : X \to Y$, where X is the set of the syntactic positions of the expressions constituting S, and Y is identical with the universum.

Ajdukiewicz's theory assumes an isomorphism between the structure of a true proposition expressed by a sentence and the ordering of the fact described in this sentence:

> The assignment of syntactic positions to objects may or may not agree with the respective positions of these objects in reality. If the sentence stating a given proposition is true, then the respective positions of the objects spoken of in this sentence in reality agree with the syntactic positions assigned to these objects in the proposition stated by the sentence. In such a case, is seems natural to call the proposition stated by the true sentence a fact (Ajdukiewicz 1967/1971, p. 124).

In short, according to Ajdukiewicz, if a sentence expresses a true proposition, the order of the expressions in this sentence corresponds to the order of their designates in the world. The proposition expressed by a sentence, in turn, is a relation assigning designates to the syntactic positions of the expressions constituting this sentence. Ajdukiewicz's syntactic analysis is based on the distinction of expressions playing the role of operators and those

playing the role of arguments. Roughly speaking, every situation comprises the "protagonists" of this situation, their properties, and the relations that bind them. The distinction into operators and arguments at the level of the sentence corresponds to the shape of the situation: the designates of the argument-expressions correspond to the "protagonists" of the situation, and the properties and relations comprising the situation constitute the designates of the operator-expressions. Given this, the general truth conditions of a given proposition can be characterized as follows:

> The proposition α expressed by a sentence S is true if and only if, for each compound expression E distinguishable in S, it is the case that the objects constituting the designates of the expressions occupying argument positions in E stand in the relation[14] constituting the designate of the expression occupying an operator position in E, and these objects stand in this relation in an order corresponding to the order determined by the numbering of the syntactic positions of the argument-expressions.

For example, the proposition expressed by the sentence *Rebecca swims* is true if and only if the designate of the expression occupying the position of the argument has the property constituting the designate of the expression occupying the position of the operator – that is, if Rebecca has the property of swimming.

Thus formulated truth conditions faithfully reflect Ajdukiewicz's approach. It is also not difficult to see that they constitute a particular version of a more general formulation according to which the proposition expressed by a sentence of the form $\varphi(a_1, a_2, \ldots, a_n)$ is true if and only if the objects $a_1, a_2, \ldots, a_n$ belong to the extension of φ – the same as the formulation entailed by King's theory.

Ajdukiewicz's theory of propositions is a rare case in that it combines characteristics of both the structural and the functional theory of propositions. On the one hand, a proposition is determined largely by the structure of an appropriate sentence – this structure determines what is bound by the proposition conceived as a special kind of relation. In other words, the set identified with a proposition contains elements corresponding to the particular constituents of the sentence expressing this proposition. This aspect of Ajdukiewicz's theory clearly brings it closer to the structural approach. On the other hand, one ought not to forget that Ajdukiewicz identifies a

[14]For brevity, I assume that properties are one-argument relations.

proposition with a certain kind of function, and this is characteristic of functional theories.

If Ajdukiewicz's theory is functional, it is certainly nonstandard. According to him, a proposition is a function relating arguments in the form of the syntactic positions of expressions constituting the sentence expressing this proposition to values in the form of the designates of these expressions. In standard functional approaches, on the other hand, a proposition is a function determining the truth value of the sentence expressing this proposition for each possible world, that is, a function from possible worlds to the two-element set containing truth and falsity. To put it another way, according to Ajdukiewicz – and structural approaches in general – a proposition is constituted by what it is about, and the truth value is predicated of the proposition. In typical functional theories, in contrast, the truth value is, in a sense, a constituent of the proposition.[15] Given this, it seems that Ajdukiewicz's conception is closer to structural standpoints than it is to functional ones. Moreover, as I intend to argue, it has much in common with King's conception.

2. Comparison

2.1. Similarities

Although the respective theories by Ajdukiewicz and King were presented at different times and against fundamentally different philosophical backdrops, I am of the opinion that they are based on the same overall idea. There are two nontrivial differences between them, but they have enough in common for their comparison to be worthwhile. I think that this comparison can help bring to light certain specific issues concerning logical propositions in general and singular propositions in particular.

Let us consider the sentence *Rebecca swims*. In King's conception, the proposition expressed by this sentence is a fact consisting in the obtaining of the relation P^{16} which binds two objects (Rebecca and the property of swimming) by means of the relations that constitute it. One could say that King begins constructing his tree by determining the relation R which binds

[15] Of course, insofar as we permit that proposition-functions be considered as ordered pairs whose elements are possible worlds and truth values.

[16] For simplicity's sake I temporarily ignore the instruction I featuring in King's model. The instruction is a part of the proposition in this model but not a part of the relation P. Since I is encoded by R, which is a part of P, this simplification seems to be acceptable.

the appropriate linguistic expressions and to which the relevant semantic relations are added during subsequent analysis so that the relation P can emerge. Let us see how an analogous procedure of a model for a proposition (and thus, of determining the structure of this proposition) looks like in the case of Ajdukiewicz's conception.

Ajdukiewicz's syntactic analysis of the sentence *Rebecca swims* looks like this:

$$\begin{array}{cc} \underline{\textit{Rebecca}} & \underline{\textit{swims}} \\ (1,1) & (1,0) \end{array}$$

As a result of this step, the syntactic relations obtaining between the expressions constituting this sentence – captured using the relation R in King's model – have been determined. We can thus move onto the next part of King's tree, namely, the part where the appropriate semantic relations are represented (i.e. the relations of reference obtaining between the name "Rebecca" and Rebecca and the expression "swims" and swimming).

According to King's approach, the expression entering the given semantic relation is identified by reference to its syntactic characteristics, that it, by determining its location on one of the branches of the tree. For example, it is indicated that Rebecca is the designate of the expression located on the left branch of Tree 1, that is, of the name "Rebecca". An analogous step can be found in Ajdukiewicz's analysis. Here, the proposition is taken to be the function relating each syntactic position distinguished in the sentence expressing it to the object constituting the designate of the expression this syntactic position. To determine the syntactic position of an expression is thus, no more no less, to identify this expression by reference to its syntactic characteristics (it is impossible for two expressions to occupy the same syntactic position). For instance, Rebecca is assigned to position (1,1) in our example because this position is occupied by the name "Rebecca" referring to Rebecca.

In King's conception, the compounding of the syntactic relation and the semantic relations – that is, the compounding of the relation R and the relation of reference – gives in effect the relation P which can be thought of as the structure of the proposition. In Ajdukiewicz's conception, the compounding of the analogous relations – that is, the assignment of syntactic positions to the expressions featured in the sentence and the subsequent assignment of appropriate designates to these positions – determines a

function identical with the set A:

$$A = \{\langle(1,1), \text{Rebecca*}\rangle, \langle(1,0), \text{swimming*}\rangle\}.^{17}$$

The set A thus plays the same role in Ajdukiewicz's conception as does Tree 2 in King's theory. Namely, both represent the relation obtaining between the syntactic positions distinguishable in a sentence and the designates of the expressions occupying these positions. This ordering of the designated objects by reference to the syntactic characteristics of the expressions designating them is at the core of a proposition, according to both conceptions.[18][19]

2.2 (Apparent) differences

Regarding the constitution of a proposition, the two conceptions differ in a twofold manner. The first difference is that Ajdukiewicz's set A is slightly poorer in information than Tree 2, its analogue in King's conception. Ajdukiewicz's model does not account for three elements considered by King: context, instruction, and the relation of joint instantiation.

As far as the relation of joint instantiation is concerned, is seems legitimate to claim that it is inscribed into A. The set is determined in such a way that it is clear that the term "Rebecca" refers to Rebecca and that it constitutes the first argument of the operator in the form of the expression referring to the property of swimming—precisely these two properties of the term "Rebecca" are captured in King's model as the relation of joint instantiation.[20]

[17] It might be worth noting that both in King's model and in Ajdukiewicz's conception the last stage of the analysis of the nature of the proposition (Tree 2 in King and Set A in Ajdukiewicz) does not feature expressions themselves. The transition from Tree 1/syntactic analysis to Tree 2/Set A consists, among other things, in the removal from the model of the names of the expressions making up the analyzed sentence and limiting the model to the syntactic characterization of the expressions on the one hand, and to their designates, on the other.

[18] This clearly differentiates the two positions from many other versions of the structural approach to propositions (e.g. those of Soams or Salmon) which assume that the structure of the proposition is somehow correlated with the structure of the sentence but do not incorporate this assumption into their actual models of propositions.

[19] It might be worth noting that the postulate to reflect the structure of the sentence in the structure of the proposition is dictated mainly by the desire to avoid the problem of an imprecise identification of propositions faced by functional theories.

[20] As mentioned earlier, joint instantiation is responsible for the compounding of the syntactic relation and the appropriate semantic relations. What is in a sense analogous to this in Ajdukiewicz's model is the apprehension of a given syntactic position and an appropriate designate as an ordered pair.

The question of context and instruction is more difficult. The instruction contains information concerning the general truth conditions of a proposition and is encoded by the syntactic relation R constituting a part of the propositional relation. Bluntly speaking, the relation R determines the way in which the designates of the expressions bound by R must be connected in order for the analyzed proposition to be true. In the case of the sentence *Rebecca swims*, its syntax determines the fact that the proposition expressed by this sentence is true if Rebecca has the property of swimming. Ajdukiewicz's method has certain advantages over King's model because here the relationship determining the truth conditions of a proposition is contained in the very syntactic analysis of the sentence expressing it, based on the distinction into arguments and operators. A proposition is true if the designate of the expression constituting the argument, or the designates of the expressions-arguments, satisfies the condition expressed by the term or phrase functioning as the operator. In order for the proposition about Rebecca to be true, Rebecca – as the designate of the expression-argument – must instantiate the property expressed by the operator "swim". This kind of relationship between the designates of the expressions making up a sentence is thus taken into account already at the level of syntactic analysis. Therefore, in Ajdukiewicz's approach there is no need to "glue" an extra instruction onto the propositional relation (this step is required in King's approach). On the other hand, each analysis carried out using Ajdukiewicz's method encodes the same kind of instruction – a proposition is true if the designates of the expressions-arguments satisfy or fall under what is expressed by the operator. It is thus impossible to encode an instruction imposing that the proposition expressed by the sentence *Rebecca swims* is true if and only if Rebecca does not instantiate the property of swimming. Owing to the fact that King treats the instruction I as encoded by the relation R, but also external and autonomous relative to it, he can successfully represent different instructions governing the truth conditions of a proposition.[21]

[21] It is legitimate to ask at this point if a theory of propositions must in fact account for the possibility to encode different instructions in a proposition. There are two sub issues here. On the one hand, there is the empirical question of whether there exists a (natural) language in which the proposition [that Rebecca swims] is true if Rebecca does not have the property of swimming. On the other hand, one can doubt, on theoretical grounds, if such a language is at all possible. Its users would certainly possess a different notion of truth from ours. One could say that, for them, the predicate "true" is a synonym of our predicate "false". This is not the right place to offer a detailed discussion of the concept of truth but, in light of the above, it is justified to claim, in my opinion, that King's concept of instruction is at best vague.

As far as accounting for dependence on context (sensitivity of the reference of given expressions to context) in a proposition is concerned, there is, in short, no room for it in Ajdukiewicz's model. However, there is no reason why appropriate contextual parameters could not be introduced into the description of the proposition – in such a case, one would assume that a given sentence expresses this or that proposition in this or that context. There is also no doubt that context must play a role in determining the designates entering into ordered pairs involving expressions sensitive to context, especially indexical expressions. There is a fundamental difference between this kind of approach and King's conception: King takes relativization to context to be a constituent of the proposition, not a part of its description. King (2007, p. 39) is convinced that not accounting for context in a proposition must yield a theory that does not permit propositions expressed by sentences containing expressions sensitive to context. I must leave the highly complex question of whether context is best seen as external to the proposition (as in Ajdukiewicz) or as an integral constituent thereof (as in King) open. Regardless of the solution, this constitutes a clear difference between the two theories.

Another difference regards, to put it in general terms, the ontological status of propositions. King is very clear that in his theory's propositions are not identified with any kind of formal constructs, and thus, in particular, that they are not identified with functions. A proposition, according to King, is a special kind of fact: the fact that a certain relation obtains between certain objects. Importantly, the proposition is not identical with this relation. For example, the propositional relation P illustrated in Tree 2, linking Rebecca and the property of swimming, is not identical with the fact-proposition [that Rebecca swims]. As has been shown, the (rough) equivalent of the propositional relation in Ajdukiewicz's theory is a function identical with the set A whose elements are ordered pairs containing specific kinds of objects. However, Ajdukiewicz does not claim that a proposition should be identified with the fact of the occurrence of this function or the fact of the existence of the set A; in his conception, the proposition expressed by the sentence *Rebecca swims* is that function, and thus, is the set A. In short, according to Ajdukiewicz, a proposition is a relation, and according to King, it is the fact that a certain relation obtains. That said, it is worth considering if this difference is in fact as fundamental as it might seem at first glance rather than being purely verbal.

Ajdukiewicz was certainly not a flippant philosopher. If he said that a proposition-function can be identified with a set of appropriate ordered pairs,

then it should be concluded that that is what he meant, not merely that the function can be so represented. If we stick to the letter of Ajdukiewicz's argumentation, then, the difference between his theory and King's might indeed be fundamental.

That said, I am of the opinion that another reading of Ajdukiewicz's conception – one close enough to his intention – is possible. It is based on an alternative interpretation of the concept of function. Ajdukiewicz assumed a commonly accepted and frequently employed set-theoretical interpretation of this concept according to which a relation, in particular a function, is a set of ordered pairs of elements constituting the arguments of this relation. However, if we treat a function in a less "logical" and more "ontological" manner, we can characterize it as a *sui generis* mechanism, process, and even fact – the fact that an assignment occurred where objects belonging to a certain set are related to objects belonging to another set. In light of this, Ajdukiewicz's proposition can thus be identified with the occurrence of an assignment of certain objects to appropriate syntactic positions.[22] If we assume that the interpretation of Ajdukiewicz's conception according to which every proposition is a proposition about (among other things) certain syntactic positions is not correct (see below), and if we accept the aforecited alternative understanding of the concept of function, we get an approach according to which a proposition is identical with the fact of the occurrence of a certain relation between objects constituting the designates of the expressions making up the sentence expressing this proposition. This relation, on the other hand, is the compound of the syntactic relation (captured via syntactic positions) binding expressions making up the given sentence and the semantic relations obtaining between the particular expressions making up this sentence and their designates. It is not difficult to see that this summary of Ajdukiewicz's standpoint overlaps with the characterization of propositions offered by King. It is doubtful that King is familiar with Ajdukiewicz's conception, but if it were the case, one could convincingly argue that his theory is a development of Ajdukiewicz's conception, as interpreted here.[23]

[22]This interpretation is supported by the expression used by Ajdukiewicz to characterize a proposition. Namely, he writes that a proposition is a function establishing the assignment of syntactic positions to designates (cf. Ajdukiewicz 1967/1971, pp. 123, 124).

[23]One should remember that the difference concerning the constituents of a proposition indicated by King but missing from Ajdukiewicz's model is still very much there, even on the alternative understanding of the concept of function outlined above.

Another difference between the discussed standpoints concerns the content of a proposition and is related to the problem of its constitution. In Ajdukiewicz's approach, propositions are constituted by individual objects and properties/relations (making up the set of values of the function identified with a given proposition) and syntactic positions (making up the domain of the function-proposition) – the latter determine the order of the terms in the proposition. Things are different in King's theory, where syntactic properties belonging to a proposition enter into it indirectly – as constituents of the propositional relation binding the appropriate constituents of the proposition, that is, the appropriate objects and properties/relations. It is worth considering if this difference is not apparent; it could be assumed that syntactic positions play a similar role in Ajdukiewicz's theory as tree branches do in King's – that is, they represent the order of the objects in a proposition. It seems that both in Ajdukiewicz's and in King's proposal, a proposition is nothing other than the objects this proposition, taken a certain way, is about (Ajdukiewicz 1967/1971, p. 124). On the other hand, it is unlikely that either Ajdukiewicz or King would be inclined to claim that propositions are about syntactic positions or tree branches (or brackets used to reflect the structure of the proposition). This assumption seems legitimate in light of the sameness of the truth conditions generated by both conceptions, as indicated above; it is also perfectly acceptable given the interpretation of Ajdukiewicz's function-proposition as the occurrence of a certain ordering of designates.

The difference between the two interpretations of Ajdukiewicz's conception – referring to two different approaches to the concept of function – can appear to be unimportant, even purely terminological. The ease with which the alternative approach has allowed us to nearly equate Ajdukiewicz's and King's conceptions might arouse skepticism. According to King, one of the most significant advantages of his conception is that it draws an equivalence between propositions and facts, as opposed to logical constructs. As it turns out, the only step that had to be made in the case of Ajdukiewicz's conception to move from an identification of a proposition with a formal construct to its identification with a fact was the assumption of an alternative understanding of the concept of function – incidentally, this understanding does not seem to be at odds with the more traditional set-theoretical approach. In light of this, it is an open question if identifying propositions with facts is in fact as significant as King makes it out to be.

3. The Benacerraf Problem

King's unwillingness to identify propositions with any kind of formal construct stems from his conviction that all conceptions that make such an identification inadvertently fall into the sort of trouble indicated by Paul Benacerraf (1965). The problem arises when one model permits two (or more) equally adequate yet mutually contradictory representations of a given phenomenon. In some theories of propositions,[24] the problem is that the ordered n-tuple identified with a given proposition is determined unequivocally. One can thus conclude that these standpoints face the Benacerraf Problem. In the case of Ajdukiewicz's theory, no oversight of this gravity can be noted. His method of syntactic analysis is sufficiently determined and based on the fundamental distinction into expressions-operators and expressions-arguments, thus yielding at least apparently unequivocal results.

However, it is not difficult to notice that, using Ajdukiewicz's method, two different and at the same time intuitively equally good syntactic description of the same sentence can be given. For example:

$$\underset{(1,1)}{\underline{\textit{Ajdukiewicz}}}\ \underset{(1,0)}{\underline{\textit{was}}}\ \underset{(1,2)}{\underline{\textit{the son-in-law of Twardowski}}}$$

$$\underset{(1,1)}{\underline{\textit{Ajdukiewicz}}}\ \underset{(1,0)}{\underline{\textit{was the son-in-law}}}\ \underset{(1,2)}{\underline{\textit{of Twardowski}}}$$

Both analyses of the sentence are correct, and it would be difficult to argue for the superiority of one over the other on purely syntactic grounds. It would thus appear that we are facing a typical example of the Benacerraf Problem here. This is not the case, however, since the analyses in question do not constitute two competing representations of the same. The syntactic ambiguity they reveal matches a semantic ambiguity in this case—according to analysis (i), the discussed sentence states Ajdukiewicz's membership in

[24] This pertains, for instance, to the twin conceptions (to my knowledge later abandoned by their authors) proposed by Salmon (1986) and Soames (2009). In these conceptions, the proposition expressed by a sentence is represented using appropriate ordered n-tuples, but it is not at all clear according to what rules these n-tuples are to be ordered. For example, the proposition expressed by the sentence *Desdemona loves Cassio* can be represented by any of the following formulas (where L refers to loving): (1) $\langle$Desdemona*, L*, Cassio*$\rangle$; (2) $\langle$Cassio*, L*, Desdemona*$\rangle$; (3) $\langle$L*, Desdemona*, Cassio*$\rangle$; and (4) $\langle$L*, Cassio*, Desdemona*$\rangle$. The problem here is not only the multiplicity of possibilities but also the fact the same four n-tuples can be considered to represent the proposition expressed by the sentence *Cassio loves Desdemona*.

the class of Twardowski's sons-in-law, and according to analysis (ii), it states that the relation of being a son-in-law obtains between two individuals: Ajdukiewicz and Twardowski.[25] The possibility of giving two different syntactic descriptions of this sentence does not indicate that Ajdukiewicz's method of syntactic analysis yields ambiguous results but that the sentence can express two different propositions:

(i) $\{\langle(1,1), \text{Ajdukiewicz}^*\rangle, \langle(1,0), \text{belonging to}^*\rangle, \langle(1,2), \text{being Twardowski's son-in-law}^*\rangle\}$[26]

(ii) $\{\langle(1,1), \text{Ajdukiewicz}^*\rangle, \langle(1,0), \text{being a son-in-law of}^*\rangle, \quad \langle(1,2), \text{Twardowski}^*\rangle\}$

It is not the case that in instances of this kind the method of syntactic analysis turns out imprecise and the standpoint faces the Benacerraf Problem. On the contrary, the method allows us to uncover a phenomenon consisting in two token sentences of the same type expressing two different propositions already at the level of syntactic properties. However, there is a certain special group of sentences whose analysis within Ajdukiewicz's theory does lead to the Benacerraf Problem. They are sentences whose main operator is an expression constituting a two-argument functor referring to some symmetrical relation (regardless of the kind of arguments is might take). Typical examples of such sentences are sentences of the form A=B, or A≠B, in which the main operator is an expression constituting a two-argument functor referring to the symmetrical relation of identity or nonidentity, respectively. The analysis of sentences of this kind can be carried out in a twofold manner: such that the first argument of the operator is A and such that it is B. As a result, we get two alternative approaches to the same proposition – and there are no criteria for the selection of one approach over the other as more adequate.

It seems that the only way to avoid the Benacerraf Problem here is to introduce a conventional rule to the effect that the first appropriate expression occurring in a sentence is to be treated as the first argument of the operator. For example, in a false sentence of the form A=B, the syntactic position (1,1) will be occupied by A. This solution cannot be considered

[25] As noted by Tałasiewicz (2003, p. 153), the indicated ambiguity is not a frequent occurrence in the vernacular. However, the distinction into membership in a class and being one of two arguments of a relation turns out significant, for example, in ontological discourse.

[26] The object constituting the designate of the expression located at (1,2) here is the extension of the predicate "being Twardowski's son-in-law", that is, a certain set.

satisfactory since it is *ad hoc* – it is difficult to point to an independent rationale for this rule other than the need to avoid the Benacerraf Problem. Moreover, the introduction of such a convention leads to an undesirable outcome: in a sentence of the form B=A, position (1,1) is occupied by B, and thus, the propositions expressed by A=B and B=A must be represented, respectively, as:

- $\{\langle(1,1), A^*\rangle, \langle(1,0), =^*\rangle, \langle(1,2), B^*\rangle\}$
- $\{\langle(1,1), B^*\rangle, \langle(1,0), =^*\rangle, \langle(1,2), A^*\rangle\}$

These representations are different, and since it would seem that, in a pair of two syntactically different sentences about the same symmetrical two-argument relation, both sentences express the same proposition, we do face the problem of "surplus" representation indicated by Benacerraf. Does this mean that King is right that standpoints identifying propositions with formal constructs ought to be rejected because of the Benacerraf Problem? The answer is yes and no.

King is wrong when he assumes that moving propositions from the domain of formal constructs to the sphere of facts will eliminate the spectre of the Benacerraf Problem. The difficulty related to sentences about symmetrical relations outlined above occurs both in the interpretation of Ajdukiewicz's theory according to which a proposition is identified with a function and in the reading according to which a proposition is identical with the fact of the occurrence of a certain assignment (function). In the first case it is not clear which of two functions should be identified with the proposition expressed by a sentence of the form A=B; in the latter case, the fact of the occurrence of which function should be considered as identical with the proposition.

As can easily be seen, King's theory faces an analogous problem. In his conception, the fact identified with a proposition is that certain objects remain in this or that relation – the relevant relation being determined through appropriate formal constructs. In order to report the proposition-fact expressed by a (true) sentence of the form A=B, we must not only identify objects A and B but also describe the propositional relation obtaining between them. The key fragment of this relation is the relation R reflecting the syntactic properties of the sentence. We therefore get, analogously to Ajdukiewicz's conception, two alternative approaches to the relation R, and thus, two propositions-facts. One consists in the occurrence of an assignment characterizable as $\{K, I, [[\text{A}^*][=^*[\text{B}^*]]]\}$; the other, in the occurrence of

$\{K, I, [[\text{B}^*][=^*[\text{A}^*]]]\}$. In short, it turns out that one can speak of two facts here, each of which can be identified with the proposition expressed by A=B.

King attempts to find a way out of this trap by means of a rather surprising proposition. Namely, he accepts the validity of the claim (deemed problematic earlier in this paper) that a sentence of the form A=B expresses a different proposition than the sentence of the form B=A. As part of his justification of this thesis, King (2007, p. 95) analyzes the pair of sentences *2=1* and *1=2*. The fact that King considers these sentences to express different propositions follows from his stance on propositions. In his theory, the aforecited sentences express propositions structured like this:

- $\{K, I, [[2^*][= ^*[1^*]]]\}$
- $\{K, I, [[1^*][= ^*[2^*]]]\}$

These propositions are made up of the same constituents but they differ in structure, that is, their constituents are ordered differently, thus constituting different propositions.

This approach raises doubt since it seems that by uttering the sentence *2=1* one conveys the same information as one does by uttering *1=2*. Analogously, it is difficult to claim that different content is expressed by uttering *Alec likes tomato soup or Alec likes spinach* and *Alec likes spinach or Alec likes tomato soup* etc. One could thus say that the fact that in King's conception propositions expressed by such pairs of sentences are considered to be different is a defect. To put it another way, one could challenge King by pointing out that his theory is too fine-grained when it comes to the identification of propositions.

Naturally, King is aware of this and does respond to the challenge. According to him, the conviction that the propositions expressed by *1=2* and *2=1* are identical is due to the following principle (P):

> (P) Sentences p and q express different propositions if there is a context C such that O_p and O_q have different truth values in C, O being a non-transparent sentential operator (King 2007, p. 96).

A non-transparent sentential operator is a one-argument sentential functor establishing a non-transparent context, that is, one where – to put it simply – the truth value of a compound sentence O_p is not a function of the truth value of the subordinate clause p falling within the scope

of this operator. Examples of such operators include expressions such as: "necessarily", "John claims that", "should be a fact" etc. Since there is no context in which the sentences John claims that 2=1 and *John claims that 1=2* would express propositions characterized by different truth values, it is standardly assumed that *2=1* and *1=2* express the same proposition. King (2007, p. 97) holds that the aforecited principle for distinguishing between propositions is incorrect for the following reason:

> We can think of the propositions expressed by sentences containing a connective in this set as consisting of a proposition (expressed by the embedded sentence) and a property of propositions (expressed by the connective). This "complex proposition"' is true at a circumstance iff the constituent proposition possesses the property in question at the circumstance. From this perspective, the claim that two sentences express the same proposition if the results of embedding them with respect to all propositional connectives have the same truth values in all circumstances (and similarly for all syntactically similar pairs), essentially amounts to the claim that propositions that possess all properties of propositions expressed by English (or natural language) propositional connectives in common at all circumstances of evaluation (and similarly for all syntactically similar pairs) are identical (King 2007, pp. 97–98).

According to King, there is no basis for accepting this thesis. He rejects as incorrect the assumption that natural-language sentential operators express all properties that a proposition can possibly have. Therefore, even if two propositions have the same set of all properties expressed by such operators, the most one can say is that these propositions have a lot in common, not that they are identical. It seems that King's argumentation can be summarized in the following manner: the notion of the identity of propositions entailed by the principle P is a flawed one since it reduces their identity to the identity of the set of expressible properties (expressible by means of sentential operators). In other words, identity, as it is construed in P, is a contingent property of propositions, and it should be a necessary one.

In light of the above, it would be an error to conclude that the fact that in King's conception propositions expressed by pairs of sentences about identity are different stems from syntactic differences between these sentences (that is, the fact that expressions located in the position of the first and the second argument are switched). Rather, one should say that the structure

of the relevant propositions is not identical, according to King, and this is reflected in the syntax of the appropriate sentences.

However, there are at least two weak points in King's argumentation.[27] First, it is legitimate to ask what – if not syntactic analysis – provides grounds for the claim that the propositions expressed by *2=1* and *1=2* are different. If we had at our disposal some language-independent means of glimpsing into the structure of a proposition, King's standpoint would be justified. But since this is beyond our capacity, the only thing we have are our linguistic intuitions, and these – or so it seems – incline us toward the claim that the sentences in question express the same proposition. In short, even if we agree with King that there is no good reason to consider the two sentences in question as expressing the same proposition, King does not offer any solid grounds for claiming that they express different propositions.

Secondly, one might say that King is tendentious in his choice of examples since the sentences *2=1* and *1=2* are false (or, alternatively, say something about two different objects). Once true sentences about identity are considered, King is bound to face the problem of doubling the same object in one proposition. For example, the structure of the propositions expressed by:

(CC) Cicero=Cicero

(CT) Cicero=Tullius

is the following according to King's theory:[28]

- S(CC) = $\{K, I, [[\textcircled{\scriptsize}][= *[\textcircled{\scriptsize}]]]\}$
- S(CT) = $\{K, I, [[\textcircled{\scriptsize}][= *[\textcircled{\scriptsize}]]]\}$

[27] The weak "initial" point, noted by an anonymous referee, as mentioned earlier, can be added to this list (I am, again, most obliged for this comment). Namely, the principle (P) is construed as a criterion for stating that two propositions are different, but King interprets it as stronger than it actually is (one might say that he reads an implication as if it were an equivalence) and assumes that if two propositions do not satisfy this criterion, then they are to be considered identical. This misuse shows that his diagnose as to the reason why it is standardly thought that *1=2* and *2=1* express the same proposition (and not two different propositions) is misplaced – the reason for this lies elsewhere. This in fact invalidates his entire argumentation.

[28] In order to avoid confusion stemming from using names in the description of propositional structure, I use a graphic representation of Cicero – after all, according to the discussed conception, Cicero himself enters into the proposition.

In order to avoid the claim that these propositions are identical, King would have to conclude that Cicero featuring in the structure of the proposition P(CT) on the left is given in some other manner than Cicero situated in the structure of this proposition on the right. However it is clear that to consider the manner in which an object is given as a part of the proposition is to abandon the conception of direct reference and singular propositions. And this is seriously at odds with the assumptions of King's theory.

***

The above discussion can be briefly summarized as follows: (i) the two theories explored here are based on the same idea that the proposition expressed by a sentence is a fact that the designates of the expressions making up this sentence remain in an order corresponding to the syntactic structure of the sentence; (ii) differences between the discussed theories turn out to be apparent once an alternative (non-set-theoretical) understanding of the concept of function (more specifically, the function identified with a proposition according to Ajdukiewicz) is introduced; (iii) although it seems at first glance that King's theory is better equipped to cope with the Benacerraf Problem, his argumentation is ultimately unconvincing, prompting the conclusion that neither of the analyzed theories is immune to this problem.

Bibliography

Ajdukiewicz, K. (1960/1985). Związki składniowe między członami zdań oznajmujących. In *Język i poznanie*, Vol. II (pp. 344–355), Warszawa: PWN.

Ajdukiewicz, K. (1967/1971). Sąd jako konotacja zdania. In J. Pelc (Ed.), *Semiotyka polska 1894–1969* (pp. 112–127), Warszawa: PWN.

Benacceraf, P. (1965). What Numbers Could Not Be. *The Philosophical Review*, *74*, 47–73.

Carnap, R. (1947/2007). Znaczenie i konieczność. In *Pisma semantyczne* (T. Ciecierski, M. Poręba, M. Sala, B. Stanosz, Trans.; pp. 201–465). Warszawa: Aletheia.

Ciecierski, T. (2003). O pojęciu sądu logicznego. *Przegląd Filozoficzny*, *48*, 125–144.

Ciecierski, T. (2012). A Problem with Structured Propositions. In P. Stalmaszczyk (Ed.), *Philosophical and Formal Approaches to Linguistic Analysis* (pp. 81–86). Berlin, Boston: De Gruyter.

Deutsch, H. (2008). Review of The Nature and Structure of Content. *Notre Dame Philosophical Reviews. An Electronic Journal*. 24.05.2008, URL: `http://ndpr.nd.edu/news/23524-the-nature-and-structure-of-content/`

King, J. (2007). *The Nature and Structure of Content*. Oxford: Oxford Univeristy Press.

King, J. (2014). Naturalized Propositions. In J. King, S. Soames, J. Speaks, *New Thinking about Propositions* (pp. 47–70). Oxford: Oxford University Press.

Kirkham, R. (2001). *Theories of Truth. A Critical Introduction*. Cambridge, Mass.: MIT Press.

Loux, M. (2003). *Metaphysics. A Contemporary Introduction*. London, New York: Routledge.

Lycan, W. (2002). *Philosophy of Language. A Contemporary Introduction*. London, New York: Routledge.

Makin, G. (2000). *The Metaphysicians of Meaning. Russell and Frege on Sense and Denotation*. London, New York: Routledge.

Odrowąż-Sypniewska, J. (2006). *Rodzaje naturalne. Rozważania z filozofii języka*. Warszawa: Semper.

Russell, B. (1903/2008). *The Principles of Mathematics*. London, New York: Merchant Books.

Salmon, N. (1986). *Frege's Puzzle*. Cambridge, Mass.: MIT Press.

Soames, S. (2009). Direct Reference, Propositional Attitudes, and Semantic Content. In *Philosophical Essays*, Vol. II (pp. 33–71). Princeton: Princeton University Press.

Tałasiewicz, M. (2003). Rodzaje niejednoznaczności opisu składniowego wyrażeń. *Przegląd Filozoficzny*, *12*, 145–154.

Originally published as "Dwa modele struktury sądu" *Studia Semiotyczne* 30/2 (2016), 99–127. Translated by Joanna Guzowska.

STUDIA SEMIOTYCZNE—ENGLISH SUPPLEMENT, vol. XXIX (2017), 107–126
ISSN 2451-2745
DOI: 10.26333/stsen.xxix.06

Bartosz Wcisło[1], **Mateusz Łełyk**[2]

Strong and Weak Truth Principles[3]

Abstract This paper is an exposition of some recent results concerning various notions of strength and weakness of the concept of truth, both published or not. We try to systematically present these notions and their relationship to the current research on truth. We discuss the concept of the Tarski boundary between weak and strong theories of truth and we give an overview of non-conservativity results for the extensions of the basic compositional truth theory. Additionally, we present a natural strong theory of truth which admits a number of apparently unrelated axiomatisations. Finally, we discuss other possible explications of the notion of 'strength' of axiomatic theories of truth.

Keywords Axiomatic truth theories, Peano arithmetic, Conservativity, Tarski boundary

1. Introduction

1.1 Axiomatic theories of truth

Formal theories of truth are a part of philosophy investigating the notion of truth with the methods of mathematical logic. One of the main methods of formalisation is to consider *axiomatic theories of truth* which are constructed in the following way:

[1]University of Warsaw, Faculty of Mathematics, Informatics and Mechanics.
E-mail: b.wcislo@mimuw.edu.pl
[2]University of Warsaw, Faculty of Philosophy and History.
E-mail:mlelyk@student.uw.edu.pl
[3]Research reported in this paper was conducted with a financial support of the Polish National Science Centre, Grant number 2014/13/B/HS1/02892.

- We fix a **base theory** B modelling the totality of our knowledge of extra-semantic facts (facts not concerning such notions as meaning or truth).

- We add to the language of that theory a new unary predicate $T(x)$ with the intended reading "x is a true sentence" and we extend B with axioms governing the new predicate.

We then investigate, how the properties of the obtained theory depend on the choice of axioms governing the truth predicate.

The base theory B is often chosen to be Peano arithmetic PA. The motivation behind this choice is that the vast majority of results concerning the relationship between truth theory and its corresponding base theory do not significantly depend on the specific choice of the latter. The only thing which we do require is that it is capable of expressing and proving basic facts concerning syntax, like: "every sentence which is built correctly has the same number of right and left parentheses." PA is more than enough to this end.

It is worth stressing that most logicians do not think of PA as a theory of numbers but rather as a more general theory of finite mathematical objects like hereditarily finite sets, finite graphs, finite strings of characters over a finite alphabet. This theory suffices to prove surprisingly many facts concerning these kinds of objects[4]. Since sentences, formulae or proofs in formal languages may also be treated as finite mathematical objects, namely strings of characters with some simple structural properties, Peano arithmetic allows us to freely speak about them. Having said that, we have to admit that the choice of PA as a base theory is somewhat arbitrary. Formulating in an abstract way the conditions guaranteeing that a base theory is strong enough from the point of view of truth theory, and which suffice to prove the results which make PA our choice, seems a rather daunting task. At this initial stage of research, we prefer the "bottom-up" strategy.

In this paper, we focus on theories describing the truth predicate for the language of the base theory. However, let us stress that the properties of self-referential truth predicates which formalise the notion of truth for all sentences of the language to which they belong are a subject of extensive studies (a good account of results concerning such theories may be found in (Halbach, 2011)).

[4]There are many sources concerning formalisation of syntax and making above comments precise. We especially recommend (Franzen, 2003).

The role of the axioms governing the truth predicate is, obviously, to capture various intuitions concerning this notion. In the studies on formal truth theories, we are trying to explain what are the relations between those intuitive properties, and what are the consequences of the fact that the predicate enjoys these features. One of the simplest conditions for the truth predicate which we may consider is as follows:

$$\varphi \equiv T(\varphi)$$

for all sentences in the language of the base theory. These axioms say that the truth predicate satisfies Tarski's biconditionals for the language of the base theory. The theory extending PA in which the only axioms governing the truth predicate are these biconditionals is called TB^{-}[5].

Another property, which should be satisfied by the truth predicate is compositionality. We express it with axioms formalising principles such as:

> (For all sentences φ and ϑ) The conjunction of sentences φ and ϑ from the language of the base theory is true if and only if both of the conjuncts are.

Or:

> (For an arbitrary variable v and an arbitrary formula $\varphi(v)$ with at most one free variable v) The universal sentence $\forall\varphi(v)$ in the language of the base theory is true if and only if for any numeral $\underline{x}$, the sentence $\varphi(\underline{x})$ is true.
>
> Let us add that by a numeral $\underline{x}$, we mean the canonical term denoting the number x, for instance $(\ldots((\overline{0}+\overline{1})+\overline{1})\ldots+\overline{1})$, where the addition symbol occurs x times (and where $\overline{0}$ and $\overline{1}$ are some fixed symbols representing 0 and 1, respectively).

Let us observe that already the theory TB^{-} can prove for any two concrete sentences from the language of arithmetic that their conjunction[6] is true if and only if both are true. However, it cannot prove the general fact about all arithmetical sentences, which is expressed by the first of the quoted axioms. The theory whose axioms say that the truth predicate is

[5]The notation $\mathsf{TB}\restriction$ is used more often in the literature.

[6]Obviously, we mean here the Gödel code representing the conjunction of these formulae. For the sake of clarity, we will use slightly imprecise expressions.

compositional is called CT^-[7]. The precise definition of this theory may be found in (Halbach, 2011). The other principle, which can be postulated, is the **extentionality principle** for the truth predicate:

> For any sentences $\varphi(t)$, $\varphi(s)$ from the language of the base theory, if the values of the terms t,s are equal, then the sentence $\varphi(t)$, is true if and only if the sentence $\varphi(s)$ is true.

Another possible requirement is that the sentences containing the truth predicate satisfy induction or, equivalently, the least number principle:

> Every nonempty subset of natural numbers defined with a formula containing the truth predicate has the least element.

In the language of first-order logic (in which all the theories considered here are formulated), the above principle can be expressed with an infinite system of axioms, the so called induction scheme for the formulae of the language extended with the truth predicate. The above principle has a more technical character than the ones which we have previously described. However, we can interpret it as follows: the properties defined using arithmetical predicates are "well defined" in the sense of not being vague.

Let us add that the theories CT^- and TB^- with the induction scheme for the sentences containing the truth predicate are called CT and TB, respectively. We hope that the Reader sees that there is a vast array of natural properties which the truth predicate should satisfy. There are even more possibilities, when we consider the self-referential truth predicate, that is, if we try to account for the behavior of the truth predicate applied to sentences in which that very predicate occurs.

1.2 Weak and strong truth theories

It is one of the very basic facts in the theory of truth that the theory CT proves certain arithmetical sentences which are not provable in PA alone. Namely, by Gödel's Second Theorem we know that if PA is consistent, then it does not prove the sentence $\mathsf{Con}_{\mathsf{PA}}$ which formalises the consistency claim for PA. However, the following fact holds:

Theorem 1 (Tarski)**.** CT *proves the sentence* $\mathsf{Con}_{\mathsf{PA}}$.

[7]Again, the theory in question is more often called $\mathsf{CT}\uparrow$. More generally, the theories which we denote Th^- are typically called $\mathsf{Th}\uparrow$.

Let us present an informal sketch of the proof of this theorem (a full proof may be found, for instance, in (Łełyk & Wcisło, 2017a)): We first show that CT proves the statement "All axioms of PA are true." Since PA has infinitely many axioms, this is not quite trivial. It is not enough to prove that every axiom separately is true (which can be done already in TB^-). We need to show the general statement. The intuition behind the proof is not terribly complicated, but it does contain some technical details, so we will only sketch it. Working in CT^-, let us fix any formula $\varphi(x)$. The sentence

$$T(\varphi(\overline{0})) \wedge \forall x(T(\varphi(\overline{x})) \rightarrow T(\varphi(\overline{x+1}))) \rightarrow \forall x T(\varphi(\overline{x}))$$

is an (actual) instance of the induction scheme (with parameter φ) for a certain formula with the predicate T, therefore it is available in CT as an axiom. Using the compositional axioms of CT^-, we obtain

$$T(\varphi(\overline{0}) \wedge \forall x(\varphi(x) \rightarrow \varphi(x+1)) \rightarrow \forall x \varphi(x))$$

and the above sentence states that the instance of the induction scheme for the formula φ is true. Since φ was arbitrary, we obtain the general sentence[8]. PA has only finitely many axioms except for the induction scheme, so by finitely many applications of compositionality of the truth predicate, we can show that all of them are true.

Having proved that the axioms of PA are true, we show by induction on the number of steps in a proof that any sentence which is derivable from the axioms of PA is true. At the same time, we can show that no sentence of the form $\varphi \wedge \neg\varphi$ is true. Therefore, no sentence of this shape is provable in PA which ends the sketch of the proof of Theorem 1.

The above theorem may be viewed philosophically important. It turns out that adjoining to PA a truth predicate satisfying very natural conditions yields a theory stronger than PA. This fact has been employed in a well-known argument against the deflationary theory of truth[9]. When Th_1, Th_2 are two theories such that $\mathsf{Th}_1 \subseteq \mathsf{Th}_2$ and there exists a sentence in the language of the theory Th_1 which is provable in Th_2, but not in Th_1, we say that Th_2 is **non-conservative** over Th_1. We say that Th_2 is conservative over Th_1 otherwise. By Theorem 1 (and Gödel's Theorem) it follows that CT is non-conservative over PA. Regardless of the philosophical importance of this

[8]We are omitting certain details here. For instance, the described argument only shows the truth of the parameter-free induction scheme. As we have said, the full proof requires us to deal with some technical issues which are not conceptually demanding.

[9]See (Ketland, 1999), (Shapiro, 1998).

specific fact, the following general question seems interesting: what properties of the truth predicate make the truth theory Th non-conservative over its base theory B? In the following paper, we describe certain results concerning this question. In other words, we try to understand what properties of the notion of truth make the truth theory "stronger" than its base theory.

2. Known results concerning conservativity

In light of Tarski's result discussed in the previous section that a compositional truth theory with full induction scheme for the whole language is not conservative over PA, it is natural to ask whether the truth theory CT^- is conservative, in which we assume only that the truth predicate is compositional. This is settled by the following theorem:

Theorem 2 (Kotlarski–Krajewski–Lachlan, Enayat–Visser, Leigh)**.** CT^- *is conservative over* PA*.*

Before we discuss the above theorem, let us comment upon its attribution. Kotlarski, Krajewski, and Lachlan (1981) proved a model-theoretic theorem which implied the conservativity of a certain theory very close to CT^-. When that paper was written, the axiomatic truth theories were not yet isolated as a separate field of research and their standard definitions were yet to be established. Therefore, the theory whose conservativity may be deduced from the Kotlarski–Krajewski–Lachlan's result is different from CT^- (it axiomatises satisfaction rather than truth) and it is not quite clear, how should we modify their proof in order to show the conservativity of CT^-. A conservativity proof of a compositional truth theory, much simpler than the argument in Kotlarski *et al.* (1981), has been obtained only by Enayat and Visser (2015). The theory which they investigated also was different from CT^-. This difference, however, was not so significant. The conservativity proof for CT^- has only been given by Leigh (2015) who used still other techniques.

2.1 Closure and Correctness Principles

Theorem 2 states that a purely compositional truth theory does not prove more arithmetical facts than PA alone. Only upon adding the induction scheme for the formulae containing the truth predicate will it allow us to prove, for instance, that PA is consistent.

Hence, we see two very natural theories CT^- and CT, only one of which is conservative. The induction for the truth predicate allows us to prove many facts about its structure. Compositionality, the basic feature of this predicate is not enough to prove new arithmetical theorems. Our question on the natural dividing line between truth theories which are conservative and not conservative over PA may be narrowed down to the following problem: what natural axioms characterising the truth predicate added to CT^- will make the resulting theory non-conservative over PA?

Ali Enayat suggested naming the dividing line between conservative and non-conservative truth theories between CT^- and CT **the Tarski boundary**[10]. Now our question may be expressed as follows: where is the Tarski boundary located? Let us discuss some natural axioms which extend CT^-, but are provable in CT. One very natural group of such axioms is the **closure and correctness principles**. Closure principles state that true sentences are closed under reasoning in a given deductive system. Correctness principles state that all sentences in a certain set are true. Let us present some principles of these sorts.

From the non-conservativity proof for CT, we may isolate one very simple correctness principle which definitely isn't conservaitve. Namely, **the principle of correctness of** PA:

> Every theorem of Peano arithmetic is true.

The above principle is also called **the global reflection principle** over PA. Let us notice that in the non-conservativeness proof for CT we have used exactly the fact that CT proves the correctness of PA.

We can isolate two further natural principles provable in CT which together imply the principle of correctness of PA. The first one is **the principle of closure under first-order logic**:

> Every sentence provable in first-order logic from true premises is true.

We can say that this principle is of more fundamental character than the principle of correctness of PA. It only says about the connection between

[10]Let us briefly justify the choice of the name. Tarski has been apparently the first one to point out the "weakness" of some arithmetic truth theories (inter alia TB^-). Besides that, CT^- is modelled after his inductive conditions defining the satisfaction relation. Ali Enayat and the authors of the paper has used this expression a few time in conference talks. However, to our best knowledge, this paper is the first place where the expression has been used in print.

truth and first-order logic and does not explicitly depend on our trust in the truth of the axioms of arithmetic. This trust is expressed by **the principle of axiomatic correctness of** PA:

Every axiom of PA is true.

Using standard proof-theoretic techniques, one can show that the arithmetical consequences of CT strictly contain the arithmetical consequences of CT^- with the principles of axiomatic correctness of PA and closure under first-order logic. Therefore, we have isolated a natural truth theory which is strictly weaker than CT but still not conservative over PA. It could seem that leaving any of these two axioms of this theory added to CT^- would also yield a non-conservative extension. However, it turns out that the principle of the axiomatic correctness of PA is one of the weak principles, which has been stated already in Kotlarski *et al.* (1981). These results have also been announced in (Enayat & Visser, 2015) and in (Leigh, 2015) (where it has been presented with proof). All of the cited sources bring different methods to demonstrate this theorem.

Theorem 3 (Kotlarski–Krajewski–Lachlan, Enayat–Visser, Leigh)**.** CT^- *with the principle of axiomatic correctness of* PA *is conservative over* PA*.*

Hence, it turns out that we can narrow down our search of the boundary between weak and strong compositional truth theories in a rather precise way. Adding to CT^- a principle that all axioms of PA are true is not enough to obtain new arithmetical consequences. On the other hand, extending this theory further with a principle that all sentences provable in PA are true, already turns out to be non-conservative.

2.2 Bounded induction scheme for the truth predicate

Another perspective which allows us to look for natural theories which are not conservative over Peano arithmetic but weaker than CT is restricting the induction scheme to some specific classes of formulae. We say that a formula is in the class Δ_0, if all quantifiers which appear in it are bounded, i.e., they are of the form $\forall x < t$, $\exists y < s$ for some terms t, s. Hence, the truth of sentences in the class depends only on objects of some fixed size. We can think of them as some special class of sentences whose truth value may be decided effectively. This class of formulae plays a very significant role in the research on metamathematical properties of arithmetic.

Another important class of formulae is Π_1. The formulae in this class are of the form

$$\forall x_1 \ldots \forall x_n \varphi,$$

where φ is a Δ_0 formula. We can think of them as purely universal formulae. They express that certain simple facts which may be decided effectively hold for all objects.

An important class of subtheories in Peano arithmetic are its fragments resulting from restricting the formulae in the induction scheme to the formulae of class Π_1 or Δ_0. We will follow this path also in the case of truth theories. By CT_1 we mean CT^- with all the instances of the induction scheme

$$\varphi(\overline{0}) \wedge \forall x(\varphi(x) \to \varphi(x+1)) \to \forall x \varphi(x)$$

in which the formula φ is an arbitrary formula of the class Π_1.

Let us note that the restriction to formulae in the class concerns only the formulae containing the truth predicate, since already CT^- contains all instances of the induction scheme for the arithmetical formulae, as an extension of PA.

The theory CT_1 is rather natural in the context of our research, since by inspection of the non-conservativity proof for CT, we can conclude that we in fact used only the axioms available in CT_1. We reach the following conclusion:

Theorem 4. CT_1 *is not conservative over* PA.

Indeed, one can easily show that the principle of correctness of PA, and consequently the principle of axiomatic correctness of PA, are provable in CT_1, as is the principle of closure under first-order logic. Therefore, we have reached another perspective allowing us to narrow down our search for the natural principles which make truth theory significantly stronger than its base theory. In particular, it is natural to ask about the conservativity of the theory CT_0 which results from restricting the induction scheme for formulae containing the truth predicate to Δ_0 formulae.

Let us add that TB (i.e., TB^- with the full induction scheme) is conservative over Peano arithmetic which is a significantly simpler result than Theorem 3. It follows that a truth theory with natural axioms including the full induction scheme does not automatically have to prove new arithmetical facts. Moreover, we can show examples of fully inductive theories nontrivially extending TB and based on some variants of Tarski's equivalences which are

still conservative. Therefore, the fact that we consider compositional truth theories is crucial for our results.

3. Discovering the Tarski Boundary

In this section we present the main known facts on the contour of the Tarski Boundary – most of the theorems, that we have stated, are yet unpublished, so the content of this section is to be treated as a report on a work in (hopefully) progress.

Let us start with briefly completing what we have already presented: in the last section we observed that (over CT^-) the closure under first-order logic principle in conjunction with the principle of axiomatic correctness of PA proves the principle of correctness of PA. It transpires that the first principle alone is capable of doing this: working in CT^- extended with the closure under the first-order logic principle, we will prove that each axiom of PA is true. The proof of this fact is rather standard and very intuitive: finitely many axioms fixing the basic rules of addition, multiplication and ordering are already true in CT^-. What remains are induction axioms: PA (even interpreted in a non-standard model) "thinks that" objects which it talks about are ordered as natural numbers, meaning that from 0 up to an element b there are only finitely many steps (obviously exactly b, which in a non-standard model can be a non-standard number). Working in CT^- with the closure under the first order logic principle and assuming $\varphi(\overline{0})$ and $\forall x(\varphi(x) \to \varphi(x+1))$ are true for a fixed arithmetical formula $\varphi(x)$, for an arbitrary a we will build a proof of $\varphi(\overline{a})$ in pure first order logic (we use the dictum de omni and modus ponens rules a many times). The proof runs in parallel to a classical argument that in the standard model of arithmetic the axioms of induction are true, with the only difference that in a non-standard model we use the induction axiom for a formula

$$\vartheta(x) := \mathsf{Prov}_{\mathsf{PA}}(\varphi(\overline{x})).$$

Since the set of true sentences is closed under reasoning in first-order logic, we can conclude that $\varphi(\overline{a})$ is true. Since a was arbitrary, we conclude that for all a $\varphi(\overline{a})$ is true, hence (on the basis of the compositional axioms) that the sentence $\forall x(\varphi(x))$ is true. This concludes our proof.

In an analogous way one can show (bypassing one additional difficulty – we refer the reader to (Cieśliński, 2010b) where this was proved for the first

time) that the principle of correctness of PA is equivalent to the following, much more restricted, correctness principle:

All validities of first-order logic are true.

The above principles might be naturally grouped into these claiming that the set of true sentences is closed under certain rules of inference (reasoning in first-order logic, for example) and these claiming that all sentences from a certain set are true (e.g. theorems of PA or theorems of first-order logic). Intuitively, the principles of the first kind say something more than their counterparts of the second kind. However, last year Cezary Cieśliński presented an insightful proof that over CT^- these principles are equivalent. Let us isolate it as separate

Theorem 5 (Cieśliński)**.** CT^- *extended with the axioms "All theorems of first-order logic are true" proves the closure under first-order logic principle.*

Searching for weaker principles provable in CT, but properly extending CT^- , let us extract from the closure under first-order logic **the principle of closure under propositional logic**:

> Each sentence provable in classical propositional calculus from true premises is true.

Obviously, over CT^- the above sentence is provable from the closure under first-order logic principle. As was shown by Cezary Cieśliński (2010a) CT^- , extended with the principle of closure under propositional logic, is equivalent to the compositional theory of truth with bounded induction, i.e. CT_0.

Theorem 6 (Cieśliński)**.** *The principle of closure under propositional logic is provable in* CT_0*. Each axiom of* CT_0 *is provable in* CT^- *extended with the principle of closure under propositional logic.*

Long before this paper of Cieśliński, Henryk Kotlarski (1986) published a proof that CT_0 proves the principle of correctness of PA. The argument, despite being concise, seemed correct and convincing enough to be cited also in Cieśliński (2010b) and Halbach (2011). However, in 2008 Albert Visser and Richard Heck noticed a gap in the proof of Koltarski: the problem was to demonstrate that CT_0 proves the sentence:

> Each axiom of first-order logic is true.

Kotlarski's proof worked fine for CT_0 extended with the above sentence. It is quite clear that the above can be proved with the help of the induction for Π_1 formulae. Reconstructing this proof with induction only for bounded formulae seemed so undoable that many logicians (including the authors of this paper) started searching for the proof that CT_0 lies on the conservative side of the Tarski Boundary.

Finally, it was shown (the proof appeared in (Łełyk & Wcisło, 2017b)) that CT_0 proves the same arithmetical sentences as CT_0 with the principle of correctness of PA added. Remarkably, in this proof only two very natural principles (provable in CT_0 but not in CT^-) were used: the first one was, introduced previously, the principle of axiomatic correctness of PA, the second was the generalized commutativity with the disjunction principle, called the disjunctive correctness principle:

> For all x and for every sequence of x sentences $\varphi_0, \ldots, \varphi_x$, their disjunction is true if and only if one of $\varphi_0, \ldots, \varphi_x$ is true.

Let us clarify this a little bit: for every natural number n, CT^-, using compositional axioms, will be able to prove that a disjunction of n sentences is true exactly when one of these sentence is. However, it will not be able to prove the above general statement[11]. It came as a surprise that such a simple generalization of compositional axioms, together with a (conservative when considered separately) principle of axiomatic correctness of PA, gives a theory which proves the same arithmetical sentences as CT^- with the PA correctness principle. Let us summarize this in the following:

Theorem 7 (W)**.** CT^- *extended with the principles of disjunctive correctness and the axiomatic correctness of* PA *proves the same arithmetical sentences as* CT^- *extended with the principle of correctness of* PA*.*

It is worth emphasizing that these results are not self-evident: so far it turns out that all the principles that we know to be located on the non-conservative side of the Tarski Boundary prove (at least) the consequences of the principle of correctness of PA (henceforth let us denote this principle with TPA). Let us observe that this set contains much more PA-unprovable sentences than simply the sentence naturally expressing the consistency of PA (abbreviated as $\mathsf{Con}_{\mathsf{PA}}$). We have already seen that this sentence is provable in $\mathsf{CT}^- + \mathsf{TPA}$. This is an arithmetical sentence, hence, applying

[11] This was first shown by Kotlarski *et al.* (1981). One can give an alternative proof based on Enayat and Visser methods of constructing full satisfaction classes.

finitely many times the compositional axioms, we can show that it is true. Since this theory proves the closure under first-order logic principle and we know that the axioms of $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}}$ are true, therefore no false sentence can be a consequence of this theory (in particular $0 = 1$ cannot). Thus we have just proved the consistency of theory $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}}$ i.e. the sentence:

$$\mathsf{Con}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}}}.$$

Nothing stops us from iterating this process further, this way proving stronger and stronger consistency assertions

$$\mathsf{Con}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}}}}$$

$$\mathsf{Con}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}}}}}$$

$$\mathsf{Con}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}}}}}}$$

and so on.

The arithmetical capacities of $\mathsf{CT}^- + \mathsf{TPA}$ does not stop there. It is not hard to convince oneself that it proves all sentences with the form

$$\forall x(\mathsf{Prov}_{\mathsf{PA}}(\varphi(x)) \to \varphi(x)) \tag{$*$}$$

for an arbitrary arithmetical formula $\varphi(x)$. The set of all sentences of this form is called the uniform reflection principle over PA[12]. A small subset of this set (for Π_1 formulae) is sufficient to prove all the above iterations of consistency statements. And that's not all: the set of (Gödel codes of) sentences of the above form is recursive (hence strongly representable in PA), hence in arithmetic we can define a theory

$$\mathsf{PA}^1 := \mathsf{PA} + \forall x(\mathsf{Prov}_{\mathsf{PA}}(\varphi(x)) \to \varphi(x))$$

for which the standard provability predicate will satisfy the Gödel–Löb conditions. In $\mathsf{CT}^- + \mathsf{TPA}$ we will prove all sentences of the form $(*)$ for PA^1, i.e. all sentences

$$\forall x(\mathsf{Prov}_{\mathsf{PA}^1}(\varphi(x)) \to \varphi(x)),$$

where $\varphi(x)$ ranges over arithmetical formulae with at most one free variable. In the next step we can define the theory PA^2, replacing in $(*)$ PA with PA^1. Iterating this process in the infinite, in the limit step taking

$$\mathsf{PA}^\omega := \bigcup_{n\in\omega} \mathsf{PA}^n$$

[12]It can be seen right now why we called the principle of correctness of PA the "global" reflection – in the presence of the truth predicate we can express the above principle in a single sentence.

we will obtain an arithmetical axiomatization of the arithmetical consequences of $\mathsf{CT}^- + \mathsf{TPA}$. A very elegant proof that PA^ω is really sufficient for deducing all arithmetical consequences of this theory of truth, was given by Henryk Kotlarski (1986).

The situation starts looking as if every "natural" theory of truth which proves the consistency of arithmetic, proved at the same time all the sentences from PA^ω. Obviously one can cook-up some artificial counterexamples to this "theorem": for example theory CT^- extended with the axiom "$\mathsf{Con}_{\mathsf{PA}}$ is true" is non-conservative over PA and much weaker than the considered "natural" theories (for example, it does not prove the sentence $\mathsf{Con}_{\mathsf{PA}+\mathsf{Con}_{\mathsf{PA}}}$). Obviously "natural" is not a formal notion, but it expresses a certain heuristics: it helps to temporarily block the ad hoc counterexamples. Right now we are trying to find a "natural" counterexample, possibly in the meantime realizing that no such counterexample can exist. Then we will probably understand what "natural" means.

Let us stress that in the above we did not say that CT_0 proves the principle of correctness of PA. The proof of non-conservativity of this theory consists in constructing a formula $T'(x)$ which, provably in CT_0 behaves like a predicate satisfying both CT_0 and the principle of correctness of PA. Using a definition the introduction of which we postpone for a moment (Definition 13), it has been shown that CT_0 augmented with the principle of correctness of PA is relatively truth definable in CT_0. However, we still didn't know whether the constructed formula $T'(x)$ *provably* in CT_0 had the same extension as the "original" truth predicate. Stating this less formally: it could be the case that CT_0 is able to "upgrade" its own truth predicate but cannot prove that its own truth predicate is as good (i.e. satisfies the principle of correctness of PA). In the meantime Ali Enayat[13] showed that focusing on the extension of CT^- with the principles of the axiomatic correctness of PA and disjunctive correctness, we were not in fact working with a weaker theory. He proved the following:

Theorem 8 (Enayat)**.** CT^- *extended with the principle of axiomatic correctness of* PA *and the disjunctive correctness principle proves* CT_0*.*

It turned out that, up to deductive equivalence and looking only at the theories that we can prove to be non-conservative, there are only two minimal theories above the Tarski Boundary: CT_0 and $\mathsf{CT}^- + \mathsf{TPA}$, and, moreover, that they are mutually relatively truth definable (Definition 13).

[13]Personal communication.

After all, it has been shown that this picture is even simpler: a direct fix to the old proof of Kotlarski was discovered. Let us summarize our findings in the following:

Theorem 9 (Cieślinski, Enayat, Kotlarski, Ł)**.** *The following theories are equivalent:*

1. CT_0.
2. CT^- *with the principle of correctness of* PA.
3. CT^- *with the principle of closure under first-order logic.*
4. CT^- *with the principle of closure under propositional logic.*
5. CT^- *with the principle of correctness of first-order logic.*
6. CT^- *with the principles of disjunctive correctness and axiomatic correctness of* PA.

So far the situation on the Tarski Boundary looks as if there were the least ("natural") theory which admits many different axiomatizations. Obviously, as the careful Reader has certainly noticed, some questions have been left unanswered in the above considerations. This was not accidental: as for this moment we still do not know whether the extension of CT^- only with the principle of disjunctive correctness is conservative over PA or not[14]. Intuitively, it should be a weak extension of PA. However it would not be the first time when our intuitions have failed...

It is worth mentioning, at least concisely, the possible impact of the above theorem on the philosophical debate over deflationism[15]. Assuming that the deflationist should present a theory which both proves some general facts about the truth predicate and is conservative over PA, it follows that his options are rather limited. He cannot, for example, demand both classical compositionality and closure under propositional logic from the truth predicate axiomatized by such a theory. Furthermore, he cannot even demand generalized compositionality (which would imply the disjunctive correctness principle) and the principle of correctness of PA. It might seem that this

[14] Telling the truth: we did not know this when writing the polish version of this paper. Recently, however, Fedor Pakhomov presented an insightful proof that CT^- with the principle of disjunctive correctness is actually the same theory as CT_0. This is a highly unexpected result.

[15] We thank the anonymous referee for the suggestion of adding this remark to the paper.

situation is hopeless. To obtain this conclusion, however, we need to assume that the deflationary theory of truth needs to prove general facts about the behavior of the truth predicate and be conservative. This requirement is based on yet another assumption: we have to agree that provability in a theory is a good enough explication of the notion of justification or explanation (depending on how the thesis of deflationism is formulated). This view has been recently ciriticised at length in (Cieśliński, 2017) and we have to admit that right now we are unconvinced as to whether Theorem 9 can really play a role in this debate.

3.1 The Tarski Boundary and different truth theories

Let us observe that we can also ask about the contour of the Tarski Boundary with respect to theories of truth different from CT^-. For example, we can start from the least (thus far) "natural" non-conservative theory of truth i.e. CT_0 and weaken the compositional axioms, modelling them not after the classical logic, but, for example, on strong Kleene logic. In such a theory, known as PT_0[16] we do not have a global axiom for the negation, i.e.

> (For every sentence φ) The negation of φ is true if and only if φ is not true.

Instead for every connective (negation included) and quantifier we say separately when the negation of a sentence beginning with this connective is true. For example, the following sentence is an axiom of PT_0

> (For all arithmetical sentences φ, ψ) The negation of the conjunction of φ and ψ is true if and only if the negation of either of φ or ψ is true.

We can now ask: does the contour of the Tarski Boundary depend on which logic we choose for the compositional and Δ_0 inductive truth predicate? This question is one of the topics of our current research[17].

[16] More precisely, in the literature only the non-inductive version of this theory, denoted PT^- (or $\mathsf{PT}\restriction$) is known, but PT_0 is simply this theory with axioms of induction for the Δ_0 formulae of the extended language.

[17] Now we know that some extensions of PT^- (compare footnote 16) are strong but weaker than CT_0. The question whether every natural strong extension of CT^- proves Global Reflection is still open.

4. Other Notions of Conservativity

The question about the conservativity of a given theory is just the first step into differentiating various axiomatic theories of truth. It can be taken as the first approximation, classifying the theories as either strong or weak. More generally we can ask which theories are stronger than other theories (in particular: comparing the non-conservative theories). In this the following (obvious in fact) generalization of the notion of conservativity can be used:

Definition 10. A theory Th_1 is syntactically stronger than Th_2 if and only if the arithmetical consequences of Th_2 form a proper subset of the set of arithmetical consequences of Th_1.

One can show, for example, that CT_1 is syntactically stronger than CT_0 and CT is stronger than CT_1. Non-stratified, compositional and fully inductive theories of truth are usually still much stronger, for example FS is stronger than CT, KF than FS and VF than KF[18].

Observe, however, that distinguishing theories only on the basis of their arithmetical consequences blurs the differences between many theories, whose axioms have intuitively a very different character. For example, both CT^- and TB are syntactically conservative over PA, hence they cannot be told apart solely on the base of their arithmetical consequences. One can consider also a different measure which would enable us to differentiate between theories with the same syntactical strength. This measure is based on a, well-known from the literature, notion of semantical conservativity:

Definition 11. A theory Th is semantically conservative over PA if and only if every model of PA can be expanded (with preservation of the universe and arithmetical functions) to a model of Th.

The philosophical intuition motivating this notion is as follows: we think about models of a theory as "possible worlds" ("possible" from the point of view of the considered theory). If a model of PA cannot be extended to a model of Th, it means that such possibility, while admitted by PA, is excluded by Th. It is worth noticing that semantical conservativity implies syntactical conservativity (by Completeness Theorem), but it does not reverse: neither TB, nor CT^- is semantically conservative. This notion can be generalized in the following way:

[18]These names are standard in the literature. Definitions of KF and FS can be found in (Halbach, 2011), whereas VF was introduced in Cantini's paper (1990).

Definition 12. A theory Th_1 is semantically stronger than a theory Th_2 if and only if the class of models of PA that can be expanded to models of Th_1 is a proper subclass of the class of models of PA, that can be expanded to models of Th_2.

Basing on the intuition just introduced, we can say that Th_1 is semantically stronger than Th_2, if Th_1 eliminates more "possible worlds", than Th_2. Using this distinction one can prove that TB is semantically weaker than CT^-, which matches our intuitions that compositional axioms "say more" about the notion of truth, than Tarski biconditionals (even when considered in the presence of full induction).

The most fine-grained relation which can differentiate between various truth theories, was introduced by Kentaro Fujimoto in (2010) and is known as relative truth definability.

Definition 13. Let Th_1 and Th_2 be two truth theories. We say that Th_1 is relatively truth definable in Th_2 if and only if there exists a formula $\varphi(x)$ such that Th_2 proves the axioms of Th_1 with $\varphi(x)$ substituted for the truth predicate of Th_1.

To say the same things in simple words (perhaps less precisely): Th_1 is relatively truth definable in Th_2 if Th_2 can define the truth predicate which satisfies the axioms of Th_1. We shall say that Th_2 is Fujimoto-stronger than Th_1 if Th_1 is relatively truth definable in Th_2 but not vice-versa. The proof of Theorem 7 shows that CT_0 together with the principle of correctness of PA is relatively truth definable (hence not Fujimoto-stronger than) in CT^- with the principles of disjunctive correctness and axiomatic correctness of PA. There are theories which can be distinguished only by the above relation, for example TB^- and UTB^-[19].

5. Summary and Open Problems

We began the paper with introducing the most basic measure of strength of axiomatic theories of truth, according to which a theory is classified as strong if it proves some sentences which are unprovable in PA. The boundary between strong and weak axiomatic theories of truth was called the Tarski Boundary. The most important discovery concerning the contour of the Tarski Boundary can be summarized as follows: each "natural" theory of

[19]Defined in Halbach (2011) as $\mathsf{TB}\restriction$ and $\mathsf{UTB}\restriction$.

truth, which up till now has proved to be strong, proves CT^- with the principle of correctness of PA[20]. Moreover this last theory admits many different axiomatizations, one of them being CT^- augmented with a scheme of induction for bounded formulae with the truth predicate (this theory was called CT_0). Lastly, we showed that there exist two interesting strengthenings of the introduced measure, which help us to discern between the strength of truth theories for which the basic notion was too coarse-grained. It is worth emphasizing that still there are many interesting open questions concerning the "strength" of axiomatic truth theories. We list some of them below:

1. Is CT^- with the principle of disjunctive correctness conservative over PA[21].

2. Is CT^- with the principle of the correctness of propositional logic conservative over PA? Let us notice that the above additional axiom is a correctness principle corresponding to the principle of closure under propositional logic. The latter one (over CT^-) is equivalent e.g. to the Global Reflection principle, hence is very strong.

3. Is CT^- semantically stronger than UTB? We know that $\mathsf{CT}^-$is at least as strong as UTB (i.e. every model which can be expanded to a model of CT^-, can be expanded to a model of UTB; it was proved in Łełyk and Wcisło 2017a).

4. Does CT^- relatively truth define UTB?[22]

Bibliography

Cantini, Andrea (1990). A theory of formal truth arithmetically equivalent to ID_1. *Journal of Symbolic Logic*, *55*(1), 244–259.

Cieśliński, Cezary (2017). *The epistemic lightness of truth. Deflationism and its logic*. Cambridge: Cambridge University Press.

[20] Actualization: now we know quite a few such theories, one of them being PT_0, compare footnote 17. The question is still open for extensions of CT^-.

[21] Actualization: this question was already answered by Fedor Pakhomov, compare footnote 14.

[22] Recently, this has been answered in the negative by Albert Visser (private communication).

Cieśliński, Cezary (2010a). Deflationary truth and pathologies. *Journal of Philosophical Logic*, *39*(3), 325–337.

Cieśliński, Cezary (2010b). Truth, conservativeness, and provability. *Mind*, *119*(474), 409–422.

Enayat, Ali, Visser, Albert (2015). New constructions of satisfaction classes. In Theodora Achourioti, Henri Galinon, José Martinez Fernández (Eds.), *Unifying the philosophy of truth* (pp. 321–335). Dordrecht: Springer Netherlands.

Franzen, Torkel (2003). *Inexhaustibility. A Non-Exhaustive Treatment*. Association for Symbolic Logic. Wellesley: A K Peters.

Fujimoto, Kentaro (2010). Relative truth definability of axiomatic truth theories. *Bulletin of Symbolic Logic*, *16*(3), 305–344.

Halbach, Volker (2011). *Axiomatic theories of truth*. Cambridge: Cambridge University Press.

Kaufmann, Matt, Schmerl, James (1987). Remarks on weak notions of saturation in models of Peano Arithmetic. *Journal of Symbolic Logic*, *52*(1), 129–148.

Kaye, Richard (1991). *Models of Peano Arithmetic*. New York: Clarendon Press.

Ketland, Jeffrey (1999). Deflationism and Tarski's paradise. *Mind*, *108*(429), 69–94.

Kotlarski, Henryk (1986). Bounded induction and satisfaction classes. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, *32*(31–34), 531–544.

Kotlarski, Henryk, Krajewski, Stanisław, Lachlan, Alistair (1981). Construction of satisfaction classes for nonstandard models. *Canadian Mathematical Bulletin*, *24*, 283–293.

Leigh, Graham (2015). Conservativity for theories of compositional truth via cut elimination. *Journal of Symbolic Logic*, *80*(3), 845–865.

Łełyk, Mateusz, Wcisło, Bartosz (2017a). Models of weak theories of truth. *Archive for Mathematical Logic*, *56*(5–6), 453–474.

Łełyk, Mateusz, Wcisło, Bartosz (2017b). Notes on bounded induction for the compositional truth predicate. *The Review of Symbolic Logic*, *10*(3), 455–480.

Shapiro, Stewart (1998). Proof and truth: through thick and thin. *Journal of Philosophy*, *95*(10), 493–521.

Originally published in *Studia Semiotyczne* 31/1 (2017), 23–44.

STUDIA SEMIOTYCZNE—ENGLISH SUPPLEMENT, vol. XXIX (2017), 127–158
ISSN 2451-2745
DOI: 10.26333/stsen.xxix.07

Olga Poller[1]

All The Superhero's Names

Abstract In this paper I concern myself with *The Superman Puzzle* (the phenomenon of the substitution failure of co-referential proper names in simple sentences). I argue that the descriptive content associated with proper names, besides determining the proper name's reference, function as truth-conditionally relevant adjuncts which can be used to express a manner, reason, goal, time or purpose of action. In that way a sentence with a proper name 'NN is doing something' could be understood as 'NN is doing something as NN' (which means 'as-so-and-so'). I argue that the substitution of names can fail on modified readings because the different descriptive content of proper names modifies the main predicate differently. Here I present a formal representation of modified predicates which allows one to model intuitively the different truth-conditions of sentences from *The Puzzle*.

Keywords The Superman Puzzle, proper names, substitution failure, qualifying prepositional phrases, modified predicates, descriptivism, adjuncts, pseudonyms, simple sentences

1. Introduction: Double life

By the 1970s, Romain Gary, the French novelist, was a literary celebrity. A decorated war pilot and diplomat he won the Prix Goncourt in 1956, at the beginning of his career as a novelist. But twenty years later, critics and readers were sated with the books of a fading literary star. So, while still publishing as Romain Gary, he created a new identity, that of a young Algerian student, Émile Ajar, who had fled to Brazil to escape jail and from

[1]Institute of Philosophy, Jagiellonian University
E-mail: kukuxa@gmail.com

where he was sending his manuscripts. In 1975, the second of Ajar's novels became a literary sensation and the Académie Goncourt awarded the prize to the author whilst knowing nothing about his real identity. In such a way Gary became the only person to win the Prix Goncourt twice. Knowing that Gary and Ajar is one and the same person, consider:

(1) Romain Gary won the Prix Goncourt in 1956.

(1') Émile Ajar won the Prix Goncourt in 1956.

(2) Émile Ajar, not Romain Gary, won the Prix Goncourt in 1975.

(2') Romain Gary, not Émile Ajar, won the Prix Goncourt in 1975.

Sentences (1) and (2) are true but our intuitions about the truth-value of (1') and (2') are mixed. On the one hand, Gary and Ajar is one and the same person and it is true about this person that he won the prize in 1956 and in 1975, but on the other hand, while being Romain Gary, he didn't win the prize as Romain Gary in 1975, and didn't win it as Ajar in 1956.

Here is another story. The greatest boxer Muhammad Ali lost five fights in his boxer-career but he never lost a fight before he changed his name. Consider:

(3) Cassius Clay was never beaten, whereas Muhammad Ali lost five times.

(4) Muhammad Ali lost more fights than Cassius Clay.

Sentence (3) and (4) could be true (actually that is how people complain on Ali's fanpages) but again you may have mixed intuitions about their truth-value. Sentences (1)–(4) exemplify three main cases of *The Superman Puzzle* – the phenomenon of the substitution failure in simple sentences which occurs when a change from one co-referential name to another affects the truth-value of a sentence in an extensional context. By Case 1 (C1) I will understand a situation in which one and the same person (or object) with names 'NN' and 'MM' simultaneously does something as NN and does something else as MM (or does something as NN but does not act as MM (while still being MM)). C1 is represented by sentence (2) – the same person, Romain Gary, won the Prix as Émile Ajar, not as Romain Gary, while still being Romain Gary. By Case 2 (C2) I will understand a situation in which the same person (or object) does something as NN at one time and does something as MM at another time (sentence (3)). Finally, all sentences with

comparative quantifiers (e. g. 'more than' in sentence (4)) will constitute Case 3 (C3).

I have presented three cases of *The Puzzle* using genuine proper names, not pseudonyms, but most examples you can find in the philosophical literature concern the names of superheroes:

C1:
'While talking on the phone to Superman, Lois looked through the window at Clark Kent' (Moore, 1999, p. 102).
'Clark Kent went into phone booth and Superman came out' (Saul, 1997, p. 102).

C2:
'I never made it to Leningrad, but I visited St Petersburg last week' (Saul, 1997, p. 103).

C3:
'Superman was more successful with women than Clark Kent' (Saul, 1997, p. 103).
'Superman leaps tall buildings more frequently than Clark Kent' (Moore, 1999, p. 92 n. 1).
'Hammurabi saw Hesperus more often than he saw Phosphorus' (Crimmins, 1998, p. 19).

Note that *The Puzzle* appears only for those who know that names 'NN' and 'MM' refer to one and the same person (Moore (1999) proposed calling such people 'enlightened'). So if you are enlightened it seems that you have to choose between two ways of explaining why intuitively the substitution of co-referential proper names fails in sentences (1)-(4) and why sentences as (2), (3) and (4) seems true. You can say that the truth-values of (1) and (1') differ because these sentences express different propositions (that is exactly why (2), (3) and (4) are true – they express a proposition other than an analytically false one). Or, on the contrary, you can say that sentences (1) and (1') semantically expresses one and the same proposition but pragmatically convey different ones, and that is why people have mixed intuitions about the truth-conditions of sentences (1)-(4). I will call a view of the former type semantic and of the latter type pragmatic.

The plan of this paper is as follows. In the next section I shall explain why *The Puzzle* puzzles. In section I briefly explain my proposal of semantics

for qualifying prepositional 'as'-phrases ('as so-and-so') which I will analyze in a similar way as adverbs are treated – as predicate modifiers. In section 4 I lay the groundwork for my own proposal. I will develop a hypothesis that the descriptive content of proper names could behave as truth-conditionally relevant adjuncts and be an additional contribution of proper names to the truth-conditions. Finally, in the Appendix, I will present a formal semantics for predicate modifiers and a model for one of *The Puzzle* sentences.

2. Why *The Puzzle* puzzles

Let me start from the semantic type of view. The proponents of such a view assume that sentences from *The Puzzle* express different propositions so the core of the puzzle lies in giving a semantic explanation as to why sentences which seem simple, differing in co-referring proper names only, nevertheless express different propositions. Let us have a closer look at such a sentence. Consider: 'Superman is successful with women but Clark Kent is not'. It seems at first glance that if you accept the Leibniz Law of the Indiscernibility of Identicals you face a dilemma: either you have to give up names' co-referentiality, or have to accept the view that such sentences are always false. Link – who was trying to solve the similar puzzle of substitution failure between co-referential group terms and between coextensive plural terms – expresses the former possibility in the following way (1983, p. 304): 'So if we have, for instance, two expressions a and b that refer to entities occupying the same place at the same time but have different sets of predicates applying to them, then the entities referred to are simply not the same'. If you give up the co-referentiality of names then the problem of substitutivity failure becomes trivial. In the case of *The Superman Puzzle*, David Pitt (2001), Bjørn Jespersen (2006) and (a contextual version of it) Joseph Moore (1999) hold such a view. According to them, sentences from *The Puzzle* express different propositions because proper names are not genuinely co-refering (they refer to different fusions of time-slices (Pitt) or to different aspects (Moore) of the same individual, or they refer to different individual concepts (Jespersen)). It is little wonder that giving up co-referentiality leads to problems with identity statements. Identity statements expressed by sentences of the form 'NN is MM' come out false (or at least are false in some contexts). Besides this unintuitive consequence, this type of a solution blocks the substitution of proper names in situations

in which it is intuitively allowed (Predelli, 2004, p. 110; Saul, 2000, p. 256; Saul, 2007, pp. 33–34).

So perhaps it would be better to keep the co-referentiality of names and, in order to explain how sentences from *The Puzzle* could express different propositions, to give up the claim that sentences are simple (to give up the principle called by Predelli (2004, p. 108) 'Syntactic Innocence'). Such a line of explanation was taken by Graeme Forbes (1997, 1999, 2006) who noticed that sentences from The Puzzle, as for example, 'Lex fears Superman', could be paraphrased with the pronoun 'such', 'Lex fears Superman as such' (2006, pp. 157–58). According to Forbes, in the case of substitution failure, simple sentences should be understood as containing the covert prepositional phrase 'as such' in which the pronoun 'such' should be treated as a case of logophora (a special case of anaphora in which an expression serving as antecedent is taken itself as a referent of an anaphoric pronoun). In a nutshell, the Forbesean idea was to treat dossiers of information (or, more precisely, a capacity to activate a certain dossier) as a representation of Fregean modes of presentation (2006, p. 158). A speaker could create different dossiers in which he stores different information about one and the same object. A proper name serves as a label for somebody's dossier; so if you substitute one proper name in a sentence for a different but co-referential one, you will change the reference of a covert pronoun while the referent of a name will remain the same. The new label will activate a different dossier so all you have to do to get a difference in truth-conditions is to connect expressions and dossiers (modes of presentation) with a special function which induces opacity and makes a mode of presentation which is connected with a name as part of the truth conditions (2006, pp. 158–59).

Mark Crimmins (1993, p. 273) raised an objection to the general version of this view (which covers belief ascriptions) and proposed the consideration of a story in which Lois encounters Superman in both guises but does not know either of his names. We can report for example: 'Lois believes that Clark is in the building, but doesn't believe that Superman is in the building'. Intuitively, this sentence is true, but the possibility of using Lois's unlabelled dossiers is ruled out on Forbes' account.

So perhaps a better idea would be to preserve both co-referentiality and syntactic simplicity and shift the criteria of evaluation. Stefano Predelli (2004) followed this line and noticed that sentences from The Puzzle could be uttered in different contexts with different focuses of conversation. It could be so that, due to a special focus of a conversation in a context, some contextually salient circumstances should be taken into account in order to

decide if a proposition expressed by a sentence in this context is true or not. The use of a name in a context triggers some features of the name's bearer which are of importance due to the focus of a conversation. Taking these features into account, the conversation participants decide if a referent of a proper name belongs to the extension of a predicate or not. Note that we are talking about the features of one and the same referent of both names and, once these features are taken into account, nothing prevents the substitution of proper names (if they all contribute to truth-conditions as their referent). Saul (2007, pp. 55–56) objected that it is not clear what these circumstances are and how to use them in order to solve examples of C3.

Let us leave the semantic camp and see what the proponents of the pragmatic view would propose. According to such views, sentences from *The Puzzle* semantically express one and the same proposition but pragmatically convey different ones. Alex Barber (2000) tried to explain *The Puzzle* using Gricean notion of implicature. A speaker uttering 'Superman is more successful with women than Clark Kent' semantically expresses an analytically false proposition but his conversational partner assumes that the speaker is preserving the Cooperative Principle and is talking as if he is one who is unaware that Superman is Clark Kent. Those who are unaware (unenlightened speakers) would, under foreseeable epistemic conditions (for example taking into account attributes of appearing), utter what the speaker uttered (2000, pp. 303–304).

But what about truth-conditions? As we know, an implicature is not a part of the truth-conditions of a proposition literary expressed. Consider:

(5) If Clark Kent didn't ever pick up a woman and Superman did, then Clark Kent is more successful with women then Superman.

We could have mixed intuitions about 'Superman is more successful with women than Clark Kent' but sentence (5) strikes us as false (or even inconsistent). But it should be true on Barber's account (because it is an implication from false to false). So it seems that the pragmatic view leads to a dilemma: either the information pragmatically conveyed is a part of what is said and affects the truth-conditions or the truth-conditions of what is said differs radically from our intuitions. Note that if you accept the former claim (as Recanati (2012, p. 203 n. 5) did) you will owe the same explanation as the proponents of a semantic view.

So the main problem for a real pragmatist is to provide semantically adequate truth-conditions for sentences from *The Puzzle*. It has to be said that a lot of people have an intuition similar to Barber's in that enlightened speakers uttering such sentences somehow pretend. Thomas Zimmermann

(2005) elaborated this intuition and tried to fix the problem with the right truth-conditions. In a nutshell, what makes speakers unenlightened is the lack of knowledge that NN and MM is one and the same person. So when enlightened speakers utter sentences from The Puzzle they pretend and talk as if they were unenlightened: 'If I believed that NN is not MM then I would say that NN is Q'. Zimmermann calls such utterances 'counterfactual speech acts' (2005, pp. 77-78). According to him, in our conversational practice we naively assume that no two names of our language have the same bearer (Principle of Uniqueness (UP), 2005, p. 70). This assumption is rather a naive belief, nevertheless, according to Zimmermann, it is a cornerstone of our conversational behavior and constitutes one of the conversational principles. So when one enlightened speaker talks to another and uses two co-referential names, he violates one of the conversational principles and this in turn triggers an implicature that the speaker does so in order to convey another proposition. But what about truth-conditions? Let us recall Frege's criterion of thought difference (1892/1984, p. 162): 'Anybody who did not know that the evening star is the morning star might hold the one thought to be true, the other false'. According to this criterion, two sentences with co-referential names express two different thoughts (which are the same in terms of truth-value) and for somebody one of the thoughts could be true and the other could be false with respect to the things he believes. Zimmermann uses this criterion: sentences from *The Puzzle* have the same objective truth-value but could differ in truth-value with respect to somebody's doxastic perspective (differ in a subjective truth-value). So when an enlightened speaker violates UP he switches his language to the subjective language of unenlightened speakers who believe wrongly that 'NN' and 'MM' refer to different people. 'Switching languages' is expressed formally as changing the context of uttering to another which is exactly the same except for the language it is spoken in. So we get intuitively right truth-conditions (sentence (5) appears false) in a subjective language of those who believes that names 'NN' and 'MM' refer to different people. This last claim makes this solution similar to the proposal of all of those from the semantic camp who assume that proper names do not genuinely co-refer and that is why they have a similar problem with the falsity of identity statements (2005, pp. 94–95).

I hope I have convinced you that *The Puzzle* puzzles and now I intend to present my solution to it.

3. Modified predicates

I will remain in the semantic camp and develop an idea similar to the Forbesean. I take *The Superman Puzzle* to be a case of a broader phenomenon of substitution failure of co-referential nominal phrases: apart from proper names, this phenomenon concerns co-referential group terms (The Committee Puzzle), plural terms, definite descriptions and natural kind terms (Link, 1983; Landman, 1989; Szabó, 2003). In (Poller, 2016) I raised a hypothesis that the role of a descriptive content associated with proper names (and other terms) could not be the only reference determining but this content could also serve as a truth-conditionally relevant adjunct used to express a manner, reason, goal, time or purpose of action. The idea in a nutshell is to treat identifying descriptions 'the so-and-so' associated by speakers with a proper name as qualifying prepositional phrases 'as so-and-so'. In such a way, a sentence containing a proper name 'NN is doing something' could be understood as 'NN is doing something as NN' (which means as so-and-so). I present the semantics of prepositional 'as'-phrases briefly (elaborated version of it you can find in (Poller, 2016)) and then turn to a way of how it could be used to solve *The Puzzle*.

Consider the following sentence:

(6) The papal nuncio supported an anarchist protest.

Although nothing prevents one understanding (6) as saying the papal nuncio supported an anarchist protest as a private person you understand (6) rather as (6'):

(6') The papal nuncio supported an anarchist protest as the papal nuncio.

We could paraphrase (6') as (6"):

(6") The papal nuncio as such supported an anarchist protest.

I agree with Forbes who noticed that all the sentences which form *The Puzzle* could be paraphrased with the pronoun 'such' (cf. 'Lex fears Superman *as such*', 2006, p. 158) and espouse the view that 'as'-phrase invokes a mode of presentation connected with an expression. But contrary to Forbes, who treats 'such' as a case of logophora, I think of 'such' as an adjectivally anaphoric pronoun standing for a property (after (Carlson, 1980), (Landman & Morzycki, 2003), (Landman, 2006), (Siegel, 1994), (Wood, 2002)) and see no reason to think that the preposition as induces opacity. I propose

analyzing prepositional 'as'-phrases in a similar way to that in which adverbs are analyzed – as predicate modifiers[2].

In my analysis of 'as'-phrases, I followed Romain Clark (1970) who proposed a semantics for adverbs and prepositional phrases which was an alternative to events semantics proposed by Donald Davidson (1967/2001). The core of Clark's proposal is the idea that predicates could be built recursively out of n-place predicate constants by adding modifiers which have i places in total. So for example take 'stroll'. It is a 1-place predicate. Take the adverb 'slowly'. If you add this adverb to 'stroll' (getting 'slowly stroll') you would not increase the number of argument places. So 'slowly' is 0-place modifier (as are many other adverbs). The extension of 'slowly stroll' is a subset of the extension of 'stroll' (Clark, 1970, p. 325) and that is why you can infer from 'Sebastian slowly strolled' that 'Sebastian strolled' but not the other way around. This type of adverbial entailment failure is known as *Non-Entailment* (Davidson, 1967/2001; Katz, 2008) and we will see that it is a key property in solving the failure of the substitution puzzle. Now take 'at' and 'through'. Each of them are 1-place modifiers and if you add them to 'stroll' (getting 'stroll-through-at') you will increase the number of argument-places and will get a new 3-place predicate out of a 1-place initial one. You can infer from 'Sebastian strolled through the streets of Bologna at 2 a.m.' (Davidson, 1967/2001, p. 167) that 'Sebastian strolled' because the new 3-place predicate is connected with the initial 1-place predicate 'stroll' by a requirement that an object occupying the first place of the triple (Sebastian) should belong to the extension of 'stroll' (this type of entailment is called *Drop*).

I propose treating prepositional 'as'-phrases as 0-place predicate modifiers. Unlike other prepositional phrases, 'as'-phrases do not increase the number of argument-places, and, unlike adverbs, do not modify a predicate with all its argument places as a whole, they modify it on one argument-place only. Note that if you know that d is doing A and B and is φ, you can't infer that either A or B is done by d as φ (by *Non-Entailment*). This

[2]An anonymous referee noted that a placing syntactically 'as'-phrase as a predicate modifier (John as a miner supported a protest) seems unintuitive, and the 'as'-phrase should be analyzed as a name-modifier instead (John as a miner supported a protest). Such a line of analysis was used by Landman (1989). Szabó (2003, p. 391) raised convincing syntactical objections against such a view: modified names ('John-as-a-miner') do not coordinate with other names, cannot form possessives and cannot be given as an answer for 'who'-questions. Taking these arguments into account I analyze 'as'-phrases as predicate modifiers. I answered syntactic and semantic objections raised by Szabó against such a view in (Poller, 2016).

entailment failure shows that the extension of a modified predicate doing A as φ although dependent on the extensions of A and φ (by *Drop*), is not fully determined by them.

Let me briefly go through some syntactic and semantic definitions. For a modifier we will understand all predicates abstracted[3] from an atomic formula or a conjunction of atomic formulas with one free variable, e.g. $\lambda x.Q(x)$, $\lambda x.(P(x) \wedge Q(x))$. An n-place predicate constant Q could be modified by a modifier $(\lambda x.\varphi)$ on its ith argument place; we write this new modified predicate as '$Q^i_{\lambda x.\varphi}$'. For example greet is a two-place predicate, $\varphi(x)$ is a formula with one free variable in which φ means 'a host of a party'. $greet^1_{\lambda x.\varphi}$, $greet^2_{\lambda x.\varphi}$ are predicates built via modification from the predicate constant greet; we read them 'as a host of a party x greets y' (modification on the 1st argument place) and as x greets y as a host of a party' (modification on the 2nd argument place). We will use a simplifying convention and in the case that a modifier is a predicate abstracted from an atomic formula, $P(x)$, we will simply write 'Q^i_P' instead of '$Q^i_{\lambda x.P(x)}$ and in the case that Q is 1-place predicate we will write 'Q_P' instead of 'Q^1_P'.

I limit predicates abstracts which could be modified to predicates abstracted from atomic formulas and their negations, $\lambda x.Q(z_1, \ldots, z_n)$ and $\lambda x.\tilde{Q}(z_1, \ldots, z_n)$. A modifier $\lambda y.\psi$ modifies a predicate abstract on ith argument place of Q (written '$(\lambda x.\varphi)^i_{\lambda y.psi}$' in general notation). I preserve an intuition that a modified predicate abstract $(\lambda x.Q(z_1, \ldots, z_n))^i_{\lambda y.\psi}$ and a predicate abstracted from a formula with a modified predicate $(\lambda x.Q^i_{\lambda y.\psi}(z_1, \ldots, z_n))$ are one and the same predicate (so you can take a modifier 'in and out' of a predicate abstract, see (Poller 2016) for proof). Formulas with all kinds of predicates (predicate constants, predicate abstracts, modified predicates and modified predicate abstracts) are built in a standard way.

Let Q and P be 1-place predicates. I defined an interpretation of modified predicate Q_P ('Q as P') as a subset of a conjunction of interpretations Q and P: $I(Q_P) \subseteq (I(Q) \cap I(P))$. So, for example, d could be the papal nuncio $(d \in (P))$ and could support an anarchist protest $(d \in I(Q))$, but could support an anarchist protest not as the papal nuncio $(d \notin I(Q_P))$. (For the general definition of an interpretation of a modified predicate see Def. III.S7 in Appendix). A modified predicate is still a predicate, it is interpreted as a subset of a predicate being modified, that is, a set of n-tuples such that

[3]In using 'predicates abstracted from a formula', 'predicate abstracts' I followed Fitting and Mendelsohn (1998, pp. 194, 196, Definition 9.4.2)

every ith element in n-tuple fulfils the descriptive content φ. Modifiers are closed under the conjunction: $I(Q^i_{\lambda.x(\varphi\wedge\psi)}) = I(Q^i_{\lambda x.\varphi}) \cap I(Q^i_{\lambda x.\psi})$.

My analysis covers uses of 'as'-phrases as adjuncts of manner ('I will use the rest of the olive oil as a base for salad dressing'), time ('Ann was fat as a child'), reason ('As a firefighter, John was asked to help in the rescue action') and purpose ('They hired him as a launching engineer'). But it doesn't cover uses of 'as'-phrases as adjuncts of comparison ('His mother still treats him as a child') when we compare two things A and B under respect C and do not say that A is B (contrary to requirements of our semantic definition).

4. Names and pseudonyms as modifiers

Let us return to sentence (1'), 'Émile Ajar won the Prix Goncourt in 1956'. The reason why we may have mixed intuitions about its truth conditions lies in the ambiguity between modified and unmodified readings. You can say, 'It's true that Ajar won the Prix Goncourt in 1956, but Ajar won the Prix not as such but as Romain Gary'. The possibility of replacing a proper name with the adjectivally anaphoric pronoun 'such' supports the claim that a proper name in an 'as'-phrase ('didn't win as Émile Ajar') is understood as standing for a property, so the predicate 'win' is modified not by a proper name but by the descriptive content of a proper name. The idea standing behind the modification of predicates by names is simple: the modifying content of a proper name n is a predicate $\lambda x.\varphi$ abstracted from the formula φ of a definite description $\iota y.\varphi$ connected with a proper name n.

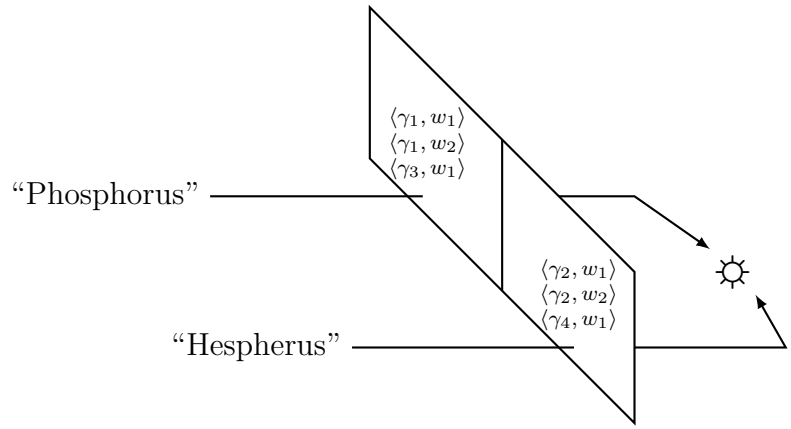
Despite being a descriptivist (in my opinion, speakers do associate definite descriptions with proper names) I do not think that the phenomenon of predicate modification by a descriptive content of names should be understood as evidence supporting descriptivism. Possibly you can accept this phenomenon without accepting any version of descriptivism (however, you will need an additional explanation of what kind of descriptive content should be semantically connected with names and why). Because of my claim that the modifying content of a proper name is a property expressed by a description connected with the name, I need to briefly explain my proposal of the formal representation of proper names in accordance with the descriptive theory of reference (descriptions are used to fix a name's reference, a full version of this proposal can be found in (Poller 2014)). In a nutshell I represent proper names formally as a special kind of term (which I call 'name-terms') which designates via sets of definite descriptions. By 'definite description' I understand a special kind of iota-terms of the form

$\iota x.[_i]\varphi$, where '$[_i]$' is a notational variant of $then_i$ operator ('true at t_i') taken after (Rini & Cresswell, 2012). Time operator $[_i]$ fixes a time of evaluation, so a definite description $\iota x.[_i]\varphi$ designates with respect to any time t the object designated by iota-term $\iota x.\varphi$ with respect to time t_i (I call definite descriptions $\iota x.[_i]\varphi$ actual with respect to t_i). I am trying to catch the idea that a definite description designates contingently with respect to possible worlds but if it designates in a world, it designates in that world one and the same object with respect to any time. That is why a iota-term representing a definite description should have a fixed time-parameter (e.g. 'the present Pope', 'the Pope in 1967').

My account of modified predicates is not general so the most complicated modifier could be a predicate abstracted out of a conjunction of atomic formulas with one free variable. That is why I will use only some of the iota-terms $\iota x.[_i]\varphi$, such that φ is a conjunction of atomic formulas. To avoid circularity (to be sure that definite descriptions $\iota x.[_i]\varphi$ used to determine a name-term's reference contain no name-terms) I need two languages, $\mathcal{L}$ and $\mathcal{L}^+$ ($\mathcal{L} \subset \mathcal{L}^+$). Let me start from language $\mathcal{L}$ which contains only variables and iota-terms as terms. The idea is to let name-terms designate through equivalence classes of descriptions designating one and the same object. But descriptions designate different objects with respect to different worlds so we need to define an equivalence relation not on a set of descriptions but on a set of pairs containing a description and a world in which the description designates. In order to be able to formally distinguish two co-referential names I have added a set of predicates $(N_1, N_2, N_3, \ldots)$ to $\mathcal{L}$ which we will read as 'called α', 'called β' etc. where 'α', 'β' are string of sounds or inscriptions[4]. I will use symbol '$!x.\varphi$' for iota-terms $\iota x.\varphi$ with only one variable x which occurs free in φ. Letting the formula φ in a description $!x.[_i]\varphi$ have a form of a conjunction of a distinguished predicate and a 1-place undistinguished predicate ($\varphi = (N_i(x) \wedge Q(x))$, e.g. 'a planet called [fɒs fə rəs]') we can define an equivalence relation in such a way that two description-world pairs belong to the same class when their descriptions designate the same object and contain the same predicate N_i. So for example, take two descriptions, 'the planet called [fɒs fə rəs]', 'the planet called [hɛs pə rəs]' (we name them γ_1, γ_2 respectively). Both descriptions γ_1, γ_2 designate in our world w, but pairs $\langle\gamma_1, w\rangle$, $\langle\gamma_1, w\rangle$ will belong to different equivalence classes because γ_1 contains predicate 'called [fɒs fə rəs]' while γ_2 contain a

[4]Arguments supporting such a view of verbs of naming can be found in (Geurts, 1997), see also (Matushansky, 2008).

different predicate 'called [hɛs pə rəs]'. This idea is represented schematically in Graph 1 below:



Graph 1

I won't go into formal details (full versions of theses definitions can be found in Appendix) and instead will just explain the key steps. In order to define an interpretation of a name-term n_i I need two functions – one which connects n_i with an equivalence class (function $\mathbb{Q}^{\leq}$) and the other which takes an equivalence class and gives the object designated by every description in the class (function F). I have presented this idea in Graph 2 below:

$$\mathbb{Q}^{\leq}(n_i) \qquad I^{\leq}_{\langle w,t\rangle}(n_i) = \mathbb{F}\left(\mathbb{Q}^{\leq}(n_i)\right)$$

$$\mathbb{F}\left(\boxed{\textit{equivalence class}}\right) = \text{☼}$$

Graph 2

In effect, name-terms designate rigidly (see (Poller, 2014) for proof) and are not synonymous with descriptions (this is exactly what a descriptive theory of reference postulates). As I have said, the idea of the predicate modification by a descriptive content of a proper name is simple: if we say that NN is doing something as NN, we mean that there is a (unspecified) way of describing NN such that NN is doing something in that way. Let φ stand

for an atomic formula or a negation of an atomic formula. We will take any name-term n to be a modifier, and write '$(\lambda x.\varphi)^i_n$' for a modified predicate abstract and '$(\lambda x.\varphi)^i_n(n)$' for a formula (a name-term and a predicate abstract modified by a name-term could form a formula iff the name-term occupying an argument place of this predicate is the same as the modifying name-term). The idea of predicate modification by a descriptive content of a proper name is represented formally as a requirement that a formula $(\lambda x.\varphi)^i_n(n)$ is satisfied in a model with respect to a world w and a time t_j iff there is a description $!y.[_j]\psi$ in the set of descriptions for the term n and the world w such that the model satisfies $(\lambda x.\varphi)^i_{\lambda y.\psi}(n)$ with respect to $\langle w, t_j\rangle$. Note that we drop a fixing-time operator $[_j]$, so our modifying descriptive content ($\lambda y.\psi$ obtained from a definite description $!y.[_j]\psi$) is sensitive to scope differences of temporal and modal operators. Take a formula with a name and a predicate. On an unmodified reading, $(\lambda x.\varphi)(n)$, all that the descriptive content of a proper name does is just pick up the reference, that is why a change of a proper name to a different but co-referential one is without significance because all you need for truth-conditions is just the name's referent and a property named by a predicate. But on a modified reading, $(\lambda x.\varphi)^i_n$, we want the descriptive content of a name to be taken into account as a circumstance of action (expressing a manner, goal, reason or time), so we make it a part of a predicate. When we say that NN is doing something as NN we understand by it that NN is doing something in a descriptive way ψ actual with respect to a time (and a world) of evaluation. So, for example, by saying (3), 'Cassius Clay was never beaten, whereas Muhammad Ali lost five fights', we convey that the greatest boxer was never beaten at a period of time when he was a boxer called 'Cassius Clay' and he lost five fights after changing his name to 'Muhammad Ali'.

It has to be said that on this account a descriptive content of co-referring genuine proper names differs only in naming predicates ('called α', 'called β'). Intuitively the difference in descriptive content between 'Superman' and 'Clark Kent' is deeper. I take expressions such as 'Superman' or 'Batman' to be pseudonyms and think that the semantics of proper names differs from the semantics of pseudonyms (cf. Katz, 2001). Let us have a closer look at pseudonyms. They are broadly understood as the names that people assume for a particular purpose (Room, 2010, p. 3). In American copyright law it is underlined that a pseudonym should be fictitious (nicknames and other diminutive forms of legal names are not considered as fictitious, cf. Copyright Office Fact sheet FL101). Usually people take pseudonyms for their activity as artists, writers, political and religious leaders, gamers, secret agents and

so on. It is a remarkable fact about pseudonyms that they can become an adopted new name whenever a person becomes mainly or solely known by their pseudonym (Room, 2010, p. 4). I take this feature of pseudonyms – to be assumed for a particular purpose – as a key feature that distinguishes pseudonyms from genuine names.

As I explained earlier, I represent genuine proper names as name-terms which designate via sets of definite descriptions of the form $!x.[_i](N(x) \wedge Q(x))$. The key difference between the formal representation of pseudonyms and names lies in representing pseudonyms as terms (called 'pseudonym-terms') which designate via sets of definite descriptions of the form $!x.[_i](N_P(x) \wedge Q_P(x))$. Every description in such a set contains a modified distinguished predicate N_P, which we read as 'named α as P' (e.g. 'called [benɪdɪkt] as a pope', 'called [ɹɑkɪt] as a hockey player'), and contains a 1-place undistinguished predicate modified by the same predicate P. By such a formal representation of pseudonyms I am trying to express their key feature of being assumed for a particular purpose. So I want the descriptive content of a pseudonym to describe an individual as doing everything with this particular purpose (e.g. 'called [bɛnɪdɪkt] as a pope', 'sends a message to the faithful as a pope', 'publishes a work as a pope' etc.). The other key feature of pseudonyms, their possibility of becoming genuine names (e.g. 'John Wayne'), when a person starts to use a pseudonym not only for a particular purpose, will not be formally represented[5][6].

[5]However, the possibility of pseudonyms to become genuine names could be formally represented. In order to represent it we could add a special operator 'only' ($*$) operating on a modifier ('only as P'). For example, at the beginning of his actor career Marion Morrison was named [dʒɒnwān] only as a film actor but from a time t_i he was named [dʒɒnwān] not only as an actor. So if we let pseudonym-terms designate via sets of definite descriptions of the form $!x.[_j](N_{*P}(x) \wedge Q_P(x))$ (containing a distinguished predicate modified by the 'only as P' modifier, N_{*P}), then from the time t_i it would be false that Morrison is named [dʒɒnwān] only as a film actor. A pseudonym-term (formal representation of 'John Wayne') is obstinately rigid and designates Morrison with respect to any time and world but from the time t_i (in our world w) it has no descriptive content which could modify a predicate (since t_i it is false that he is named [dʒɒnwān] only as an actor which in turn means that there is no description of the form $!x.[_k](N_{*P}(x) \wedge Q_P(x))$, where $i \leq k$, connected with the pseudonym-term). Letting name-terms designate via descriptions containing modified predicates we will get a name-term formally representing the name 'John Wayne' (not the pseudonym 'John Wayne') which would designate via descriptions with fixing-time operators $[_k]$, where $i \leq k$. This means that at any time later than t_i Morrison would not do anything under the pseudonym but under the name 'John Wayne'.

[6]I need to note that things are not so simple from the formal side. Imagine that Smith decided to be named 'Rocky' as a boxer. Intuitively, besides the pseudonym 'Rocky', he

As I said earlier, to avoid circularity I need two languages, $\mathcal{L}$ and $\mathcal{L}^+$ ($\mathcal{L} \subset \mathcal{L}^+$). Language L contains only variables and iota-terms as terms and language $\mathcal{L}^+$ contains additionally a set of name-terms $\mathcal{N} = \{n_1, n_2, n_3, \ldots\}$ and a set of pseudonym-terms $\mathcal{M} = \{m_1, m_2, m_3, \}$. Pseudonym-terms are interpreted in the same way as name-terms – via equivalence classes of description-world pairs, $I^{\leq}_{\langle w,t\rangle}(m_i) = \mathbb{F}\left(\mathbb{Q}^{\leq}(m_i)\right)$, which means that pseudonym-terms are obstinately rigid. A formula with a pseudonym-term is satisfied in a standard way when the referent of a pseudonym belongs to the extension of a predicate. However, a pseudonym-term has a specific feature which distinguishes it from a name-term: in all possible worlds such that a set of descriptions determining the pseudonym's reference is non-empty a pseudonym-term's referent would have a property 'P' besides a property 'called α'. Let me illustrate this specific feature by the following example. Consider four possible worlds w_1, w_2, w_3, w_4. In world w_1 Joseph Ratzinger became pope and as pope was called [benɪdɪkt sɪksti:nɵ]. On becoming pope, he visited Germany first. In world w_2 he, Benedict XVI, visited France first. In world w_3 Ratzinger failed to get into theological school and became a cigarette smuggler who always left sixteen cigarettes in his abandoned caches and as a result was known in the criminal underworld as Benedict 16. In world w_3 the police were unable to catch him but in world w_4 Ratzinger, called [benɪdɪkt sɪksti:nɵ] as a smuggler, was arrested. Formally we will have two pseudonym-terms representing Benedict XVI-a pope and Benedict-16-a smuggler pseudonyms. In all worlds such that Ratzinger is called [benɪdɪkt sɪksti:nɵ] as a smuggler he is a smuggler. Contrary to pseudonyms, proper names have no specific property besides 'called α' which is preserved in possible worlds in which a set of descriptions determining the name's reference is non-empty and that is why it is easier to construct *The Puzzle* using pseudonyms than proper names.

I defined predicate modification by a descriptive content of a proper name as a requirement that a formula $(\lambda x.\varphi)^i_n(n)$ is satisfied in a model with respect to a world w and a time t_j iff there is a description $!y.[_j]\psi$ in the set

did not take a new name 'Rocky'. Formally we will have descriptions designating Smith with 'named [ɹɑkɪ] as a boxer'-predicate and with unmodified 'named [ɹɑkɪ]'-predicate. Due to this besides a pseudonym-term designating Smith we will have a name-term designating him via descriptions containing 'named [ɹɑkɪ]'-predicate. In effect we will have name-terms which do not model any proper names from a natural language. In order to prevent such consequences we need to 'throw away' intuitively 'rubbish' descriptions containing the unmodified predicate 'named [ɹɑkɪ]' and designating Smith (see Def. VI.S(c), S(d) and Δ^*). I have elaborated upon the problem of 'rubbish' descriptions in my PhD thesis (2014).

of descriptions for the term n and the world w such that the model satisfies $(\lambda x.\varphi)^{i}_{\lambda y.\psi}(n)$ with respect to $\langle w, t_j\rangle$. It seems that there is no reason for an intended definition of modification by a descriptive content of a pseudonym $(\lambda x.\varphi)^{i}_{m}(m)$ to be different. But, as we remember, the account of modified predicates presented here is not general and the most complicated modifier is a predicate abstracted from a conjunction of formulas containing unmodified atomic predicates. Every definite description connected with a pseudonym-term contains predicates modified by some predicate P, $!x.[_i](N_P(x) \wedge Q_P(x))$, and predicate abstracted from it can't be used as a modifier. That is why a definition of predicate modification by a descriptive content of a pseudonym differs from a definition of a modification by a descriptive content of a proper name: a formula $(\lambda x.\varphi)^{i}_{m}(m)$ is satisfied in a model with respect to a world w and a time tj iff there is a description $!y.[_j](N_P(x) \wedge Q_P(x))$ in the set of descriptions for the term m and the world w such that the model satisfies $(\lambda x.\varphi)^{i}_{\lambda y.P(y)}(m)$ with respect to $\langle w, t_j\rangle$. Having no modification of a predicate by an already modified predicate (having no iteration) we cannot, for example, express that Superman is entering the phone booth dressed as a superhero (predicate 'entering' is modified by the adjunct 'dressed' which in turn is modified by the 'as'-phrase). Instead we express the fact that Superman is entering the phone booth as a superhero (predicate 'entering' is modified by the 'as'-phrase).

In the Appendix I have presented the formal semantics for modified predicates and have modeled sentences with names and pseudonyms representing C1. I have not presented a model for C2 sentences (sentences such as 'I have never made it to Leningrad, but I visited St. Petersburg last week') because they are easy to explain: intuitively such sentences are true because it is not the case that Petersburg is officially called [lenɔngræd] anymore, so you can't visit it as such. Nor have I presented a model for C3 sentences with comparative quantifiers such as (4). Intuitively in (4) we compare the cardinality of sets of fights that the greatest boxer won as Muhammad Ali and won as Cassius Clay. The cardinality of these sets differs and that is why (4) is true.

Conclusion

I treat the phenomenon of the substitution failure of co-referential proper names in simple sentences as a special case of the broader phenomenon of a lack of substitutivity between two co-referential nominal phrases. I argue that the descriptive content associated with proper names, besides

determining the proper name's reference, functions as truth-conditionally relevant adjuncts which could be used to express a manner, reason, goal, time or purpose of action. In that way a sentence with a proper name 'NN is doing something' which could be understood as 'NN is doing something as NN' (which means as-so-and-so). I propose to analyze qualifying 'as'-phrases as predicate modifiers and present a formal representation of modified predicates. According to my view, sentences from The Superman Puzzle are ambiguous between modified and unmodified readings and this assumption explains why speakers have mixed intuitions about such examples. Whereas nothing prevents the substitution of co-referential proper names on unmodified readings, the substitution of names can fail on modified readings because the different descriptive content of proper names modifies the main predicate differently, so in effect sentences can have different truth conditions. I treat names such as 'Superman' and 'Batman' as pseudonyms and argue that the semantics for pseudonyms differs in some respect to the semantics for genuine proper names. Intuitively, the key difference between names and pseudonyms lies in a pseudonyms' feature of being assumed for a particular purpose and I reflect this feature in a formal representation of pseudonyms.

Acknowledgements

I am especially indebted to Katarzyna Kijania-Placek who provided advice and encouragement at every step of the preparation of this paper. I owe many thanks also to Professor Adam Olszewski for his remarks about the formal semantics presented here. I have also benefited greatly from many discussions with members of Department of Epistemology, Institute of Philosophy, Jagiellonian University and especially with Professor Tomasz Placek, Leszek Wroński, Jacek Wawer, Juliusz Doboszewski and Maja (Sherlock) Białek. A version of this paper was presented at the PhiLang 2011 conference held at the University of Łódź (12.05.–14.05.2011) and I would like to thank all of the audience for their valuable comments on this occasion and especially to Stefano Predelli and Tadeusz Ciecierski.

This project has been supported by the National Science Centre of Poland, DEC-2012/05/N/HS1/01429.

Appendix: The formal representation of names, pseudonyms and modified predicates

The languages $\mathcal{L}$ and $\mathcal{L}^+$ are based on first-order predicate logic with identity and descriptions (I followed Fitting & Mendelsohn, 1998). I will skip all standard definitions and present the definitions that are specific for a formal representation of modified predicates, names and pseudonyms.

Let me start from the language $\mathcal{L}$ which contains only two sorts of terms, variables and iota-terms.

Definition I: The alphabet of $\mathcal{L}$

A first-order language $\mathcal{L}$ contains the following symbols: sentential connectives $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$, $\sim$; quantifiers $\exists$, $\forall$; an infinite set of individual variables $x_1, x_2, x_3, \ldots$; an infinite set of predicate constants $P_1, P_2, P_3, \ldots$, with a positive integer (an arity) assigned to each of them; identity sign $=$; the definite descriptions operator ι; the abstraction operator λ; temporal operators of past $\mathbf{P}$ and future $\mathbf{F}$; an infinite set of temporal operators $[_i]$ ('true at t_i'), where $i \in \mathbb{N}$; modal operators $\Box$, $\Diamond$; an infinite set of distinguished predicate constants $N_1, N_2, N_3, \ldots$; a set of numerical symbols for natural numbers; the left parenthesis (, the right parenthesis).

Definition II: The syntax of $\mathcal{L}$

Predicate constants and, defined below, predicate abstracts, modified atomic predicates and modified predicate abstracts are predicates of $\mathcal{L}$. An atomic predicate of $\mathcal{L}$ is any predicate constant. The notions of a formula, a term, a predicate and a free variable occurrence are defined as follows:

The notions of a variable ($R1$), a predicate constant ($R2$), an atomic formula ($R3$), $\sim \varphi$ ($R4$), $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, $(\varphi \rightarrow \psi)$, $(\varphi \leftrightarrow \psi)$ ($R5$), $\mathbf{P}\varphi$, $\mathbf{F}\varphi$, $[_i]\varphi$ ($R6$), $\Box\varphi$, $\Diamond\varphi$ ($R7$), $\forall x\varphi$, $\exists x\varphi$ ($R8$), $\iota x.varphi$ ($R9$), $(\lambda x.\varphi)$ ($R10$) are defined in a standard way;

$R11$. if Q is a 1-place predicate constant and x is a variable, then $(\lambda x.Q(x))$ is a modifier. Modifiers contain no free variable occurrences;

$R12$. if $(\lambda x.\varphi)$, $(\lambda x.\psi)$ are modifiers, then $(\lambda x.(\varphi \wedge \psi))$ is a modifier;

$R13$. if Q is a n-place predicate constant and $(\lambda x.\varphi)$ is a modifier then $Q^i_{\lambda x.\varphi}$ is n-place atomic predicate modified by $(\lambda x.\varphi)$ on ith argument place of Q (where $1 \leq i \leq n$);

$R14$. if $(\lambda x.Q(z_1, \ldots, z_n))$ is a predicate abstract and $(\lambda y.\psi)$ is a modifier, then $(\lambda x.Q(z_1, \ldots, z_n))^i_{\lambda y.\psi}$ is a predicate abstract modified by $(\lambda y.\psi)$

on ith argument place of Q (where $1 \leq i \leq n$); the free variable occurrences in $(\lambda x.Q(z_1, \ldots, z_n))^i_{\lambda y.\psi}$ are those of $(\lambda x.Q(z_1, \ldots, z_n))$;

*R*15. if $(\lambda x. \sim Q(z_1, \ldots, z_n))$ is a predicate abstract and $(\lambda y.\psi)$ is a modifier, then $(\lambda x. \sim Q(z_1, \ldots, z_n))^i_{\lambda y.\psi}$ is a predicate abstract modified by $(\lambda y.\psi)$ on ith argument place of Q (where $1 \leq i \leq n$); the free variable occurrences in $(\lambda x. \sim Q(z_1, \ldots, z_n))^i_{\lambda y.\psi}$ are those of $(\lambda x.Q(z_1, \ldots, z_n))$;

*R*16. if Q is a n-place predicate constant, $Q^i_{\lambda x.\varphi}$ is n-place modified predicate and $z_1, \ldots, z_n$ is an n-element sequence of variables, then $Q^i_{\lambda x.\varphi}(z_1, \ldots, z_n)$ is a formula in which all variable occurrences in the n-element sequence are free;

*R*17. if $(\lambda x.\varphi)$ is a predicate abstract and s is a term, then $(\lambda x.\varphi)(s)$ is a formula; the free occurrences of variables in $(\lambda x.\varphi)(s)$ are those of $(\lambda x.\varphi)$ together with those of s;

*R*18. if $(\lambda x.\varphi)^i_{\lambda y.\psi}$ is a modified predicate abstract and s is a term, then $(\lambda x.\varphi)^i_{\lambda y.\psi}(s)$ is a formula; the free occurrences of variables in $(\lambda x.\varphi)^i_{\lambda y.\psi}(s)$ are those of $(\lambda x.\varphi)^i_{\lambda y.\psi}$ together with those of s;

*R*19. nothing else is a formula, a term, a predicate, a modifier and a free occurrence of a variable.

Notational convention:

– if Q is a 1-place predicate constant and β is a modifier, then instead of 'Q^1_β' we will write 'Q_β';

– if Q is an n-place predicate constant and $(\lambda x.P(x))$ is a modifier, then instead of '$Q^i_{\lambda x.P(x)}$" we will write 'Q^i_P'.

Definition III: The semantics of $\mathcal{L}$

A varying domain first-order model $\mathfrak{M}$ for $\mathcal{L}$ is a structure $\mathfrak{M} = \langle \mathcal{D}, T, <, W, I \rangle$, such that:

– $\mathcal{D}$ is a domain function mapping pairs of possible world and time $\langle q, t \rangle$ to non-empty sets. The domain of the model is the set $\bigcup\{\mathcal{D}_{\langle w,t \rangle} : w \in W, t \in T\}$. We write $\mathcal{D}_\mathfrak{M}$ for the domain of the model $\mathfrak{M}$ and $\mathcal{D}_{\langle w,t \rangle}$ for a value of the function $\mathcal{D}$ for an argument $\langle w, t \rangle$;

– T is a set of natural numbers and $<$ ('earlier than') is a linear order defined on elements of T (a set $(T, <)$ is thought as a flow of time);

– W is a non-empty set of possible worlds;

– I is a function which assigns an extension to each pair of an atomic predicate or modified atomic predicate of $\mathcal{L}$ and a pair $\langle w, t\rangle$, where $w \in W, t \in T$, in the following way:

– if Q is an n-place predicate constant, then $I_{\langle w,t\rangle}(P) \subseteq \mathcal{D}^n_{\mathfrak{M}}$;[7]

– $I_{\langle w,t\rangle}(=) = \{\langle d, d\rangle \in \mathcal{D}_{\mathfrak{M}}\}$;

let g be a variable assignment (a mapping that assigns to each free variable x some member $g(x)$ of the model domain $\mathcal{D}_{\mathfrak{M}}$) and let $I^g_{\langle w,t\rangle}$ be a function which assigns an extension to each pair of an atomic predicate, a modified predicate or a term of $\mathcal{L}$ and a pair $\langle w, t\rangle$, where $w \in W, t \in T$:

– if x a variable, then $I^g_{\langle w,t\rangle}(x) = g(x)$ for any $\langle w, t\rangle$;

– $I \subseteq I^g$ for any g;

the notion of interpretation of terms other than variables and interpretation of modified predicates and satisfaction of formulas in $\mathfrak{M}$ are defined as follows:

*S*1. if Q is an n-place predicate constant and $y_1, \ldots, y_n$ are variables, then $\mathfrak{M}^{g\ w\ t} \models Q(y_1, \ldots, y_n)$ iff $\langle g(y_1), \ldots, g(y_n)\rangle \in I_{\langle w,t\rangle}(Q)$; the notions of satisfaction of $\sim \varphi$ (*S*2), $(\varphi \wedge \psi)$ (*S*3), $(\varphi \vee \psi)$ (*S*4), $(\varphi \rightarrow \psi)$ (*S*5), $(\varphi \leftrightarrow \psi)$ (*S*6) are defined in a standard way;

*S*7. if Q is an n-place predicate constant, P is a 1-place predicate constant and x is a variable, then $I_{\langle w,t\rangle}(Q^i_{\lambda x.P(x)}) \in \mathscr{P}(\{\langle d_1, \ldots, d_i, \ldots, d_n\rangle \in I_{\langle w,t\rangle}(Q) : d_i \in I_{\langle w,t\rangle}(P)\})$;

*S*8. if $Q^i_{\lambda x.P(x)}$, $Q^i_{\lambda y.P(y)}$ are n-place atomic predicates modified by $\lambda x.P(x)$, $\lambda y.P(y)$ on ith argument place and x, y are variables, then $I_{\langle w,t\rangle}(Q^i_{\lambda x.P(x)}) = I_{\langle w,t\rangle}(Q^i_{\lambda y.P(y)})$;

*S*9. if Q is an n-place predicate constant, x is a variable, and $(\lambda x.\varphi)$, $(\lambda x.\psi)$ are modifiers, then $I_{\langle w,t\rangle}(Q^i_{\lambda x.(\varphi\wedge\psi)}) = I_{\langle w,t\rangle}(Q^i_{\lambda x.\varphi}) \cap I_{\langle w,t\rangle}(Q^i_{\lambda x.\psi})$;

[7]This definition is taken after Fitting & Mendelsohn (1998, p. 103, Definition 4.7.3). I accept the authors' reasoning behind it.

*S*10. if $Q(z_1, \ldots, z_n)$ is an atomic formula and $(\lambda x.Q(z_1, \ldots, z_n))^i_{\lambda y.\psi}$ is a modified predicate abstract, then $I_{\langle w,t\rangle}g((\lambda x.Q(z_1, \ldots, z_n))^i_{\lambda y.\psi}) = \left\{d \in \mathcal{D}_{\mathfrak{M}} : \mathfrak{M}^{g\ \binom{d}{x}\ w\ t} \models Q^i_{\lambda y.\psi}(z_1, \ldots, z_n)\right\}$;

*S*11. if $(\lambda x. \sim Q(z_1, \ldots, z_n))$ is a negation of an atomic formula and $(\lambda x. \sim Q(z_1, \ldots, z_n))^i_{\lambda y.\psi}$ is a modified predicate abstract, then $I^g_{\langle w,t\rangle}((\lambda x.\sim Q(z_1, \ldots, z_n))^i_{\lambda y.\psi}) = \left\{d \in \mathcal{D}_{\mathfrak{M}} : \mathfrak{M}^{g\ \binom{d}{x}\ w\ t} \not\models Q^i_{\lambda y.\psi}(z_1, \ldots, z_n)\right\}$;

*S*12. if Q is an n-place predicate constant, $(\lambda x.\varphi)$ is a modifier and $Q^i_{\lambda x.\varphi}$ is an n-place modified predicate, then $\mathfrak{M}^{g\ w\ t} \models Q^i_{\lambda x.\varphi}(z_1, \ldots, z_n)$ iff $\langle g(z_1), \ldots, g(z_n)\rangle \in I_{\langle w,t\rangle}(Q^i_{\lambda x.varphi})$; the notions of satisfaction $\mathbf{P}\varphi$ (*S*13), $\mathbf{F}\varphi$ (*S*14) are defined in a standard way;

*S*15. if φ is a formula, then $\mathfrak{M}^{g\ w\ t_j} \models [_i]\varphi$ iff $\mathfrak{M}^{g\ w\ t_i} \models \varphi$; the notions of satisfaction $\Box\varphi$ (*S*16), $\Diamond\varphi$ (*S*17), $\forall x\varphi$ (*S*18), $\exists x\varphi$ (*S*19) are defined in a standard way;

*S*20. if $\mathfrak{M}^{g\ \binom{d}{x}\ w\ t} \models \varphi$ for exactly one $d \in \mathcal{D}_{\mathfrak{M}}$ then $I^g_{\langle w,t\rangle}(\iota x.\varphi) = d$; if it is not the case that $\mathfrak{M}^{g\ \binom{d}{x}\ w\ t} \models \varphi$ for exactly one $d \in \mathcal{D}_{\mathfrak{M}}$, then $\iota x.\varphi$ fails to designate at $\langle w, t\rangle$ in $\mathfrak{M}$ with respect to g; the notion of satisfaction of $(\lambda x.\varphi)(s)$ (*S*21) is defined in a standard way;

*S*22. if a term s designates at $\langle w, t\rangle$ in $\mathfrak{M}$ with respect to g and $(\lambda x.\varphi)^i_{\lambda y.\psi}$ is a modified predicate abstract, then $\mathfrak{M}^{g\ w\ t} \models (\lambda x.\varphi)^i_{\lambda y.\psi}(s)$ iff $I^g_{\langle w,t\rangle}(s) \in I^g_{\langle w,t\rangle}((\lambda x.\varphi)^i_{\lambda y.\psi})$; if a term s fails to designate at $\langle w, t\rangle$ in $\mathfrak{M}$ with respect to g, then $\mathfrak{M}^{g\ w\ t} \not\models (\lambda x.\varphi)^i_{\lambda y.\psi}(s)$.

I will use symbol '$!x.\varphi$' for a special case of $\iota x.\varphi$ terms with only one variable x which occurs free in φ. There are no free variable occurrences in $!x.\varphi$ and due to this if $I^g_{\langle w,t\rangle}(!x.\varphi)$ is defined then $I^g_{\langle w,t\rangle}(!x.\varphi) = I^{g'}_{\langle w,t\rangle}(!x.\varphi)$ for any assignments g and g'. That is why instead of '$I^g_{\langle w,t\rangle}(!x.\varphi)$' we will write '$I_{\langle w,t\rangle}(!x.\varphi)$' which should be understood as '$I^g_{\langle w,t\rangle}(!x.\varphi)$' where g is any assignment.

Now I will expand language $\mathcal{L}$ to $\mathcal{L}^+$ by adding name-term and pseudonym-terms. I will skip all syntactical and semantic definitions of $\mathcal{L}^+$ duplicating the definitions of $\mathcal{L}$ and will write below only new ones.

Definition IV: The alphabet of $\mathcal{L}^+$

A first-order language $\mathcal{L}^+$ contains all symbols of $\mathcal{L}$ with the addition of an infinite set of name-terms $\mathcal{N} = \{n_1, n_2, n_3, \ldots\}$ and an infinite set of pseudonym-terms $\mathcal{M} = \{m_1, m_2, m_3, \ldots\}$.

Definition V: The syntax of $\mathcal{L}^+$

$R1$. the same as $R1$. of $\mathcal{L}$;

$R2$. a name-term or a pseudonym-term is a term with no free variable occurrences;

$R3$. - $R12$. are the same as $R2$. - $R11$. of $\mathcal{L}$;

R13. s is a modifier, where s is a name-term or a pseudonym-term;

$R14$. - $R16$. are the same as $R12$. - $R14$. of $\mathcal{L}$;

$R17$. if $(\lambda x.Q(z_1, \ldots, z_n))$ is a predicate abstract and s is a name-term or a pseudonym-term, then $(\lambda x.Q(z_1, \ldots, z_n))^i_s$ is a predicate abstract modified by s on ith argument place of Q (where $1 \leq i \leq n$); the free variable occurrences in $(\lambda x.Q(z_1, \ldots, z_n))^i_s$ are those of $(\lambda x.Q(z_1, \ldots, z_n))$;

$R18$. if $(\lambda x. \sim Q(z_1, \ldots, z_n))$ is a predicate abstract and s is a name-term or a pseudonym-term, then $(\lambda x. \sim Q(z_1, \ldots, z_n))^i_s$ is a predicate abstract modified by s on ith argument place of Q (where $1 \leq i \leq n$); the free variable occurrences in $(\lambda x. \sim Q(z_1, \ldots, z_n))^i_s$ are those of $(\lambda x. \sim Q(z_1, \ldots, z_n))$;

$R19$. - $R21$. are the same as $R16$. - $R18$. of $\mathcal{L}$;

$R22$. if $(\lambda x.\varphi)^i_{s_j}$ is a modified predicate abstract and s_k is a name-term or a pseudonym-term, then $(\lambda x.\varphi)^i_{s_j}(s_k)$ is a formula iff $k = j$; the free variable occurrences in $(\lambda x.\varphi)^i_{s_j}(s_k)$ are those of $(\lambda x.\varphi)$;

$R23$. the same as $R19$. of $\mathcal{L}$.

Definition VI: The semantics of $\mathcal{L}^+$

Let $\mathfrak{M} = \langle \mathcal{D}, T, <, W, I \rangle$ be a model of $\mathcal{L}$. A varying domain first-order model $\mathfrak{M}^{\leq}$ for $\mathcal{L}^+$ is a structure $\mathfrak{M}^{\leq} = \langle \mathcal{D}, T, <, W, I^{\leq} \rangle$, where $I^{\leq} \restriction \mathcal{L} = I$.

Using already defined properties of $\mathfrak{M}$ (*Definition III*) we define the following sets, relations and functions.

$S(a)$: set $\Gamma_{\mathcal{L}}$
Set $\Gamma_{\mathcal{L}}$ is a set of iota-terms $!x.[_i]\varphi$ of $\mathcal{L}$. $!x.[_i]\varphi \in \Gamma_{\mathcal{L}}$ iff 1) there is a world $w \in W$ such that for every time $t \in T$ $!x.[_i]\varphi$ designates at $\langle w, t\rangle$ in $\mathfrak{M}$; 2) $\varphi = (N_i(x) \wedge (x))$ or $\varphi = (N_{i\ \lambda y.Q(y)}(x) \wedge Q(x))$ or $\varphi = (N_{i\ \lambda y.Q(y)}(x) \wedge P_{\lambda y.Q(y)}(x))$, where N_i is a distinguished predicate and P, Q are undistinguished predicates. (I will use symbols 'γ_i', 'γ_j' for members of $\Gamma_{\mathcal{L}}$)

$S(b)$: set Δ
$\Delta \subseteq \Gamma_{\mathcal{L}} \times W$. $\langle \gamma_i, w\rangle \in \Delta$ iff for any time $t \in T$ $I_{\langle w,t\rangle}(\gamma_i)$ is defined.

$S(c)$: set $\boldsymbol{D}$
$\boldsymbol{D} \subseteq \Delta$. $\langle \gamma_i, w\rangle \in \boldsymbol{D}$ iff there is a predicate $N_{i\ \lambda y.Q(y)}$ and a time $t \in T$ such that $I_{\langle w,t\rangle}(\gamma_i) \in I_{\langle w,t\rangle}\left(N_{i\ \lambda y.Q(y)}\right)$ and γ_i contains Q or N_i, where N_i is a distinguished predicate and Q is an undistinguished predicate.

$S(d)$: set $\boldsymbol{D}^*$
$\boldsymbol{D}^* \subseteq \boldsymbol{D}$. $\langle !x.[_i]\varphi, w\rangle \in \boldsymbol{D}^*$ iff $\varphi = \left(N_{i\ \lambda y.Q(y)(x) \wedge Q(x)}\right)$ or $\varphi = \Big(N_{i\ \lambda y.Q(y)}(x) \wedge P_{\lambda y.Q(y)}(x)\Big)$, where N_i is a distinguished predicate and P, Q are undistinguished predicates.

Let $\Delta^* = \Delta \setminus (\boldsymbol{D} \setminus \boldsymbol{D}^*)$.

$S(e)$: relation $\mathbb{R}$
$\mathbb{R} \subseteq \Delta^{*2}$. $\langle \gamma_i, w\rangle \mathbb{R} \langle \gamma_j, w'\rangle$ iff for any time $t \in T$ $I_{\langle w,t\rangle}(\gamma_i) = I_{\langle w',t\rangle}(\gamma_j)$ and there is either the same predicate N_k or the same predicate N_k $\lambda y.Q(y)$ in γ_i, γ_j.

Let $\Delta^*/\mathbb{R}$ be a partition of set Δ^* by equivalence relation $\mathbb{R}$ and $[\langle \gamma_i, w\rangle]_{\mathbb{R}}$ be an equivalence class from $\Delta/\mathbb{R}$.

$S(f)$: function $\mathbb{F}$
$\mathbb{F} : \Delta^*/\mathbb{R} \to \mathcal{D}_{\mathfrak{M}}$. For any $[\langle \gamma_i, w\rangle]_{\mathbb{R}} \in \Delta^*/\mathbb{R}$, $\mathbb{F}([\langle \gamma_i, w\rangle]R) = d$, where for any time $t \in T d = I_{\langle w,t\rangle}(\gamma_j)$ for any $\langle \gamma_j, w\rangle \in [\langle \gamma_i, w\rangle]_{\mathbb{R}}$.

Let $\leq$ be any well-order relation on a set $\Delta^*/\mathbb{R}$ and let $\langle \Delta^*/\mathbb{R}, \leq\rangle$ be well-ordered set.

$S(g)$: function $\mathbb{Q}^{\leq}$
$\mathbb{Q}^{\leq} : \{\mathcal{N} \cup \mathcal{M}\} \to \Delta^*/\mathbb{R}$. Function $\mathbb{Q}^{\leq}$ for an argument gives an equivalence class $[\langle \gamma_i, w\rangle]_{\mathbb{R}}$ in the following way:

– for n_1 $\mathbb{Q}^{\leq}$ gives the least element of $\langle(\Delta^*/\boldsymbol{D})_{\mathbb{R}}, \leq \cap((\Delta^*/\boldsymbol{D})_{\mathbb{R}})^2\rangle$;

– for every next element of $\mathcal{N}$ (with respect to an index) $\mathbb{Q}^{\leq}$ gives next element of $\langle(\Delta^*/\boldsymbol{D})_{\mathbb{R}}, \leq \cap\langle(\Delta^*/\boldsymbol{D})_{\mathbb{R}})^2\rangle$;

– in case there is no next element in $\langle(\Delta^*/\boldsymbol{D})_{\mathbb{R}}, \leq \cap\langle(\Delta^*/\boldsymbol{D})_{\mathbb{R}})^2\rangle$ then for a next element of $\mathcal{N}$ $\mathbb{Q}^{\leq}$ gives the least element of $\langle(\Delta^*/\boldsymbol{D})_{\mathbb{R}}, \leq \cap\langle(\Delta^*/\boldsymbol{D})_{\mathbb{R}})^2\rangle$;

– for m_1 $\mathbb{Q}^{\leq}$ gives the least element of $\langle\boldsymbol{D}^*/\mathbb{R}, \leq \cap(\boldsymbol{D}^*/\mathbb{R})^2\rangle$;

– for every next element of $\mathcal{M}$ (with respect to an index) $\mathbb{Q}^{\leq}$ gives next element of $\langle\boldsymbol{D}^*/\mathbb{R}, \leq \cap(\boldsymbol{D}^*/\mathbb{R})^2\rangle$;

– in case there are no next element in $\langle\boldsymbol{D}^*/\mathbb{R}, \leq \cap(\boldsymbol{D}^*/\mathbb{R})^2\rangle$ then for a next element of $\mathcal{M}$ $\mathbb{Q}^{\leq}$ gives the least element of $\langle\boldsymbol{D}^*/\mathbb{R}, \leq \cap(\boldsymbol{D}^*/\mathbb{R})^2\rangle$.

$S(h)$: relation $\mathbb{S}$
$\mathbb{S} \subseteq \Delta^{*2}$. $\langle\gamma_i, w\rangle\mathbb{S}\langle\gamma_j, w'\rangle$ iff $\langle\gamma_i, w\rangle$, $\langle\gamma_j, w'\rangle$ belong to the same equivalence class $[\langle\gamma_i, w\rangle]_{\mathbb{R}}$ and $w = w'$.

$S(i)$: function $h^{\leq}$
$h^{\leq} : \{\mathcal{N} \cup \mathcal{M}\} \times W \to \Delta^*/\mathbb{S}$. For any $n_i \in \mathcal{N}$, $w \in W$ $h^{\leq}(n_i, w) = [\langle\gamma_j, w\rangle]_{\mathbb{S}} \subseteq \mathbb{Q}^{\leq}(n_i)$ if there is such an equivalence class, otherwise $h^{\leq}(n_i, w)$ is undefined. For any $m_j \in \mathcal{M}$, $w \in W$ $h^{\leq}(m_j, w) = [\langle\gamma_i, w\rangle]_{\mathbb{S}} \subseteq \mathbb{Q}^{\leq}(m_j)$ if there is such an equivalence class, otherwise $h^{\leq}(m_j, w)$ is undefined.

Semantic rules $S1.$–$S20.$ of language $\mathcal{L}^+$ are the same as rules $S1.$–$S20.$ of language $\mathcal{L}$ (except of talking about $I^{\leq}$ instead of I);

$S21.$ if n_i is a name-term and $\Delta^*/\boldsymbol{D} \neq \emptyset$, then $I^{\leq}_{\langle w,t\rangle}(n_i) = \mathbb{F}\left(\mathbb{Q}^{\leq}(n_i)\right)$; if $\Delta^*/\boldsymbol{D} = \emptyset$, then n_i fails to designate in $\mathfrak{M}^{\leq}$ (at any $\langle w', t'\rangle$);

$S22.$ if m_i is a pseudonym-term and $\boldsymbol{D}^* \neq \emptyset$, then $I^{\leq}_{\langle w,t\rangle}(m_i) = \mathbb{F}\left(\mathbb{Q}^{\leq}(m_i)\right)$; if $\boldsymbol{D}^* = \emptyset$, then mi fails to designate in $\mathfrak{M}^{\leq}$ (at any $\langle w', t'\rangle$);

$S23.$ if a term s designates at $\langle w, t\rangle$ in $\mathfrak{M}^{\leq}$ with respect to g, then $\mathfrak{M}^{\leq g\ w\ t} \models (\lambda x.\varphi)(s)$ iff $\mathfrak{M}^{\leq g\ \binom{d}{x}\ w\ t} \models \varphi$, where $d = I^{\leq g}_{\langle w,t\rangle}(s)$; if a term s fails to designate at $\langle w, t\rangle$ in $\mathfrak{M}^{\leq}$ with respect to g, then $\mathfrak{M}^{\leq g\ w\ t} \not\models (\lambda x.\varphi)(s)$;

*S*24. if a term s designates at $\langle w, t\rangle$ in $\mathfrak{M}^{\leq}$ with respect to g and $(\lambda x.\varphi)^{i}_{\lambda y.\psi}$ is a modified predicate abstract, then $\mathfrak{M}^{\leq g\ w\ t} \models (\lambda x.\varphi)^{i}_{\lambda y.\psi}(s)$ iff $I^{g}_{\langle w,t\rangle}(s) \in I^{g}_{\langle w,t\rangle}((\lambda x.\varphi)_{\lambda y.\psi})$; if a term s fails to designate at $\langle w, t\rangle$ in $\mathfrak{M}^{\leq}$ with respect to g, then $\mathfrak{M}^{\leq g\ w\ t} \not\models (\lambda x.\varphi)^{i}_{\lambda y.\psi}(s)$;

*S*25. if n_k is a name-term and $(\lambda x.\varphi)^{i}_{n_k}$ is a predicate abstract modified by n_k, then $\mathfrak{M}^{\leq g\ wt_j} \models (\lambda x.\varphi)^{i}_{n_k}(n_k)$ iff there is a description $!y.[_j]\psi \in \pi_1\left(h^{\leq}(n_k, w)\right)$, such that $\mathfrak{M}^{\leq g\ w\ t_j} \models (\lambda x.\varphi)^{i}_{\lambda y.\psi}(n_k)$;

*S*26. if m_k is a pseudonym-term and $(\lambda x.\varphi)^{i}_{m_k}$ is a predicate abstract modified by m_k, then $\mathfrak{M}^{\leq g\ w\ t_j} \models (\lambda x.\varphi)^{i}_{m_k}(m_k)$ iff there is a description $!y.[_j]\psi \in \pi_1\left(h^{\leq}(m_k, w)\right)$, such that $\lambda z.Q(z)$ is a modifier of a predicate N_i $\lambda z.Q(z)$ from the description $!y.[_j]\psi$ and $\mathfrak{M}^{\leq g\ w\ t_j} \models (\lambda x.\varphi)^{i}_{\lambda y.\psi}(m_k)$.

In (Poller, 2016) I have proven that you can take a modifier 'in and out' of a predicate abstracted from an atomic formula or a negation of atomic formula, $\mathfrak{M}^{\leq g\ w\ t} \models \left(\lambda x.Q^{i}_{\lambda y.\psi}(z_1, \ldots, z_n)\right)(s)$ iff $\mathfrak{M}^{\leq g\ w\ t} \models (\lambda x.Q(z_1, \ldots, z_n))^{i}_{\lambda y.\psi}(s)$, $\mathfrak{M}^{\leq g\ w\ t} \models \left(\lambda x.{\sim}Q^{i}_{\lambda y.\psi}(z_1, \ldots, z_n)\right)(s)$ iff $\mathfrak{M}^{\leq g\, w\, t} \models (\lambda x.\sim Q(z_1, \ldots, z_n))^{i}_{\lambda y.\psi}(s)$, which is very useful in proofs (I will refer to it as Theorem). Now I will model a sentence from *The Puzzle*.

Let $\mathfrak{M}^{\leq}$ be a model of $\mathcal{L}^{+}$, $W = \{w\}$, $\mathcal{D}_{\langle w,t_1\rangle} = \{$🧍‍♂️, 🧍, [su:pəmæn], [kla:k kɛnt], [ləʊ.ɪs]$\}$, $\mathcal{D}_{\langle w,t_j\rangle} = \emptyset$ for $j \neq 1$. Let us use symbols "R'l ("reporter"), "S" ('superhero'), "P" ('talks on the phone with'), "L" ("look through the window at") instead "P_1", "P_2", "P_3", "P_4" of $\mathcal{L}^{+}$. Let use symbol "N_1" for "called [kla:k kɛnt]", symbol "N_2" for "called [su:pəmæn]" and symbol "N_3" for "called [ləʊ.ɪs]". Let $I^{\leq}$ be defined in following way:

	S	R	P	L	P_i $i \geq 5$	N_1	N_2	N_3	N_i $i \neq 1$
$I^{\leq}_{\langle w,t_1\rangle}$	{🧍‍♂️}	{🧍‍♂️, 🧍}	⟨🧍‍♂️, 🧍⟩ ⟨🧍, 🧍‍♂️⟩	⟨🧍‍♂️, 🧍⟩ ⟨🧍, 🧍‍♂️⟩	$\emptyset$	{🧍‍♂️}	{🧍‍♂️}	{🧍}	$\emptyset$
$I_{\langle w,t_i\rangle}$ $i \neq 1$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$

	$P^2_{N_1}$	P^2_S	L^2_S	L^2_R	$L^2_{N_1}$	$N_{2\ S}$
$I^{\leq}_{\langle w,t_1\rangle}$	$\emptyset$	$\langle$,  $\rangle$	$\emptyset$	$\langle$  ,  $\rangle$	$\langle$  , $\rangle$	$\{$  $\}$
$I_{\langle w,t_i\rangle}$ $i \neq 1$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$

For a predicate Q and any time $I^{\leq}_{\langle w,t\rangle}(Q_Q) = I^{\leq}_{\langle w,t\rangle}(Q)$. For predicates other than those mentioned above and $\langle w,t\rangle$, where t is any time, function $I^{\leq}$ gives $\emptyset$.

Set $\Gamma_{\mathcal{L}}$ (*Def.* VI. $S(a)$):

γ_1	$!x.[_1](S(x) \wedge N_1(x))$	γ_4	$!x.[_1](R(x) \wedge N_2(x))$
γ_2	$!x.[_1](S(x) \wedge N_2(x))$	γ_5	$!x.[_1](R(x) \wedge N_3(x))$
γ_3	$!x.[_1](R(x) \wedge N_1(x))$	γ_6	$!x.[_1](S(x) \wedge N_{2\ S}(x))$

	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6
$I^{\leq}_{\langle w,t\rangle}$						

Set Δ (*Def.* VI. $S(b)$)

$\langle\gamma_1, w\rangle$	$\langle\gamma_3, w\rangle$	$\langle\gamma_5, w\rangle$
$\langle\gamma_2, w\rangle$	$\langle\gamma_4, w\rangle$	$\langle\gamma_6, w\rangle$

Set $\boldsymbol{D}$ ((*Def.* VI. $S(c)$)

$\langle\gamma_1, w\rangle$	$\langle\gamma_2, w\rangle$
$\langle\gamma_4, w\rangle$	$\langle\gamma_6, w\rangle$

Set $\boldsymbol{D}^*$ (*Def.* VI. $S(d)$)

$\langle\gamma_6, w\rangle$

Set Δ^* (*Def.* VI. $S(e)$)

$\langle\gamma_3, w\rangle\langle\gamma_5, w\rangle\langle\gamma_6, w\rangle$

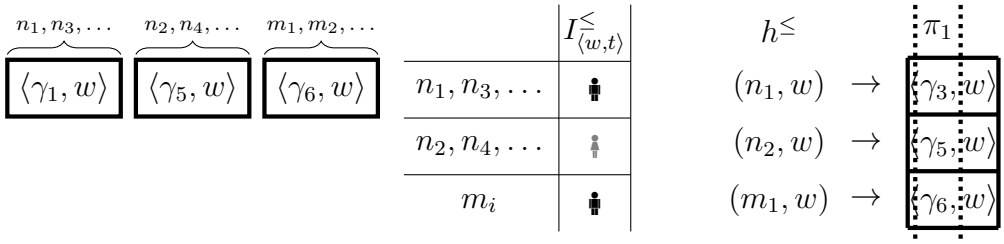
$\Delta^*/\mathbb{R}$ (*Def.* VI. $S(c)$))

$\langle\gamma_3, w\rangle$	$\langle\gamma_5, w\rangle$	$\langle\gamma_6, w\rangle$

Function $\mathbb{F}$ (*Def.*VI. $S(f)$)

$\mathbb{F}$	$[\langle\gamma_3, w\rangle]_{\mathbb{R}}$	
	$[\langle\gamma_3, w\rangle]_{\mathbb{R}}$	
	$[\langle\gamma_3, w\rangle]_{\mathbb{R}}$	

Function $\mathbb{Q}^{\leq}$ (*Def.* VI. $S(f)$) $I^{\leq}_{\langle w,t\rangle}(s) = \mathbb{F}\left(\mathbb{Q}^{\leq}(s)\right)$, where s is n_j or m_i Function $h^{\leq}$ (*Def.* VI. $S(i)$)



Let us see what the value of the following sentences is:

(a) While talking on the phone to Superman (as Superman), Lois looked through the window at Clark Kent (as Clark Kent);

(b) While talking on the phone to Clark Kent (as Clark Kent), Lois looked through the window at Superman (as Superman).

(a) $\lambda y.\left((\lambda x.P(y,x))^2_{m_1}(m_1)\right)(n_2) \wedge \lambda y.\left((\lambda x.L(y,x))^2_{n_1}(n_1)\right)(n_2)$;

(b) $\lambda y.\left((\lambda x.P(y,x))^2_{n_1}(n_1)\right)(n_2) \wedge \lambda y.\left((\lambda x.L(y,x))^2_{m_1}(m_1)\right)(n_2)$.

(a)

$\mathfrak{M}^{\leq gwt_1} \models \lambda y.\left((\lambda x.P(y,x))^2_{m_1}(m_1)\right)(n_2) \wedge \lambda y.\left((\lambda x.L(y,x))^2_{n_1}(n_1)\right)(n_2)$ (*Def.* VI.S3) iff

$\mathfrak{M}^{\leq gwt_1} \models \lambda y.\left((\lambda x.P(y,x))^2_{m_1}(m_1)\right)(n_2)$ and

$\mathfrak{M}^{\leq gwt_1} \models \lambda y.\left((\lambda x.L(y,x))^2_{n_1}(n_1)\right)(n_2)$ (*Def.* VI.S23) iff

$\mathfrak{M}^{\leq g\binom{d}{y}wt_1} \models (\lambda x.P(y,x))^2_{m_1}(m_1)$, where $d = I^{\leq}_{\langle w,t_1\rangle}(n_2)$ and

$\mathfrak{M}^{\leq g\binom{e}{y}wt_1} \models (\lambda x.P(y,x))^2_{n_1}(n_1)$, where $e = I^{\leq}_{\langle w,t_1\rangle}(n_2)$ (*Def.* S26, S25) iff

there is a description $!y.[_1]\varphi \in \pi_1(h^{\leq}(m_1,w))$, such that $\lambda z.Q(z)$ is a modifier of a predicate $N_{i\ \lambda z.Q(z)}$ from the description $!y.[_1]\varphi$ and $\mathfrak{M}^{\leq g\binom{d}{y}wt_1} \models (\lambda x.P(y,x))^2_{\lambda z.Q(z)}(m_1)$, where $d = I^{\leq}_{\langle w,t_1\rangle}(n_2)$ and

there is a description $!z.[_1]\psi \in \pi_1\left(h^{\leq}(n_1,w)\right)$, such that $\mathfrak{M}^{\leq g\binom{e}{y}wt_1} \models (\lambda x.L(y,x))^2_{\lambda z.\psi}(n_1)$, where $e = I^{\leq}_{\langle w,t_1\rangle}(n_2)$ (Theorem) iff

there is a description $!y.[_1]\varphi \in \pi_1\left(h^{\leq}(m_1,w)\right)$, such that $\lambda z.Q(z)$ is a modifier of a predicate $N_{i\ \lambda z.Q(z)}$ from the description $!y.[_1]\varphi$ and $\mathfrak{M}^{\leq g\binom{d}{y}wt_1} \models$

$\left(\lambda x.P^2_{\lambda z.Q(z)}(y,x)\right)(m_1)$, where $d = I^{\leq}_{\langle w,t_1\rangle}(n_2)$ and
there is a description $!z.[_1]\psi \in \pi_1\left(h^{\leq}(n_1,w)\right)$, such that $\mathfrak{M}^{\leq g\binom{e}{y}wt_1} \models \left(\lambda x.L^2_{\lambda z.\psi}(y,x)\right)(n_1)$, where $e = I^{\leq}_{\langle w,t_1\rangle}(n_2)$ $^{(Def.\ \text{VI.S23})}$ iff
there is a description $!y.[_1]\varphi \in \pi_1\left(h^{\leq}(m_1,w)\right)$, such that $\lambda z.Q(z)$ is a modifier of a predicate $N_{i\ \lambda z.Q(z)}$ from the description $!y.[1]\varphi$ and $\mathfrak{M}^{\leq g\binom{d}{y}\binom{d_1}{x}wt_1} \models P^2_{\lambda z.Q(z)}(y,x)$, where $d = I^{\leq}_{\langle w,t_1\rangle}(n_2)$, $d_1 = I^{\leq}_{\langle w,t_1\rangle}(m_1)$ and
there is a description $!z.[_1]\psi \in \pi_1\left(h^{\leq}(n_1,w)\right)$, such that $\mathfrak{M}^{\leq g\binom{e}{y}\binom{e_1}{x}wt_1} \models L^2_{\lambda z.\psi}(y,x)$, where $e = I^{\leq}_{\langle w,t_1\rangle}(n_2)$, $e_1 = I^{\leq}_{\langle w,t_1\rangle}(n_1)$ $^{(Def.\ \text{VI.S12})}$ iff
there is a description $!y.[_1]\varphi \in \pi_1\left(h^{\leq}(m_1,w)\right)$, such that $\lambda z.Q(z)$ is a modifier of a predicate $N_{i\ \lambda z.Q(z)}$ from the description $!y.[_1]\varphi$ and $\langle d,d_1\rangle \in I^{\leq}_{\langle w,t_1\rangle}\left(P^2_{\lambda z.Q(z)}\right)$, where $d = I^{\leq}_{\langle w,t_1\rangle}(n_2)$, $d_1 = I^{\leq}_{\langle w,t_1\rangle}(m_1)$ and there is a description $!z.[_1]\psi \in \pi_1\left(h^{\leq}(n_1,w)\right)$, such that $\langle e,e_1\rangle \in I^{\leq}_{\langle w,t_1\rangle}\left(L^2_{\lambda z.\psi}\right)$, where $e = I^{\leq}_{\langle w,t_1\rangle}(n_2)$, $e_1 = I^{\leq}_{\langle w,t_1\rangle}(n_1)$.

It is so that $I^{\leq}_{\langle w,t_1\rangle}(n_2) = d = e =$ 🧍, $I^{\leq}_{\langle w,t_1\rangle}(m_1) = d_1 =$ 🦸, $I^{\leq}_{\langle w,t_1\rangle}(n_1) = e_1 =$ 🦸.
Let $!z.[_1] \in \pi_1\left(h^{\leq}(n_1,w)\right)$ be $!x.[_1]\left(R(x) \wedge N_1(x)\right)$, γ_3. It is so that $\langle$🧍, 🦸$\rangle \in I^{\leq}_{\langle w,t_1\rangle}(L^2_R)$, $\langle$🧍, 🦸$\rangle \in I^{\leq}_{\langle w,t_1\rangle}(L^2_{N_1})$, $^{(Def.\ \text{VI.S9})}$ so, $\langle$🧍, 🦸$\rangle \in I^{\leq}_{\langle w,t_1\rangle}\left(L^2_{\lambda x.(R(x)\wedge N_1(x))}\right)$.
Modifier $\lambda x.S(x)$ is a modifier of a predicate $N_{i\ \lambda z.Q(z)}$ from any description $!y.[_1]\varphi \in \pi_1\left(h^{\leq}(m_1,w)\right)$. It is so that $\langle$🧍, 🦸$\rangle \in I^{\leq}_{\langle w,t_1\rangle}(P^2_S)$. This means that formula (a) is satisfied.

(b)

$\mathfrak{M}^{\leq gwt_1} \models \lambda y.\left((\lambda x.P(y,x))^2_{n_1}(n_1)\right)(n_2) \wedge \lambda y.\left((\lambda x.L(y,x))^2_{m_1}(m_1)\right)(n_2)$ $^{(Def.\ \text{VI.S3})}$ iff
$\mathfrak{M}^{\leq gwt_1} \models \lambda y.\left((\lambda x.P(y,x))^2_{n_1}(n_1)\right)(n_2)$ and
$\mathfrak{M}^{\leq gwt_1} \models \lambda y.\left((\lambda x.L(y,x))^2_{m_1}(m_1)\right)$ $^{(Def.\ \text{VI.S23})}$ iff
$\mathfrak{M}^{\leq gwt_1} \models (\lambda x.P(y,x))^2_{n_1}(n_1)$, where $d = I^{\leq}_{\langle w,t_1\rangle}(n_2)$ and
$\mathfrak{M}^{\leq gwt_1} \models (\lambda x.L(y,x))^2_{m_1}(m_1)$, where $e = I^{\leq}_{\langle w,t_1\rangle}(n_2)$ $^{(Def.\ \text{VI.S25, S26})}$ iff
there is a description $!z.[_1]\psi \in \pi_1\left(h^{\leq}(n_1,w)\right)$, such that $\mathfrak{M}^{\leq g\binom{d}{y}wt_1} \models (\lambda x.P(y,x))^2_{\lambda z.\psi}(n_1)$, where $d = I^{\leq}_{\langle w,t_1\rangle}(n_2)$ and
there is a description $!y.[_1]\varphi \in \pi_1\left(h^{\leq}(m_1,w)\right)$, such that $\lambda z.Q(z)$ is a modifier of a predicate $N_{i\ \lambda z.Q(z)}$ from the description $!y.[_1]\varphi$ and

$\mathfrak{M}^{\leq g\binom{e}{y}wt_1} \models (\lambda x.L(y,x))^2_{\lambda z.Q(z)}(m_1)$, where $e = I^{\leq}_{\langle w,t_1\rangle}(n_2)$ $\overset{\text{(Theorem)}}{\text{iff}}$
there is a description $!z.[_1]\psi \in \pi_1\left(h^{\leq}(n_1,w)\right)$, such that $\mathfrak{M}^{\leq g\binom{d}{y}wt_1} \models \left(\lambda x.P^2_{\lambda z.\psi}(y,x)\right)(n_1)$, where $d = I^{\leq}_{\langle w,t_1\rangle}(n_2)$ and
there is a description $!y.[_1]\varphi \in \pi_1\left(h^{\leq}(m_1,w)\right)$, such that $\lambda z.Q(z)$ is a modifier of a predicate $N_{i\ \lambda z.Q(z)}$ from the description $!y.[_1]\varphi$
and $\mathfrak{M}^{\leq g\binom{e}{y}wt_1} \models \left(\lambda x.L^2_{\lambda z.Q(z)}(y,x)\right)(m_1)$, where $e = I^{\leq}_{\langle w,t_1\rangle}(n_2)$ $\overset{(Def.\ \text{VI.S23})}{\text{iff}}$
there is a description $!z.[_1]\psi \in \pi_1\left(h^{\leq}(n_1,w)\right)$, such that $\mathfrak{M}^{\leq g\binom{d}{y}\binom{d_1}{x}wt_1} \models P^2_{\lambda z.\psi}(y,x)$, where $d = I^{\leq}_{\langle w,t_1\rangle}(n_2)$, $d_1 = I^{\leq}_{\langle w,t_1\rangle}(n_1)$ and
there is a description $!y.[_1]\varphi \in \pi_1\left(h^{\leq}(m_1,w)\right)$, such that $\lambda z.Q(z)$ is a modifier of a predicate $N_{i\ \lambda z.Q(z)}$ from the description $!y.[_1]\varphi$ and
$\mathfrak{M}^{\leq g\binom{e}{y}\binom{e_1}{x}wt_1} \models L^2_{\lambda z.Q(z)}(y,x)$, where $e = I^2_{\langle w,t_1\rangle}(n_2)$, $e_1 = I^2_{\langle w,t_1\rangle}(m_1)$ $\overset{(Def.\ \text{VI.S12})}{\text{iff}}$
there is a description $!z.[_1]\psi \in \pi_1\left(h^{\leq}(n_1,w)\right)$, such that $\langle d,d_1\rangle \in I^{\leq}_{\langle w,t_1\rangle}\left(P^2_{\lambda z.\psi}\right)$, where $d = I^{\leq}_{\langle w,t_1\rangle}(n_2)$, $d_1 = I^{\leq}_{\langle w,t_1\rangle}(n_1)$ and
there is a description $!y.[_1]\varphi \in \pi_1\left(h^{\leq}(m_1,w)\right)$, such that $\lambda z.Q(z)$ is a modifier of a predicate $N_{i\ \lambda z.Q(z)}$ from the description $!y.[_1]\varphi$ and
$\langle e,e_1\rangle \in I^{\leq}_{\langle w,t_1\rangle}\left(L^2_{\lambda z.Q(z)}\right)$, where $e = I^{\leq}_{\langle w,t_1\rangle}(n_2)$, $e_1 = I^{\leq}_{\langle w,t_1\rangle}(m_1)$.

It is so that $I^{\leq}_{\langle w,t_1\rangle}(n_2) = d = e =$ 🧍, $I^{\leq}_{\langle w,t_1\rangle}(m_1) = e_1 =$ 🧍, $I^{\leq}_{\langle w,t_1\rangle}(n_1) = d_1 =$ 🧍. Every description from the set $\pi_1\left(h^{\leq}(n_1,w)\right)$ contains the predicate N_1. It is so that $I^{\leq}_{\langle w,t_1\rangle}(P^2_{N_1}) = \emptyset$, which means that $\langle$🧍, 🧍$\rangle \notin I^{\leq}_{\langle w,t_1\rangle}(P^2_{N_1})$. This in turn means that for any description $!z.[_1]\psi \in \pi_1\left(h^{\leq}(n_1,w)\right)$ $\langle$🧍, 🧍$\rangle \notin I^{\leq}_{\langle w,t_1\rangle}(P^2_{\lambda z.\psi})$ (*Def.* VI.S9). The first part of (*b*) formula's conjunction is not satisfied, so (*b*) is not satisfied.

Bibliography

Barber, A. (2000). A Pragmatic Treatment of Simple Sentences. *Analysis*, *60*(4), 300–308.

Carlson, G.N. (1980). Such. In Gregory N. Carlson, *Reference to Kinds in English* (pp. 219–236). New York: Garland.

Clark, R. (1970). Concerning the Logic of Predicate Modifiers. *Noûs*, *4*(4), 311–335.

Crimmins, M. (1993). So-Labeled Neo-Fregeanism. *Philosophical Studies*, *69*(2/3), 265–279.

Crimmins, M. (1998). Hesperus and Phosphorus: Sense, Pretense, and Reference. *The Philosophical Review*, *107*(1), 1–47.

Davidson, D. (1967/2001). The Logical Form of Action Sentences. In *Essays on Actions and Events* (pp. 105–121). Oxford: Claredon Press.

Fitting1998 Fitting, M., Mendelsohn, R.L. (1998). *First-Order Modal Logic*. Dordrecht: Kluwer Academic Publishers.

Forbes1997 Forbes, G. (1997). How Much Substitutivity?. *Analysis*, *57*(2), 109–113.

Forbes, G. (1999). Enlightened Semantics for Simple Sentences. *Analysis*, *59*(2), 86–91.

Forbes, G. (2006). *Attitude Problems. An Essay on Linguistic Intensionality*. Oxford: Claredon Press.

Frege, G. (1892/1984). Sense and Meaning. In B. McGuiness (ed.), *Collected Papers on Mathematics, Logic, and Philosophy* (pp. 157–177). Oxford: Blackwell.

Geurts, B. (1997). Good News about the Description Theory of Names. *Journal of Semantics*, *14*, 319–348.

Jespersen, B. (2006). The Phone Booth Puzzle. *Organon F*, *13*(4), 411–438.

Katz, J.J. (2001). The End of Millianism: Multiple Bearers, Improper Names, and Compositional Meaning. *The Journal of Philosophy*, *98*(3), 137–166.

Katz, G. (2008). Manner Modification of State Verbs. In L. McNally, C. Kennedy (eds.), *Adjectives and Adverbs* (pp. 220–248). New York: Oxford University Press.

Landman, F. (1989). Groups II. *Linguistics and Philosophy*, *12*(6), 723–744.

Landman, M.L. (2006). Adjectival Anaphora. In M.L. Landman, *Variables in Natural Language*, Dissertation (pp. 44–110). University of Massachusetts.

Landman, M.L., Morzycki, M. (2003). Event-Kinds and the Representation of Manner. In N.M. Antrim, G. Goodal, M. Schulte-Nafeh, V. Samiian (eds.), *Proceedings of the Western Conference in Linguistics (WECOL) 2002*, vol. 14 (pp. 136–147). Fresno: California State University.

Link, G. (1983). The Logical Analysis of Plurals and Mass Terms: A Latticetheoretical Approach. In R. Bäuerle, C. Schwarze, A. von Stechow (eds.), *Meaning, Use, and the Interpretation of Language* (pp. 302–323). Berlin: Walter de Gruyter.

Matushansky, O. (2008). On the Linguistic Complexity of Proper Names. *Linguistics and Philosophy*, *31*(5), 573–627.

Moore, J.G. (1999). Saving Substitutivity in Simple Sentences. *Analysis*, *59*(2), 91–105.

Pitt, D. (2001). Alter Egos and Their Names. *The Journal of Philosophy*, *98*(10), 531–552.

Poller, O. (2014). Formal Representations of Proper Names in Accordance with a Descriptive Theory of Reference. *Polish Journal of Philosophy*, *8*(1), 37–52.

Poller, O. (2016). The Descriptive Content of Names as Predicate Modifiers. *Philosophical Studies*. DOI doi:10.1007/s11098-016-0801-5.

Predelli, S. (2004). Superheroes and Their Names. *American Philosophical Quarterly*, *41*(2), 107–123.

Recanati, F. (2012). *Mental Files*. Oxford: Oxford University Press.

Rini, A.A. and Cresswell, M.J. (2012). *The World-Time Parallel: Tense and Modality in Logic and Metaphysics*. New York: Cambridge University Press.

Room, A. (2010). *Dictionary of Pseudonyms*, 5th ed. Jefferson, North Carolina, London: McFarland & Company.

Saul, J.M. (1997). Substitution and Simple Sentences. *Analysis*, *57*(2), 102–108.

Saul, J.M. (2000). Did Clinton Say Something False?. *Analysis*, *60*(3), 255–257.

Saul, J.M. (2007). *Simple Sentences, Substitution and Intuitions*. Oxford: Oxford University Press.

Siegel, M.E.A. (1994). Such: Binding and the Pro-adjective. *Linguistics and Philosophy*, *17*, 481–497.

Szabó, Z.G. (2003). On Qualification. *Philosophical Perspectives*, *17*, 385–414.

Wood, J.L. (2002). Much about Such. *Studia Linguistica*, *56*(1), 91–115.

Zimmermann, T.E. (2005). What's Two Names?. *Journal of Semantics*, *22*, 53–96.

Originally published in *Studia Semiotyczne* 30/2 (2016), 11–44.

STUDIA SEMIOTYCZNE—ENGLISH SUPPLEMENT, vol. XXIX (2017), 159–179
ISSN 2451-2745
DOI: 10.26333/stsen.xxix.08

Krzysztof Posłajko[1]

Semantic Deflationism, Public Language Meaning and Contextual Standards of Correctness[2]

Abstract The paper aims at providing an argument for a deflationary treatment of the notion of public language meaning. The argument is based on the notion of standards of correctness; I will try to show that as correctness assessments are context-involving, the notion of public language meaning cannot be treated as an explanatory one. An elaboration of the argument, using the notion of ground is provided. Finally, I will consider some limitations of the reasoning presented.

Keywords deflationism, meaning, public language, grounding

Introduction

The aim of this paper is to provide an argument for the idea that the notion of public language meaning should be treated in a deflationary fashion. The argument is based on the notion of contextual standards of correctness. The argument is also intended as a partial response to the recent objection to deflationism raised by Stephen Schiffer.

First, I am going to provide a working definition of deflationism as applied to semantic notions in general, and elucidate the notion of public language meaning. Then, I am going to present Schiffer's objection to deflationism. After that I'll introduce the notion of standards of correctness, which plays

[1]Jagiellonian University, Faculty of Philosophy.
E-mail: krzysztof.poslajko@uj.edu.pl

[2]The work on this paper was funded by National Centre of Science, grant no. UMO-2014/15/B/HS1/01928, led by dr Paweł Grabarczyk from University of Lodz. An earlier version of this paper was presented at Context, Cognition and Communication Conference in Warsaw. I am grateful to Paweł Banaś and two anonymous referees for helpful remarks.

a central role in the debates on normativity of meaning. This notion will be crucial in the argument for deflationism. Afterwards, I am going to present an elaboration of my argument, which is based on Kit Fine's notion of ground. The next part will be devoted to an attempt to extend the line of argumentation to a broader range of phenomena, namely actions based on understanding public language expressions. In the final part of the paper I will show some limitations to the line of reasoning presented and an objection to it.

1. Definition of deflationism

"Deflationism" is a term that has come to stand for a great variety of philosophical views in different areas of inquiry. In the present paper I shall be interested in deflationism understood as a theory concerning the status of semantic notions. The paradigmatic case here is truth – most of the early versions of deflationary theories were created as theories of this concept. However, nowadays it is not uncommon to see deflationism applied to other semantic notions, like meaning or reference (many examples of such approaches and their criticisms are discussed in a collection of essays in Gross & Tebben & Williams, 2015).

In what follows I will treat deflationism as a generic position, in principle acceptable to any semantic notion. I will also assume (although it might be controversial) that one can be semantic deflationist locally – i.e. only with respect to one of the semantic notions, while adopting a substantial theory to other such notions, or remaining neutral with respect to them.

Traditionally, deflationism about "S" was understood as the idea that there is no such property as S-hood; in the paradigmatic case of truth deflationism has been for a long time defined as a theory which simply denied that there is such property as being true (see e.g. Strawson, 1950).

This definition however, has led to some serious theoretical difficulties. The critics pointed out that it is not feasible to claim that "true" does not refer to a property. If there is no such property as being true, the argument goes, then it would seem that the predicate "is true" would have an empty extension (see e.g. Wright, 1992). But this would amount to the claim that there are no true sentences whatsoever. Such a preposterous claim has never been intended to be made by deflationists, who clearly have not intended to develop a version of error theory concerning truth: i.e. a position which would claim that every sentence ascribing a value of truth to another sentence is

false. Such a radical theory would most likely be inconsistent. Moreover, an error theory about truth is irreconcilable with our folk intuitions about the truth predicate (as the folk clearly believe that there are indeed true sentences), and most deflationists has been keen on trying to preserve our commonsense intuitions about the truth predicate.

This argument seems to generalize other semantic notions. It would be quite bizarre to claim that there is no such property as reference, if this thesis were to be interpreted as the claim that nothing ever referred to anything. Again, this purely hypothetical position could be accused of inconsistency and it goes without saying that our everyday beliefs about reference defy it.

Considerations of these sorts have led most deflationists to the admission that "being true" indeed is a property, albeit only in minimal sense (see e.g. Horwich, 1998a). But once deflationists agree that there is such a property as "truth", they owe us a clear conception of what makes their theory different from substantial theories of truth, which also claim that truth is a property.

The standard move here is to claim that what is characteristic to deflationism is the claim that although "truth" (or other semantic notion) denotes a property, the property in question is not a substantive one. This idea forces the deflationist to propose a criterion for distinguishing between substantive and non-substantive properties. This distinction is usually explicated in terms of explanatory relevance (see e.g. Horwich, 1998a, Edwards, 2013): substantive properties are thought to be those which are relevant in explanations of phenomena. Conversely, if a certain property is not relevant in explanatory practice then we should treat it as non-substantive.

Generalizing the conclusion from the previous paragraph, we might say that the point of controversy between a deflationist and a proponent of substantial theory concerning a given semantic term "S" is whether one should treat the property S as explanatory relevant. The deflationist is the one who claims that although we might say that there is something like property S, and that the predicate "S" has a non-empty extension, there is no deep theoretical job for the property; we might use it as a purely logical device, but that is about it. The anti-deflationist, on the other hand, argues that the notion in question is needed for theoretical purposes and that invoking the property in question helps us to genuinely explain important phenomena.

2. Public language meaning and deflationism

The semantic notion I wish to focus on is public language meaning. I do not intend to provide a formal definition of this notion, but only an informal elucidation, which I hope would be sufficient for the purposes of the paper. The basic idea is that when we ask about a meaning of an expression, we might distinguish between an idiolectical meaning, which is specific to a single user, and a meaning which is bestowed upon an expression by a wider community. A well-worn example of that distinction is that of malapropisms: in one's idiolect it might be well true that "eventually" has the same meaning as "actually", whereas in standard public English of the educated Anglo-American population these two words have clearly different meanings.

It is important to note that the distinction between idiolectical and public language meaning is different from the well-known distinction between a speaker's meaning and semantic meaning (see Kripke, 1977). The speaker's meaning is the meaning of an expression as used by a speaker in a given context. Both idiolectical meaning and public language meaning are subspecies of semantic meaning, i.e. both are meanings which are systematically attributable to the expression, rather than being properties of individual use. The difference between them lies in the fact that while idiolectical use traces patterns of use of a single individual, the public one traces the patterns of use of a wider community.

Deflationism about public language meaning is, then, a position according to which public language meaning is a non-substantive notion, which means that public language meaning should not be treated as relevant in explanations of any interesting phenomena.

Such a position has already been presented in the literature as the minimalist interpretation of the so-called "sceptical solution" of the sceptical problem presented by Saul Kripke in his reading of Wittgenstein remarks of following a rule. As it is well known the "Kripkensteinian" sceptic questioned whether there is any fact that determines the meaning of any expression (Kripke, 1982). In the most famous example, the sceptic claimed that nothing determines whether the symbol "+" means "plus", as there is no way of excluding the possibility that this symbol denotes some other function, say quus, i.e. a function which yields the same results as plus when the arguments are lower than an arbitrary number, and yields 5 in other cases.

Kripke presents his owns answer to this challenge and dubs it the "sceptical solution". Its basic claim is that there indeed are no facts deter-

mining meaning of an expression, but nonetheless there is room for claiming that certain semantic attributions are correct (according to the communal standards of correctness). Moreover, Kripke (1982, p. 86) claims that redundancy theory of truth can be applied to such semantic attributions. Some authors have claimed that these remarks are best understood as putting forward a deflationary account of meaning (see e.g. Byrne, 1996, Kusch, 2006, Wilson, 1998). According to them, what Kripkenstein really denies is the existence of robust/substantial semantic fact, but his sceptical solution allows for the existence of deflationary/minimal semantic facts.

In my paper I am not going to engage in exegesis of Kripke's notoriously vague arguments. Instead, I am about to provide a new argument for the deflationary approach to public language meaning. Although this argument is not intended to be an interpretation of Kripke, it draws some inspiration from his work. But before I do that, I shall look at Stephen Schiffer's recent critique of deflationism, which, I believe, provides an important dialectical setting for the discussion.

But before we proceed, it is important to note that the definition of deflationism about public language meaning presented in this paper is distinct from the perhaps more well known version of deflationism about meaning, namely the one promoted by Paul Horwich. The basic tenet of Horwich's theory is that we should explain meaning in terms of a basic acceptance property, which in turns allows us to claim that the primary meaning of an expression is a concept expressed by it (Horwich, 1998b, p. 45–46). Such a theory allows him to describe meanings in terms of biconditionals like "dog" means DOG, where DOG is a concept.

The difference between the approaches preferred by Horwich and the one I want to pursue here, stems, in my opinion, from the fact that I am primarily interested in the metaphysical status of the putative meaning-property, i.e. in a meta-semantic problem, while Horwich wants to provide a (first-order) theory of meaning. In my opinion causal relevance is a plausible candidate for a test to distinguish between those properties which are to be treated in a deflationary manner and those properties which are to be treated in a strongly realist fashion. Thus, focusing on a question of whether public language meaning is causally relevant is the best way to answer the meta-semantic question whether we are dealing with a "substantial property" here.

Paradoxically enough, on my definition Horwich's theory turns out to be a non-deflationary one, as he admits that there are indeed such things as meaning-properties; moreover, these properties have an underlying nature

and those "underlying natures of meaning-properties are basic regularities of use, explanatorily fundamental generalizations about the circumstances in which words occur" (Horwich, 1995, p. 356). The complaint that Horwich's theory of meaning is not in fact deflationary was raised by Huw Price (1998, p. 111). Price claimed that Horwich's use of "deflationism" is significantly different when it is applied to "meaning" and not to "truth".

Horwich seems to be mostly interested in providing a philosophical account of meaning and less with its metaphysical implications. I have no intention of providing such an account. Hence, I will not try to engage in the debate, whether, for example, it is possible to characterize meaning of an expression using a biconditional modeled on the T-schema of Tarski (see Horwich, 1998b, p. 14). On my take, deflationism is a negative metasemantic thesis and is not inherently tied to any account of meaning. I think adopting such definition of public language deflationism, although it might differ from other accounts of that position, is theoretically fruitful, as it takes public meaning deflationism a special case of a generic position. This position deserves critical attention, especially in the light of the recent Schiffer's critique.

3. Schiffer's worry

In his *Deflationist Theories of Truth, Meaning, and Content* (forthcoming) Stephen Schiffer argues against the idea that semantic notions should be given deflationary treatment. He opposes "radical deflationism" – a strictly defined, globally applicable, hypothetical position (based, to an extent, on Harty Field's views). So, my defense of a modest, local, deflationism about public language meaning is not in direct opposition to Schiffer's work.

However, Schiffer's arguments provide, in my opinion, a substantial challenge to all forms of deflationism – even those more locally focused. The line of argumentation provided in his paper is fairly intricate, but a quite simple, yet powerful argument can be extracted from it, and it is a one that all deflationists should take seriously.

According to Schiffer, the deflationists claim that it is possible to explain human language-related behaviour without referring to any semantic properties. But for Schiffer such a project is unrealistic. In everyday practice it is perfectly normal to explain human behaviour by resorting to semantic properties of the expressions used. And there is no principled reason to treat such explanations as defective (apart from general worries about causal exclusion, which Schiffer dismisses). The other worry is that deflationism

provides us with no workable alternative to the common practice; in those cases when we normally appeal to semantic properties, we do not have any practically applicable methods of explaining human behaviour other than the ones that we actually employ, and these are laden with semantic properties.

In my paper I am going to focus on the first part of the challenge. At first glance, the Schiffer's worry might look as a pretty weak argument, as it relies on description of *de facto* existing explanatory practices (and who can be sure that our actual, pre-scientific ways of explaining phenomena are above criticism?). But I think this is indeed quite a powerful argument. It aims to show that deflationism is an under-motivated position, as it provides no reason to think that appeals to the semantic in explanatory practice are defective. And that the alternative – namely substantial theories of the semantic – have had the advantage of already being tried in working practice.

To counter this line of reasoning, a deflationist must present an argument which would provide motivation for their position. In what follows I am going to provide an argument which aims to counter the intuitions Schiffer's argument wanted to induce. At the heart of Schiffers argument seems to be that we must treat meaning as substantial as it plays an important causal-explanatory role in psychological explanations. So, my argument would aim to show that, appearances to the contrary, the notion of public language meaning plays no important role in causally explaining human-language related behaviour. This argument will be based on the notion of standards of correctness, which is central to the contemporary debates about normativity of meaning.

4. Standards of correctness

The idea that expressions of public language have conditions of correct use is central to the debates on normativity of meaning. The claim that meaning is normative, once considered obvious (see e.g. Kripke, 1982, McDowell, 1984), has been subject to many criticisms more recently (see e.g. Hattiangadi, 2006, Glüer & Pagin, 1999). At the heart of the debate lies the question whether meaning is normative in a "strong" or "philosophically interesting" sense. There are, of course, many ways one might precisify the normativity claim and different arguments has been waged for and against the normativist thesis (for a recent defense of normativism see e.g. Whiting, 2007).

What is curious about the debate is that both sides of it seem to agree on a basic intuition that there is something like correct and incorrect use of language. (The only prominent philosopher who had qualms about this

thesis was probably Davidson (2005)). The basic idea is quite straightforward: when a user of a public language uses a certain expression, we, as other users of the same language are entitled to judge this use as correct or incorrect, according to the semantic norms of the language in question. This shared assumption is central to the argument that I am going to present.

This fact might well be regarded as constitutive of notion of public language meaning. It is only possible to claim that the phenomenon we are dealing with is indeed a public language if there are standards of correct use associated with it. (This might, at least partially, explain why Davidson ended up claiming that there is no such thing as language (Davidson, 2005) – as he denied that there is such a thing as standards of correct use).

The observation that there are standards of correcteness might seem relatively trivial and not particularly relevant to the deflationism debate. But in my opinion this is a crucial fact. I claim that standards of correctness of any public language are context-involving, in the sense that they include factors external to the current, internal state of the speaker. In order to appraise someone's use of language we must look beyond what is, at the moment of an utterance, going on in the head of the speaker.

I should try to argue for this claim by way of analogy. It is widely accepted in the literature on normativity of meaning that semantic norms can be compared to institutional ones (this is accepted by normativist and anti-normativists alike). Hattiangadi (2006, p. 63) made an analogy with a theme park where there is a rule stating that only kids of a certain height can go on the ride. This example serves Hattiangadi to criticize normativism; she focuses her attention on the observation that in this case the height of a child is a purely naturalistic characteristic of her/him.

Still, this example can be used to highlight a different aspect of the correctness condition thesis. If we look only at the purely internal characteristic of the child then we are in no position to judge whether she or he is of the "right height" – we might only be able to provide with a purely physical description of the child. In order to get to know whether we are dealing with a case that is "correct" according to the rules that are in force, we must look at other factors than the subject itself (in this case we must, obviously, look at the regulations of the theme park).

I think this observation generalizes to all cases of institutional correctness. Whenever there are some institutional rules in force (no matter whether trivial or serious) that allow us to judge certain actions as correct or not, the judgment must be based on comparing an agent's actions with the rules in question. These rules must refer to at least some factors external to the

agents which are being judged as acting correctly or not. This is crucial because otherwise it would be impossible for one to act incorrectly. And this very possibility of incorrectness is something which makes the very assessment possible. For if it were impossible for one to act incorrectly, the very notion of correctness in this context would have no sense. Wittgenstein has famously described such a situation as the one in which "whatever is going to seem right to me is right. And that only means that here we can't talk about «right»" (Wittgenstein, 1953, § 258).

So, if we agree that language rules are akin to institutional rules, then we should also admit that the linguistic norms are in a way external to the speakers. This observation seems obvious for every proponent of externalist theories of meaning. According to externalism the facts that determine correctness conditions for language use are external to the speaker, as they include either social facts (as social externalist, of the type of Burge proposed) or facts about the kinds of things that are in the physical surroundings of the language user (as natural-kind externalists of the Putnamian kind assert).

But for internalists the thesis that standards of semantic correctness are context involving might not be that obvious. For the theorist of internalist inclination wants to explain meaning purely in terms of psychological states of the speakers. However, in my opinion, even someone who believes that language meaning is determined by purely psychological factors must admit that some factors, which are relevant to the assessment of expressions, are in a way external to the agent using language in a certain situation. This is due to the fact that even an internalist wants to maintain the distinction between correct and incorrect use.

Therefore, what such a theorist needs is a distinction between the psychological state of the user while making an utterance and a psychological state, which determines the correctness conditions for the use. I think that this distinction is implicit in most internalistic theories of meaning. Usually, it is introduced by postulating a time difference between the act of use (and associated psychological states) and meaning-determining psychological states. Put simply, the internalist usually claims that meaning is determined by meaning-intentions, which are made previous to the acts of use. What serves as the standards of correctness for my current use are the meaning intentions which I have made in the past. And this time difference allows for that correct/incorrect distinction. For I might presently act in a different manner than I intended in the past.

This time distinction lies at the heart of many of the examples Kripke offers in his discussion on rule-following. Even the most famous "plus" –

"quus" example plays on the fact that my present use of the symbol "+" might deviate from my previous intention: what makes the putative subject, and who uses the symbol in a quus-like way err in the fact that she is unfaithful to her previous intention to use the symbol in a "standard" way.

The general picture of meaning which I am going to presuppose in the next sections of the paper might be then described as broadly externalist. This broad conception of externalism includes many conceptions of what might have been traditionally described as internalist. The position I am putting forward is to a great extent a schematic one – it insists only on the claim that whenever we want to ascribe a public meaning to a certain expression we must implicitly accept that there is something external to the occurrent psycho-functional state of the speaker which is to be taken as a standard of correctness. But this schematic theory remains neutral to the question of what these standards of correctness are in particular cases. Even on a more general level the conception presupposed in this paper remains neutral to the question whether say, Kripke's conception of natural kind term is the correct theory of reference for terms like "gold" or "water". So, the phrase "contextual standards of correctness" should be treated as a sort of theoretical place-holder, whereby various externalist theories of meaning might fill in different ways.

It might also be useful to distinguish between two general kinds of broadly externalist approaches[3]: according to the first it is the standards of correctness that are external to the speaker. In the other what is external to the speaker are the norms stating what contextual elements are to be taken into consideration when assessing a certain utterance. My position is obviously externalist in the first sense; I claim that for each language use there is something "outside the head" of the speaker with which his use is to be compared. The second sort of externalism claims that the norms of correctness are constituted externally – for example by the societal agreement. This version of externalism seems to be plausible when we theorize about public language, but, as I want to stress, this is not an assumption which is needed in order for the argument of the next section to be sound.

5. The argument

In this section I am going to provide an argument to the effect that public language meaning should be given a deflationary treatment, which is based on the premise that meaning involves contextual standards of correctness.

[3]I am grateful to an anonymous referee for drawing my attention to this point.

First, let me introduce some definitions. I will use "E" to denote an expression fact, i.e. the fact that a particular language user used a given expression at a particular occasion. "M" will be used to denote the fact (or the totality of facts) that determines the public language meaning of the expression used in E.

This description of "M" is deliberately vague, as I want to be as non-committal as possible with regard to the different theories of meaning. I shall not argue that either of the numerous theories of meaning is correct or not. Rather, I should use "M" as a sort of place-holder, which denotes states postulated by whichever theory of meaning comes out right.

I should understand "P" as a psychological-functional state that is causally responsible for the agent's utterance in E. This again is a vague description, as the exact description of what is the character of states that are causally responsible for linguistic utterances is still largely unknown. Still, even if P and M are only vaguely characterized, I think it is possible to try to establish certain truths about relations between them.

My hypotheses concerning the relations between these two kinds of facts are the following:

First and foremost, M cannot be identified with P. This is because, as the second claim goes, P is doing all the causal-explanatory work, and M does none.

The transition from the second thesis to the first one is fairly straightforward. If two putative facts differ when it comes to their causal-explanatory role, then we might safely assume that we are indeed dealing with two different facts and any attempt to identify them would be mistaken.

So, the crucial task is to justify the second thesis that it is P that is relevant in providing causal explanation to the Es, while M is not. Again the first part of the task seems relatively easy: Ps are causally relevant to the linguistic production by definition. So, what needs to be justified is the claim that Ms are not.

This can be supported by an observation that in a given situation we can keep P fixed, while M changes – a subject can be in the same current, internal state (and thus produce the same expressions) and mean different things, depending on context.

Let us consider the plus vs. quus example. Kripke (1982, p. 8) invites us to consider a counterfactual situation in which the "+" sign really means quus not plus. In such a situation a subject might have used the sign "+" in a way as we actually use it: namely, as if the symbol denoted addiction. Such a person would commit an error according to the standards which are in

force in her public language, but it is quite possible that her internal psycho-functional state at the time of making the utterance would be identical with someone's from our linguistic community which would perform a standard addition.

So, there is a possibility of there being two persons who are identical with respect to their linguistic beahaviour and psycho-functional causes of it, but whose expressions in the relevant situations have different meanings. This is obviously a direct consequence of the claim that the standards of correctness are contextually determined. In this case these contextual factors might include community agreement, previous intentions or objective mathematical facts.

Examples might be multiplied. The famed Burge's thought experiment of arthritis can be used to prove a similar point – depending on the contextual factors, the patient who classifies any pain in the tight as "arthritis" (Burge, 1979) might be treated as using the world correctly or not. In our actual community this is of course an incorrect use, but it is not hard to imagine a different community, in which "arthritis" is used in a way the discussed subject uses it. In all such cases it is not the psycho-functional state of the speaker that influences the meaning, but rather external, contextual factors.

Now, the converse situation is also possible. We might easily imagine two subjects whose utterances have the same meanings (so we have identical Ms), while their Ps are different. This is because, once we allow for the possibility of error, we must admit that the psycho-functional which lead to correct and erroneous linguistic use are indeed quite different (the psycho-functional state which leads one to use "+" as a quus-denoting symbol is obviously rather different from one which leads the "normal" user who uses "+" to simply add). But we must admit that when we have two uses of the same expressions made in the context of the same public language then they have the same meaning, even though one of the uses is incorrect. So, in the example discussed, the fact that someone uses the "+" symbol incorrectly does not (in a normal situation) change the public language meaning of the symbol. It still means plus, even if an erratic user uses it in a quus-like pattern.

This is important, because the whole idea of meaning involving correctness conditions leads inevitably to the conclusion that even wildly erring usage does not change the meaning of the expression used. If this was not so, we would lose the possibility of error: if deviation from the standards of correctness led to the alteration of meaning of expressions then it would be impossible to use linguistic expressions incorrectly.

These considerations prove that there are two kinds of possible situations. In the first type, there are two possible subjects, who share the same P-state, but there expressions have different meanings. In the situation of the second type, there are possible subjects who use the same expression with the same meaning but their psycho-functional states differ. So, it is possible to have the same Ps with different Ms and *vice versa*. But, what is crucial, in both of the situations it is the change in Ps which causes the change in behaviour. Change of meaning, which is not accompanied by the change in the psycho-functional state of the user, has, in itself, no causal impact on linguistic behaviour. Additionally, in order to cause the change in the use, the change in psycho-functional state does not need to be accompanied by a change in meaning.

This shows clearly, in my opinion, that we should take psycho-functional states rather than public language meanings of the expressions to be the causes of linguistic behaviour. But, this conclusion seems enough to justify a deflationary approach to public language meaning (as defined in section 2).

It is important to note that the argument presented is not a straightforward variant of the causal-exclusion argument, which has been extensively discussed in the philosophy of mind. I do not intend to claim that only physical or "basic" properties are causally relevant. On the contrary, I am open to the possibility that psycho-functional characteristics might not be reducible to the physical ones. The contrast between the psycho-functional properties and the semantic ones is not the contrast between "ontological levels". It is rather a matter of granularity of descriptions. When we describe expressions as correct and incorrect, we describe them taking a broader context into account, while description of psycho-functional states abstracts from the contextual elements.

6. Elaboration

In this section, I am going to present a simple metaphysical model which is an elaboration of the argument presented above and which would aim at explaining two things. Firstly, why we should treat public language meaning in a deflationary way, and secondly, why we treat public language meaning as explanatory in our everyday practice.

The model will use Kit Fine's notion of ground (Fine, 2001). According to Fine, the relation of grounding is a basic metaphysical one: if A grounds B, then B obtains in virtue of A. This relation eschews a straightforward definition as it is metaphysically basic. Still, it can illuminate the ques-

tion of realism. According to Fine, we should treat certain propositions in a realist fashion when they are either metaphysically basic and factual (Fine, 2001, p. 17) or are grounded in some basic and factual propositions. But if there are no real grounds for certain propositions then we might claim that these are not factual propositions. It is an important feature of Fine's proposal that it makes room for grounding relation between non-factual elements as well (Fine, 2001, p. 17). So, when we are dealing with a non-factual proposition we might make hypotheses about which constitutive elements of a given proposition make it non-factual.

Let us try to apply the notion of ground to the phenomena discussed in this paper. My hypothesis is that P (a psycho-functional state of the speaker) is a partial ground for M (the meaning of the expression used). But it is important to bear in mind that it is only partial ground. The other fact that partially grounds M-facts concerns the contextual factors which serve as correctness conditions.

Both P-facts and contextual factors might be treated as factual. Yet I claim that Ms are non-factual, even though they are grounded in Ps and contextual factors, which are both factual. So, in order to support the claim that Ms are non-factual, it is necessary to postulate a non-factual element which also grounds them.

In my opinion such a non-factual element is the relation between E and, by extension, P and the contextual standards of correctness. This is a relation of "being a standard of correctness for". Even if we take that the terms of said relation to be perfectly factual in Fine's sense, there seems to be little motivation to take the relation itself to be factual. That a certain element of the context is taken to provide the benchmark of correct use of a given expression seems to be an utterly conventional matter. Moreover, this relation seems to have no causal-explanatory import.

Such relations might be taken to be grounding the putative meaning-facts. The fact that a certain expression means something is rooted in the relation which binds the expression to the standard of correctness; that my utterance of the symbol "+" in a given context, means plus, is grounded by the relation of this utterance to the standard of correctness (say, my previous meaning-intention). Should this relation be different, the meaning of my utterance would be different as well.

This model, in my opinion, allows us to elevate the worry presented by Schiffer. The question was: Why do we treat public language meaning as explanatory, when it is not, at least according to the deflationists? And the answer is: Because public language meaning is partially grounded by

something that really plays the causal explanatory work – namely the psycho-functional states of the speakers. And it seems to be quite normal that in everyday explanations we treat "broader" facts as explanations, especially in the situations when we lack access to the "fine-grained" facts. We do not normally know anything about the psycho-functional states of ourselves and fellow language users, so we resort to explanations in terms of public language meaning. In doing so, we tacitly assume that these meaning facts are somehow rooted in "something in the head" of the speaker, which is the genuine cause of their behaviour.

Schiffer might be perfectly right that we have no realistic alternative to meaning-based explanations, which could be used in everyday practice. Explanations resorting to the psycho-functional states might be practically unattainable. However, I do not think this is a fatal objection to deflationism, as it is meant to be a metaphysical position regarding the nature of semantic predicates, not a practically applicable theory.

7. Understanding and action: remaining problems

The arguments presented in this paper might be easily attacked for not being general enough. They might be said to show that public language meaning is not relevant in providing causal explanations of linguistic production. But this is not the only possible use of meaning in explaining human behaviour. To my knowledge, none of the existing inflationist theorists of meaning have treated the role of meaning in explaining linguistic production as the main reason for treating the public language meaning in substantial fashion. But, I believe, the argument I presented against treating public language meaning as explanatory in the context of language production can be applied to other cases, where one might want to treat this concept as relevant in causal psychological explanation.

For Schiffer the central observation speaking in favour of treating the semantic properties as substantial was that we explain action by reference to the fact that a person understands a certain expression in a certain way. When applied to the problem of status of public language meaning, Schiffer's insight might be understood as follows: the fact of understanding, which explains some action of some objects, stems from the fact that the expressions mean something in a given public language. Thus, public language meaning plays an important role in explaining behaviour.

This might sound terribly complicated, but the phenomenon is in fact quite easy and commonplace. For example, when we want to know why

the children in the classroom sat down it is perfectly legitimate to say that they did so because the teacher said to them "asseyez vous" and this phrase means "please sit down" in French.

Meaning can thus enter the explanation of action differently than by explaining linguistic production. But if this kind of explanation is a legitimate one, then deflationism about public language meaning is in serious trouble, because it turns out that meaning is actually needed in explanation of some language-related phenomena.

The question then arises whether the argument presented in the previous sections of this paper can also be generalized as to cover the cases in which meaning is used to explain actions which stem from understanding expressions of public language. In what follows I will try to formulate such an argument.

The reasoning will be similar in spirit to the one presented in section 6, and it will also be based on the notion of standards of correctness. This is because understanding public language expressions is subject to the assessment in terms of correctness, in a similar way linguistic production is. A subject might understand a certain public language expression correctly or not, and this observation seems to be central to the notion of a public language meaning.

The possibility of error is clearly visible when we focus on understanding expressions made in a foreign public language – it is quite common for people who are not native speakers to misunderstand expressions of a given language. But even within our own native language there is always the possibility of understanding an expression differently than in a way prescribed by standards of correctness operating in said public language.

In such situations we should distinguish between the meaning of the public expression used and the act of understanding, which is a psycho-functional state of the speaker. Again, I should argue that these two facts must be considered as distinct. Moreover, it is the psycho-functional state of understanding that is causally responsible for the actions of the users. I shall try to prove this using an example.

Take Tom, a native English user who is quite ignorant of the vernacular used to denote different kinds of seafood. He sees "crayfish" on a restaurant menu and understands this expression as a name of a kind of fish dish. As he strongly dislikes fish, he decides not to order. However, he is a great fan of seafood, and if he were to believe that the dish is a kind of seafood he would most likely order it. But misinterpreting the expression of his own public language prevented him from acting on his preferences.

Again, the examples might be multiplied, but I guess it is not necessary. It is quite easy to note that the fact that misunderstandings are possible leads to the conclusion that the subjective act of understanding must be taken to be distinct from the public language meaning and whatever constitutes it. And when it comes to its relevance in causal explanations it is the subjective act of understanding which might reasonably claim priority; in the situations when one misunderstands the expressions the subject will act on her subjective psycho-functional state.

A public language meaning that is not mediated by the act of subjective understanding seems to have no direct influence on the actions of the subject. Therefore, the deflationary position concerning this notion seems justified, even if we take into account the phenomenon of actions based on understanding expressions of public language.

8. The over-generalization challenge

The argument presented in this paper can be also attacked for being, as it were, too general. The problem is that the line of argumentation presented in section 6 can be quite easily extended to other domains, in which the conclusion might seem implausible[4].

The basic idea of the argument, to put it briefly, was that meaning is, at least partly, determined by contextual factors. And given this fact, we might observe that meaning cannot be thought to be causally responsible for actions of language users. This is because the mere change in contextual factors does not, by itself, change the behaviour. The change in linguistic action is brought upon by the change in the functional-internal state of the speaker. And this state cannot be identified with meaning.

The worry is that a similar argument can be produced in all contexts of institutional norms. Every fact that an institutional norm is in force is analogous to linguistic meaning in that respect that in involves a relation of the subject whose behaviour is governed by a given norm to some contextual standard of correctness. Let us take a standard example of institutional norms – road traffic rules. When we say that someone acted incorrectly according to the traffic rules, we compare the subject's beahviour with some contextual standard. The same behaviour can be described as correct or not depending on the context in which the assessment is made (driving on the left is correct in the UK but wrong in continental Europe and so on).

[4]I am very grateful to an anonymous referee for raising this point.

Again, as in the case of meaning, the change in context does not, by itself, change behaviour. This leads to the conclusion that institutional rules are not explanatorily relevant (when contrasted with psycho-functional states of the subjects). Consequently, we should claim that all institutional rules are to be treated in a deflationary fashion. But this might seem counterintuitive: it would mean that institutional rules do not have any impact on the actions of people.

There are two possible ways to answer this challenge. The first would be a direct rebuke to show that the analogy does not hold – that there is a deep theoretical difference between the way language operates and the way other institutional discourses and facts do. This would be a strategy of containment of deflationism to the linguistic realm. Unfortunately, I do not see how this could be done. The argument presented above in no way is based on peculiar characteristics of language. The argument relied only on the fact that the norms of language are institutional ones. So, if the argument is correct it should indeed be generalized to all forms of institutional rules. Thus, I must bite the bullet and say that my argument leads to global deflationism about the institutional.

This is certainly an implausible conclusion for many theorists, but it might be noted that a deflationary approach to institutional and legal fact would not be an entirely groundless position and there are philosophers who seem to endorse it. One recent example of such an approach might be found in Thomasson (2013). For her, deflationism about the institutional and the legal is a welcome consequence of her globally deflationary approach to metaphysics. In the context of the philosophy of law, James Coleman (1995) argued that Dworkin's views could be viewed as a form of deflationism.

Of course, the question whether the deflationary account of the legal and the institutional in general is an acceptable one is extremely puzzling, and answering it would require a separate paper. But I want to stress that even though deflationism in these areas might seem intuitively implausible, it seems to be a real option on the theoretical level.

9. Conclusions

The fact that the notion of public language meaning essentially involves standards of correctness allowed me to present an argument to the effect that public language meaning is not causally explanatory, even though in everyday use we might treat it as it actually were. This reasoning applies both to explanations of linguistic production and actions based on understanding,

which we would normally make using the notion of public language meaning. This conclusion gives us a strong reason to accept the deflationary account of public language meaning, as the point of contention between deflationists, as defined in this paper, and proponents of a substantial theory regarding a given semantic notion is whether the notion in question is relevant in causal explanations of the phenomena.

However, the line of argumentation presented in this paper has some serious limitations. Firstly, it might well be the case that there are some other ways in which the notion of public language meaning comes into explanatory practice, and the kind of argument which has been developed above has no application to them. So, the argument might be, at best, treated as a shift of burden of proof. The adherent of substantial theory of public language meaning must, in response to it, show which phenomena need explanation in terms of this notion.

The other limitation of the argument is that it is, in a way, a local one. It does not extend to semantic notions other than public language meaning, which leaves open the question of whether, for example, idiolectical meaning or semantic properties of propositional attitudes should not be treated in a substantive fashion. Moreover, the line of reasoning presented here relies on the notion of an internal psycho-functional state which is assumed to explain the language-related behaviour. But nothing in what has been argued for suggests that this kind of state cannot have semantic properties. But if this is so, the only upshot of this paper would be that substantial semantic properties must be located on the psychological level and not on the level of public language meaning.

This might seem too modest a conclusion for a deflationist, as it leaves room for a substantial account of at least some semantic properties. So, we are left with the question; is it possible to mount a more general argument which would show that semantic properties are not substantial ones? This is an extremely complicated issue and I will not try to resolve it in the present paper. Still, I believe that even a partially applicable argument can shed some light on the immensely intricate problem of deflationism.

Bibliography

Burge, T. (1979). Individualism and the Mental. *Midwest Studies in Philosophy*, *4*, 73–121.

Byrne, A. (1996). On Misinterpreting Kripke's Wittgenstein. *Philosophy and Phenomenological Research*, *56*, 339–343.

Coleman, J. (1995). Truth and Objectivity in Law. *Legal Theory*, *1*, 33–68.

Davidson, D. (2005). Nice Derangement of Epitaphs. In *Truth, Language, and History* (pp. 89–108), Oxford: Clarendon Press.

Edwards, D. (2013). Truth as a Substantive Property. *Australasian Journal of Philosophy*, *91*(2), 279–294.

Fine, K. (2001). The Question of Realism. *Philosophers' Imprint*, *1*(1), 1–30.

Glüer, K., Pagin, P. (1999). Rules of Meaning and Practical Reasoning. *Synthese*, *117*, 207–227.

Gross, S., Tebbenm N., Williams, M. (eds.) (2015). *Meaning without Representation*. Oxford: Oxford UP.

Hattiangadi, A. (2006). Is Meaning Normative?. *Mind and Language*, *21*, 220–240.

Horwich, P. (1995). Meaning, Use and Truth. *Mind*, *104*, 355–68.

Horwich, P. (1998a). *Truth*. Oxford: Oxford UP.

Horwich, P. (1998b). *Meaning*. Oxford: Oxford UP.

Kripke, S. (1982). *Wittgenstein on Rules and Private Language*. Cambridge, Mass: Harvard UP.

Kripke, S. (1977). Speaker's Reference and Semantic Reference. In P.A. French, T.E. Uehling Jr, H.K. Wettstein (eds.), *Studies in the Philosophy of Language* (pp. 255–296), Minnesota: University of Minnesota Press.

Kusch, M. (2006). *A Sceptical Guide to Meaning and Rules*. Chesham: Acumen.

McDowell, J. (1984). Wittgenstein on Following a Rule. *Synthese*, *58*, 326–363.

Price, H. (1998). What Should a Deflationist about Truth Say about Meaning? *Philosophical Issues*, *8*, 107–115.

Schiffer, S. (forthcoming). Deflationist Theories of Truth, Meaning, and Content. In B. Hale, C. Wright, A. Miller (eds.), *A Companion to the Philosophy of Language*, 2nd ed., Oxford: Willey-Blackwell.

Strawson, P. (1950). Truth. *Proceedings of the Aristotelian Society*, *24*, 129–156.

Thomasson, A. (2013). Fictionalism versus Deflationism. *Mind*, *122*, 1023–1051.

Whiting, D. (2007). The Normativity of Meaning Defended. *Analysis*, *67*, 133–140.

Wilson, G. (1998). Semantic Realism and Kripke's Wittgenstein. *Philosophy and Phenomenological Research*, *58*, 99–122.

Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell.

Wright, C. (1992). *Truth and Objectivity*. Oxford: Oxford UP.

Originally published in *Studia Semiotyczne* 31/1 (2017), 45–66.

STUDIA SEMIOTYCZNE—ENGLISH SUPPLEMENT, vol. XXIX (2017), 180–200
ISSN 2451-2745
DOI: 10.26333/stsen.xxix.09

Natalia Karczewska[1]

Evaluative Adjectives – an Attempt at a Classification[2]

Abstract In my paper, I propose a certain classification of evaluative expressions. I hypothesize that the basic criterion to distinguish between evaluative and descriptive terms is the faultless disagreement test. Next, I discuss a few kinds of phenomena which seem to render this distinction dubious: context–sensitivity, vagueness and using descriptive terms to express evaluative judgments. Further, I investigate Ch. Kennedy's proposal (2016) according to which gradable adjectives can express two kinds of subjectivity (one being generated by vagueness and one stemming from evaluativity). I modify this account by postulating another sub-class of subjective adjectives ("experiential adjectives") which are not subjective due to vagueness and which are not evaluative either as they do not necessarily encode any valence. I propose a linguistic test to identify these expressions. Finally, I check where my classification of adjectives places the predicate of personal taste "tasty". I suggest that "tasty" is both evaluative and experiential and additionally, it carries a condition of its own use, that is the information that can be used to positively assess the taste of something. This, I argue, makes it similar to thin evaluative terms as it carries no descriptive component at all.

Keywords faultless disagreement, subjectivity, evaluativity, predicates of personal taste, vagueness

---

[1]University of Warsaw, Faculty of Philosophy and Sociology.
E-mail: natalia.karczewska@gmail.com

[2]This work was supported by the National Science Centre (Poland) under Grant agreement no. 2013/11/N/HS1/04817.

0. Introduction

Some types of utterances are more controversial then others. Consider the following exchange:

(1) A: Donald Trump is handsome.
B: Not at all! Donald Trump isn't handsome!

This disagreement is definitely more difficult to solve, and exchanging arguments is likely to last longer than in the case of the conversation below:

(2) A: Donald Trump is 180 cm tall.
B: No, he's only 175 cm tall.

It is quite easy to determine the truth values of the utterances in (2). In order to do that, we could, for example, measure the US President's height with a tape measure. It is, therefore, immediately obvious that at least one of the speakers has said something false. In the case of (1), the issue is much more complicated. There are no widely available tools with which one could measure the degree to which somebody is handsome. Moreover, there is no consensus about which property or set of properties is connoted by the expression "handsome" and so, there is no consensus about what somebody has to be like to be considered handsome.

One of the important differences between (1) and (2) is that the former, and dialogues similar to it, can be classified as cases of faultless disagreement, at least at first sight. Faultless disagreement is a situation which invokes two conflicting intuitions in an observer: that the participants of the conversation are disagreeing in some fundamental way (that they are in disagreement or in conflict) and that neither of them has made a mistake in their utterance (since both have expressed a proposition they were justified in expressing)[3]. It seems that this feature which dialogues like (1) have, is connected with the

[3]The faultless disagreement problem has received many formulations in the last couple of years. Only some of them are neutral with respect to a semantic theory. E.g., Max Kölbel (2003) understands faultless disagreement as a situation consisting in a pair of contradictory propositions expressed in utterances or beliefs (p and not-p). Some contextualists believe that this formulation does not allow them to satisfactorily explain faultless disagreement and therefore decide to adopt its different construal (see, e.g., Lopez de Sa, 2015, Marques, 2016). Here, I am neutral about the definition of faultless disagreement and I restrict myself to characterising its pre-theoretical intuition in the most general way possible.

fact that the speakers express their opinion or assessment, that is, instead of attributing some easily verifiable, speaker-independent quality to something or someone, they attribute value to it. They do this by using expressions belonging to a certain class, namely the class of evaluative expressions.

In the present paper, I will take a closer look at evaluative expression and sketch their classification, hoping that a better understanding of their semantics will prove helpful in the discussion on faultless disagreement. I will be mostly interested in the distinction between evaluative and descriptive expressions. I will try to show that this division is not exhaustive. In particular, it does not allow for an adequate description of predicates of personal taste such as "tasty" which have been at the center of attention of the faultless disagreement theorists for the last couple of years but which have not received many analyses in terms of lexical semantics. Finally, I will propose my own characterization of these terms.

1. Evaluative and descriptive terms

As I have mentioned above, the lexical items used by speakers to express evaluation are evaluative terms (evaluatives). It should be stressed that this group comprises not only adjectives and adverbs ("beautiful", "beautifully"), but also nouns ("genius", "athlete"). There are reasons to think that such verbs as "to wail" (in the sense of singing badly), "to scribble" or "to rumble" contain a lexically encoded element of evaluation too and therefore can be considered evaluative as well.

Evaluative terms are to be distinguished from descriptive terms. The latter are such expressions that denote objective qualities of objects. For the purposes of the present paper, I will define descriptive terms as those which do not encode evaluation in their lexical meaning, for instance: "170 cm tall", "green" or "a hundred years old". Using such expressions is usually aimed at describing the world by invoking intersubjective properties of objects.

(3) John is 1.8 m tall.

"is 1.8 m tall" is a descriptive term since, firstly, it is true if and only if John is 1.8 m tall no matter who utters (3) and secondly, there is no element of evaluation encoded in its lexical meaning. The division between evaluative and descriptive terms is not as sharp as the comparison between (1) and (3) would suggest. Many terms are composed of both descriptive and evaluative information.

1.1. Thin and thick terms, all-purpose terms

In the present paper, I assume the distinction between thin and thick terms, which is widely recognized in ethics (lately also in aesthetics and philosophy of language), formulated by Bernard Williams (1985). An expression is thin with respect to evaluation when it is used only to express evaluation and no description. Some paradigmatic thin terms are: "good", "bad", "beautiful", "right", "wrong", etc. Saying about some act that it is good does not tell us anything more than that the speaker values it positively (or claims that someone, everyone or some society recognizing a certain standard would value it positively). Thick evaluatives, on the other hand, are, e.g., "cruel", "mean", "nice", "generous", "lewd" or "chaste". They carry both evaluative and descriptive content. To call somebody cruel means to say that this person tends to inflict suffering on others for her own pleasure and to value her negatively because of that. In contrast, to call someone bad, does not inform the hearer in what way this person is bad. The element of evaluation contained in a thick evaluative becomes obvious when we notice that the same behaviour might be evaluated in different ways and therefore, two different evaluatives would be used to talk about it. For example, a person who gives a lot to other people without expectation of reward might be evaluated positively in virtue of that and called generous or evaluated negatively and called profligate. Similarly with the pairs: "courageous"/"reckless", "audacious"/"assertive", etc.

Thin terms can in a sense function as thick ones. This is a property of a class of terms sometimes called *all-purpose terms*. For example, in:

(4) John is a good father.

"good" does not serve its usual role of a thin evaluate since the speaker of (4) as it communicates something different than that John is a father and that John is good. Instead, "good" functions as an all-purpose term. To say that someone is a good father it to say that he has some particular traits, e.g., that he spends a lot of time with his kids, that he is affectionate, and so on, which relates to the positive assessment of this person. To say that a knife is good is to say that it is sharp and durable, which translates to a positive assessment. The semantics of such expressions is so underdetermined that it is often uncertain what properties exactly the speaker wishes to invoke. Calling someone a good father meant something else two hundred years ago in a rural area than it means nowadays in a big European city. Simplifying it a bit, we can agree, however, that if two speakers occupy the same cultural

and linguistic context, then by using the term "good father" they will refer to some, if not most, of the same substantive properties.

I take the faultless disagreement test to be the basic diagnostic which can be used to determine if a given expression is evaluative or purely descriptive. If two speakers can faultlessly disagree about a property to which a given expression refers, it should be considered evaluative – this can be seen in dialogue (1) above. There are, however, exceptions from this rule. For this, and other reasons, the intuitive distinction between descriptive and evaluative terms may raise doubt. In the next section, I will describe these qualms, try to find their source and decide whether they are grounded.

1.2 Context sensitivity

There are adjectives which may seem problematic for the intuitive evaluative-descriptive distinction, e.g. "tall", "old" or "worn off". They are problematic for a few reasons. One might be tempted to argue that the difference between (5) and (6) below:

(5) John is 1.80 m tall.

(6) John is tall.

consists in the fact that (5) constitutes an objective description of reality, while (6) carries something more – the speaker of (6) *decides* that John is tall enough to be classified as a tall person. Should we say then that "tall" contains some evaluative component? I believe the explanation should be sought elsewhere.

It may happen that uttering (6) in one context, we mean something else than in another. John does not always have to be at least 1.80 to be considered tall. For instance, if he is a six year old, it is enough if he measures 1.15 m for (6) to be true. On the other hand, If John is a basketball player, his 1.80 might not be enough for (6) to be true. The property of the expression "tall" described here is common to all relative gradable adjectives (Kennedy, 2007).

The fact that specifying truth conditions of a sentence containing a gradable adjective requires taking into account a comparison class ("for a six year old", "for a basketball player"), explains why some descriptive adjectives may seem slightly less subjective than others. The speaker must choose a comparison class and recall an approximate threshold from which the object is considered to have the property in question. Deciding whether the value

that the object exemplifies is higher or equal to the threshold, is, to a certain extent, a judgment call. It is not, however, the kind of evaluation that is delivered while attributing properties expressed by evaluatives. Nothing from the lexical meaning of "tall" suggests that it is either good or bad to be tall. Moreover, when the speakers coordinate, i.e., agree on the specific comparison class that is invoked in the context, the disagreement about whether someone is tall will turn out to be an easily solved misunderstanding.

1.3 Vagueness

Another reason why relative gradable adjectives are problematic for our distinction is that according to some philosophers (Wright, 1997, Barker, 2002, Richard, 2004) in some situations speakers may faultlessly disagree about whether someone is tall.

"Tall", like other vague predicates, has clear classes. For instance, a seventeen year old who measures 1.90 m is definitely tall but his 1.20 m peer – is definitely not tall. In the case we invoke a different comparison class (say, of NBA basketball players), then a 2.10 m tall sportsman will be considered definitely tall, while his 1.90 m teammate will be described as average. What is important, even once we establish what comparison class we are invoking in a given conversation, the threshold above which we will consider someone tall will still fall on a slightly different point of the scale for each speaker. Nevertheless, most competent speakers will agree about the applicability of the predicate for the vast majority of cases. That is why the disagreement about whether some 2.10 m tall basketball player is tall or not, will not seem faultless.

The situation is different, according to some, when the speakers are disagreeing about borderline cases of adjectives, for instance when the term "tall" is used in the following case: (Reggie Bullock – a player from the Phoenix Suns – is 201 cm tall which puts him in the mean of height of NBA players):

(7) A: Reggie Bullock is tall.

B: No, Bullock is not tall.[4]

[4]If the Reader is not keen on agreeing that (7) seems faultless at all, I suggest to imagine a context in which A and B are looking at Reggie Bullock (and not e.g., just talking about him on the phone), in order to rule out the situation in which one of them makes a mistake stemming from the fact that she does not remember how tall the player is, or has false information about it.

Chris Barker (2013) argues that in a situation in which the disagreement is about a borderline case of a vague predicate, then – since there is no higher linguistic authority that would be able to decide who is right – neither of the speakers has made a mistake in his or her use of the predicate. Deciding whether 201 cm of height qualifies a basketball player to the extension or the anti-extension of the predicate "tall" is not possible in the framework of the used language – this language is simply not precise enough.

In the introduction of the present paper, I assumed that it is the evaluative terms and not the descriptive ones that generate faultless disagreement. Does it mean that agreeing that such dialogues as (7) are faultless disagreements implies that "tall" and other vague predicates are evaluative? Or the other way round: that we can faultlessly disagree whether someone is tall proves that we should reject this assumption? There is a third possibility and this is the one I would like to argue for. Namely, there are reasons to consider the faultless disagreement whose faultlessness stems from the speaker's *right* to express her opinion on the extralinguistic issues to be a different phenomenon than the faultlessness stemming from vagueness. I provide detailed argumentation for this claim in another paper (2016). Here, I will restrict myself to recall Christopher Kennedy's (2016) argument in a similar spirit.

1.3.1 Two kinds of subjectivity according to Kennedy

Kennedy (2016) analyses the distinction between subjective and non-subjective expressions. He proposes two tests which serve to decide to which of these two groups a given term belongs. The former has already been mentioned here: it is the answer to the question of whether this term generates faultless disagreement. The latter consists in finding out whether the expression can be embedded under the "to find x P" construction. Kennedy believes that the fact that there are expressions that pass the first test for subjectivity, that is, they generate faultless disagreement but they do not pass the other one, suggests that there is more than one kind of subjectivity.

The first test is passed by some uses of gradable adjectives, such as exemplified by the example (7). That would suggest that all such expressions are subjective (at least when they refer to borderline cases). Kennedy reminds us, however, that gradable adjectives are vague only in the positive form. Their comparative forms are no longer such. Thus, the adjective "tall" in the sentence "Adam is tall" is vague, but in "Adam is taller than Andrew",

it isn't. It stems from the fact that in a positive form the adjective expresses the property of having some degree of height which is higher or equal to some threshold (for a given comparison class). This threshold should be equated with a standard of significance which does not only depend on the facts connected with height, but also from a subjective decision of the speaker (that is the decision about what value of height is significant enough – in other words – if it sufficiently exceeds the average to call someone tall).[5] In its comparative form, on the other hand, it expresses the property of having a degree of height which is above the degree of height possessed by someone else (here: by Andrew).

Kennedy admits that a disagreement about borderline cases of vague predicates can be faultless because of their vague character. However, there are still faultless disagreements for which no such vagueness-related account can be given, for example, when the speakers are using comparative forms of adjectives. It happens when this adjective is evaluative, e.g.:

(8) A: Chocolate cake is tastier than strawberry cake.

B: That's not true. Strawberry cake is tastier.

It seems, therefore, that for some expressions, it is not vagueness that is the source of subjectivity, but their lexis. Therefore, for the subjective adjectives in the comparative form, Kennedy applies another test which consists in embedding the sentence containing such an expression under the construction "to find x P"[6]. "Find" is a verb used to express subjective judgments, so if used with an objective term, the sentence will sound infelicitous:

[5]Kennedy here invokes Delia Graff Fara's account of vagueness (Graff Fara, 2000).

[6]According to Kennedy's account, gradable adjectives denote functions from objects to degrees, so the interpretation of the adjective "tall" in type-theoretical semantics looks the following way: $[\![\text{tall}]\!] = \text{height}_{\langle e,d\rangle}$. The general schema of the morpheme of the adjective (g) in its positive form is: $[\![\text{POS}]\!] = \lambda g\langle e,d\rangle\lambda x.g(x) \geq \text{stnd}(g)$, where 'stnd($g$)' means the standard appropriate to the measurement expressed by g. The schema of the adjective morpheme in its comparative form is: $[\![_{\text{COMP}}]\!] = \lambda g\langle e,d\rangle\lambda d_{\text{than}}\lambda x.g(x) \geq d_{\text{than}}$. In order for the adjective g to say something true of the object x, the degree to which x has the property a which g denotes must be higher than the degree expressed by the component marked as dthan. Thus, putting together the adjective and the name of the object ("Adam" – a type e expression) gives us the measure to which this adjective possesses the property denoted by this adjective.
The denotation of "tall" in its positive form looks like that: $[\![\text{POS}]\!]([\![\text{tall}]\!]) = \lambda x.\mathbf{height}(x) \geq \text{stnd}(g)$.
Its denotation in its comparative form: $[\![\text{COMP}]\!]([\![\text{tall}]\!]) = \lambda d_{than}\lambda x.\mathbf{height}(x) \geq d_{\text{than}}$.

(9) Adam finds the cake tasty.

(10) #Adam finds Anna tall.

(11) Adam finds the chocolate cake tastier than the strawberry cake.

(12) #Adam finds Anna taller than Kate.

Sentences (9) and (11) sound felicitous and therefore pass the test, while (10) and (12) are more difficult to accept. Kennedy obtains the following results:

	Adjective in positive form		Adjective in comparative form	
	Faultless disagreement	find	Faultless disagreement	find
Tall	+	-	-	-
Tasty	+	+	+	+

It turns out that vagueness on its own is a sufficient condition for faultless disagreement, but it is not a necessary one (see (8)). Vagueness is not a sufficient condition for felicity under "find". Being evaluative, on the other hand, is a sufficient condition for both being acceptable under "find" and for faultless disagreement. Kennedy infers from this that even though faultless disagreement is always generated by subjective adjectives, this subjectivity comes in two kinds. One of them has to do with vagueness and is characteristic for every relative gradable adjective used to refer to a borderline case – let us call it subjectivity$_{\mathrm{V}}$. The other kind of subjectivity is shown by evaluative expressions due to their lexical meanings – subjectivity$_{\mathrm{E}}$.

This distinction is reflected in the semantics of these expressions. In line with the classical type theory, non-subjective adjectives, such as "metal" are of $\langle e, t\rangle$ type (the function takes an object, e.g., "Eiffel tower" of type e and yields the meaning of the sentence – its truth value of type t). According to a contextualist proposal by Kjell Johann Sæbø (2009), which Kennedy adopts, subjective adjectives, such as "tasty" are of type $\langle e, \langle e, t\rangle\rangle$, where the argument (type e) is the judge: a person or a group of people who consider something tasty. Sæbø's idea explains it in this way why non-subjective expressions sound odd when embedded under "find" – there is a type mismatch.

Let us sum up the results of our considerations until now. It turns out that not only evaluative expressions generate faultless disagreement.

Also descriptive adjectives (more precisely – relative gradable adjectives in their positive form) under some circumstances can generate it, but it is faultlessness of a different kind, coming from another type of subjectivity, viz. subjectivity$_{\mathrm{V}}$. The test which helps rule out descriptive subjective expressions consists in checking their acceptability under "find". It has to be stressed that if, after Kennedy, we believe that distinguishing two kinds of subjectivity is justified, the demarcation line between objective and subjective terms does not run where the demarcation line between descriptive and evaluative terms does. The set of subjective expressions contains both evaluative terms (subjective$_{\mathrm{E}}$, e.g. "beautiful"), as well as descriptive terms (subjective$_{\mathrm{V}}$, e.g. "tall"). Joanna Odrowąż-Sypniewska (2016) has a different but interesting view of this issue. She believes that all vague terms are objective in their clear cases (in this sense that they have semantic type $\langle e, t\rangle$), but become subjective in borderline cases (they change their type to $\langle e, \langle e, t\rangle\rangle$ by being enriched with a judge parameter).[7]

It seems that despite some doubts provoked by the issues connected with vagueness and context-sensitivity, we have been able to defend the intuitive division line between descriptive and evaluative expressions, which is designated by the results of the faultless disagreement test. This is not the end of the complications, however. It turns out that propositions containing only descriptive expressions may turn out to be evaluative. First, it happens sometimes that descriptive words are used to express evaluative judgments. Second, there are polysemous terms which have both descriptive and evaluative senses, while the former seems more basic. These two possibilities are discussed in sections 1.4 and 1.5 below.

1.4. Polysemous expressions

Some expressions have both a descriptive sense and an evaluative sense. By this we do not merely mean that they have different semantic components within one sense (which is the case for the thick evaluatives, such as "generous" mentioned above), but that they have two distinct senses. Consider the following example:

(13) This sauce is heavy.

[7] Odrowąż-Sypniewska does not use type theory in her text, but it seems that the interpretation presented above is in accord with her account.

"Heavy" has a non-subjective sense relating to something measurable. This sense would be expressed in (13) if the speaker was holding a weighty saucepan. "Heavy" also has a subjective sense – to call sauce heavy means saying something about its culinary value. According to Kennedy (2016), the "find" test allows for the identifying of only this subjectiveE sense of the term. Sentence (14) sounds good only if the adjective is used evaluatively:

(14) I find this sauce heavy.

On the other hand, embedding "heavy" under "find" when the former is used descriptively is no longer acceptable, e.g. when I am looking at a bag of potatoes on which the label says "10 kilos" (15) or when an airline employee puts my luggage on the belt (16):

(15) #I find this bag heavy.

(16) #Our airline finds this bag heavy.

The reason for the infelicity of these utterances is, as it has been mentioned in the previous section, the fact that gradable adjectives which are subjective only because they are vague (in my terminology they are $\text{subjective}_{\text{V}}$), do not pass the "find" test:

(17) #I find John tall.

It seems, however, that (18) is acceptable from the perspective that the bag does not weigh so much but I am really tired:

(18) I find this bag heavy.

Similarly, when I'm holding one bag in each hand and I know that they both weigh the same but one of them has a very uncomfortable handle, I can felicitously say:

(19) I find this bag heavier than the other one.

"Heavy" in (15) seems to have a meaning closer to the one in (18) and (19), that is, it operates on the scale of weight. Where does the difference in results of the test come from then? Presumably (18) expresses yet another sense of the adjective, namely, the subjective and sensual experience of heaviness. An airline cannot have this sort of experience, just like I cannot

have the experience that John is tall. I can only *consider* John tall just like an airline, as a collective subject, based on its regulations, can consider the bag too heavy.

This diagnosis seems to be in accordance with Kennedy's account, according to which such adjectives as "long", next to the objective sense, can have an evaluative – one can say – qualitative sense:

(20) I find the flight from Chicago to Hong-Kong longer than the flight from Hong-Kong to Chicago.

The utterer of (20) may be aware of the fact that the flights last the exact same time but want to express his subjective experience of their duration. That might happen, for instance if he flies business class to Hong-Kong for vacation and the flight seems long to him since he can't wait to get there. When he is coming back to Chicago, on the other hand, the journey seems short because it brings him closer to his duties (and on top of that, he only gets to fly economy class).

It seems that invoking polysemy of some expressions does not sufficiently explain the problem of untypical uses of the expressions mentioned above. In particular, it does not explain why "heavy" in utterance (13) would mean something different than in (18). This, I believe, stems from the fact that Kennedy seems to identify evaluativity of expressions with them being subjective$_{\mathrm{E}}$. Clearly, the adjective "heavy" used with reference to a bag means (in a strong sense) something else than it does when referring to a sauce, namely, in the first case it says something about weight, and in the second – about texture. Moreover, according to Kennedy, in the first case, it has a quantitative/dimensional[8]/objective sense, while in the other – a qualitative/evaluative/subjective sense. However, "heavy" in example (18) proposed above is only a special use of the term in the first sense. In my opinion, it has a qualitative, subjective and dimensional (and not evaluative) sense. This is why I would like to propose a modification – or enrichment – of Kennedy's account which consists in distinguishing two subclasses of subjective adjectives among the adjectives which are subjective$_{\mathrm{E}}$.

[8] "Dimensional/evaluative" is a vocabulary adopted by Bierwisch (1989) which Kennedy uses. It does not entirely correspond with our terminology ("descriptive/evaluative") since not all descriptive terms are dimensional, even though all dimensional terms are descriptive. It is not problematic for my analysis, especially that Bierwisch juxtaposes dimensional and evaluative adjectives.

1.5 Subjective$_{\mathrm{E}}$ adjectives – experiential vs. evaluative

I would like to argue that subjective$_{\mathrm{E}}$ adjectives can be divided into evaluative adjectives and adjectives which I will call *experiential*. Experiential adjectives are such terms which are used to describe our experiences – often sensory experiences. Some examples are: "heavy" (when talking about a bag carried by the speaker) and "long" (when talking about a flight which seems long to the speaker). The fact that they can be embedded under "find" stems from the fact that there is an experiencer parameter in their semantic structure. This person – the experiencer – experiences some quality in a subjective way peculiar to her. It is rather in accordance with experience than with the state of the world that determines the truth value of the utterance. Therefore, in order to experience the subjective quality of heaviness, I cannot just look at a bag – I need to try to lift it. Nevertheless, neither "heavy", nor "long" lexically encode the evaluation. If I say that some flight is long, I do not necessarily assert that it is good or bad. As we know, evaluative terms always carry evaluation, e.g. "tasty", "beautiful" or "handsome". Some adjectives are, I believe, both evaluative and experiential, for instance "tasty". One cannot say that something is tasty, if she has never tried this thing or something of the same kind. At the same time, calling something tasty implies an unambiguously positive evaluation of this object.[9]

Now the question is whether there is some diagnostic which would enable us to tell experiential subjective$_{\mathrm{E}}$ terms from evaluative subjective$_{\mathrm{E}}$ ones. My proposal consists in checking if a given sentence containing a subjective term is acceptable such as the construction "x seems P to me", while the

[9] A distinction similar to mine is proposed by Louise McNally and Isidora Stojanovic in their *Aesthetic Predicates* (2014) where they claim that the adjectives including an experiencer parameter (which are acceptable under "find"), should be distinguished from evaluatives (which are not acceptable under "find"). They believe that the rare acceptability of, e.g., "beautiful" under "find" stems from the fact that it is treated experientially, that is, the object in question is being compared to other objects that the speaker has already seen. I agree with their account to some extent, except that I believe that such predicates as "tasty" are both experiential and evaluative. Stojanovic and McNally, unlike me, do not analyse such adjectives as "heavy" when used in such a way, which suggests the presence of an experiencer parameter. We also differ when it comes to linguistic intuitions: according to these Authors, "I find this painting beautiful" sounds odd and is felicitous only under special circumstances. I, on the other hand, do not sense any infelicity there and I don't see a problem with using "find" with evaluatives, although, I admit that it might be the case that they are being used experimentally (e.g. "I find murder repugnant" might sound okay only because the speaker has a gut feeling that it is wrong (as moral intuitionists would say) – so a kind of experience but agreeing that this is the case requires adopting some specific metaethical stance).

resulting phrase would have to be as synonymous as possible with the same sentence embedded under "find" (in the sense that it would have to be possible to use it to express the same thing). In particular, reformulating the sentence into a construction of the "seem" kind should not lower the epistemic certainty expressed by the speaker. It seems that:

(21) I find this bag heavier than the other one.

is synonymous with:

(22) This bag seems heavier to me than the other one.

No matter how much we know about the weight of the bags. Similarly when I say: (23) I find the flight from Chicago to Hong-Kong longer than the flight from Hong-Kong to Chicago.

Knowing that the flights last the same, I can mean the same as:

(24) The flight from Chicago to Hong-Kong seems to me longer than the flight from Hong-Kong to Chicago.

The sentences with "find" which contain experiential subjective$_{\mathrm{E}}$ adjectives can be translated into those including "seem" with virtually no difference in meaning. Can this also be achieved with prototypical evaluative adjectives?

(25) I find Mona Lisa beautiful.

The sentence (25) is not synonymous with (26):

(26) #Mona Lisa seems beautiful to me.[10]

Similarly for predicates of personal taste:

(27) I find this cake tasty.

(28) #This cake seems tasty to me.

[10]Of course, it is not the case that (26) is unacceptable in general. It could be an utterance by a person manipulated by art critics who knows that according to the standard accepted in her community, Da Vinci's painting is not supposed to be considered beautiful, but she thinks that it is. I would like to thank an Anonymous Referee for bringing my attention to this possibility.

The unacceptability of (28) turns out to be problematic for my analysis since I take them to be expressed by subjective adjectives while they are both experiential and evaluative. Therefore, I should modify the interpretation of the test so that it expects that the rephrasing from "find" to "seem" is survived only by those subjective$_{\mathrm{E}}$ adjectives which are not evaluative.

The reason why the test proposed above gives the expected results remains unclear and requires further study. At this point, I would only like to suggest that adjectives that are purely experiential sound felicitous in constructions with "seem" because the speakers tend to allow their senses to be deceived by them to a certain extent and "seem" generally lowers epistemic certainty of the judgment it precedes. At the same time, they refer to the qualities which are, in a sense, measurable, and as such we will usually agree with others (and if we don't, others can correct us if we are wrong). If we say that something seems so-and-so to us, we allow for the fact that it might be different in reality. Expressing evaluation – so asserting that something has a positive or negative value – allows for a higher certainty of the speaker. After all, it is to these standards that one has to make the utterance true or false (at least in their view). This sketch of an explanation is just a suggestion requiring additional philosophical and linguistic reflection.

Before going any further, I should verify the hypothesis I tentatively posed at the beginning. In the introduction I stated that evaluative adjectives, unlike descriptive ones, generate faultless disagreement. The first qualification to this claim was made in section 1.3.1: some descriptive adjectives, namely, subjective$_{\mathrm{V}}$ terms generate it when they are used to refer to borderline cases. Another caveat of the results from the analysis presented in this part of the text: not only evaluative and subjective$_{\mathrm{V}}$ adjectives give rise to faultless disagreement. Experiential subjective$_{\mathrm{E}}$ terms are capable of that too. In a sense, it is not a new achievement – these terms have been considered subjective. I, however, distinguish a separate class of subjective$_{\mathrm{E}}$ terms which are not evaluative.

Someone, however, might want to object to the claim that such utterances as (13) are not evaluative at all. Perhaps the cakes we make are sometimes praised precisely because they are fluffy and light? Aren't sauces criticized as being heavy or, worse still, burned? My answer is: certainly. This does not, however, go against the conclusions reached in this section. Descriptive expressions are sometimes used to make evaluative claims. I tackle this phenomenon in the next part.

1.6. Descriptive expressions which are used to make evaluative judgments

Descriptive expressions have this deceptive property that they can be used to express evaluative judgments. For instance, when I describe a piece of art in such words as: "dynamic", "sombre", "harmonious" or "powerful", I often thereby express some evaluation even though I am not using evaluative terms *sensu stricto*, i.e. such us "beautiful". According to Louise McNally and Isidora Stojanovic (2014), (29) might, but does not have to express an evaluative aesthetic judgment:

(29) Picasso's Guernica is dynamic.

Sentence (29) expresses an aesthetic judgment because the predicate "dynamic" denotes an aesthetic concept. Nevertheless, whether this judgment is evaluative or not, depends on the context which includes a speaker's intentions.

Similarly, in some contexts saying of a man that he is tall and fit may carry a positive judgment of this individual:

(30) A: Is Bob handsome?
B: He's tall and fit.[11]

This does not mean that adjectives like "tall", "harmonious" or "fit" are themselves evaluative. They do not lexically encode any reference to value (e.g. aesthetic value). It is the context which makes it possible to use them to express evaluation. The fact that "harmonious" usually has a positive connotation, for instance when used to talk about a piece of music or a ballet recital, does not mean that for some current trends in art, e.g. contemporary performance, it cannot be used to express negative judgment.

Other examples of evaluative uses of descriptive terms have been shown by Cepollaro and Stojanovic (2016). One of them is the term "gypsy" which is a descriptive noun denoting a member of a certain ethnic group of Indian origin (its homonym being a related adjective). Consider:

(31) Our neighbourhood is ethnically diverse. Czechs, Vietnamese, Gypsies and Nigerians live here.

[11] Certainly, we cannot exclude that the context of (30) is such that B, not wanting to be rude, points to other characteristics of Bob, conversationally implicating that he does not have a nice-looking face. Here, however, we assume that B gives a positive answer to the question (nodding and using approving tone). I will say a bit more about implicatures below.

In (31) the term "Gypsies" was used descriptively. Sometimes, however, it is used as a slur. Nevertheless, just as "harmonious" or "tall", it does not semantically encode evaluation (as its at-issue content)[12].

It is certainly not the case that one can always express evaluation with a descriptive term. If someone asked me if I like my neighbours' kids and replied that I don't because they are loud, then I would not only describe them some way but I would also express my negative attitude to loud people. On the other hand, if I replied that I do not like them because they are tall, it would make no sense. The question arises then, what semantic or pragmatic mechanisms decide if a given expression can be used to express an evaluative judgment in a given context. On one understanding, it is the question which Stojanovic and McNally pose at the end of their 2014 paper.

My attempt at an answer is as follows: an evaluative judgment expressed with the use of descriptive terms is expressed only indirectly. Many evaluative terms – both thick and thin – are semantically underdetermined. This means that if I say that some painting is beautiful, I express a positive evaluation but I do not say what aesthetic properties make me judge it this way. In other words, my interlocutor does not have to know, what makes the painting beautiful to me, although if she is a competent speaker, she should know which properties are relevant for the semantics of "beautiful". Such a competent speaker knows that the size of the painting, the thickness of its frame, the smell of the canvas or the use of acrylic paint will not be relevant to my use of this predicate. It seems, therefore, that if someone asks me, like in example (30), whether Bob is handsome and I reply that he is tall and fit, then I communicate something along the lines: "I don't know exactly what 'handsome' means to you but Bob has such-and-such qualities, now you decide if that is relevant to your question". In everyday life, when we talk with people whom we know, we can assume that we know, more or less, what some evaluative predicates mean for them in various contexts. We can, therefore, answer directly. Similarly, we tend to be more straightforward when we intend to say that someone is handsome according to our standards and criteria.

[12]The claim that slurs does not carry evaluation as truth-conditional content is not uncontroversial or universally shared. There are so-called semantic accounts of slurs (see, e.g. Hom, 2008). Cepollaro and Stojanovic adopt an account according to which, evaluation is communicated via semantic presupposition. In the present text, I do not take a stance on this problem. I only use the example to illustrate a wider scale on which descriptive terms can be used to express evaluation.

The pragmatic mechanism I propose works well also in explaining some conversational implicatures. In a famous example by Grice (1989), a professor is writing a recommendation letter for his student. The letter is supposed to answer the question whether the student is a good researcher. The professor, who does not think very highly of his charge, lists his traits such as punctuality and good handwriting and does not mention his academic merits. A linguistically competent addressee of the letter understands that semantics of the expression "good scientist" does not include such character traits as punctuality, and therefore is able to calculate the implicature.

It seems, therefore, that the problems connected with context sensitivity and vagueness do not change much about where the intuitive demarcation line between descriptive and evaluative expression runs. These issues are nevertheless responsible for clouding this division. I hope to have clarified the picture a little. In the last part of the paper, I would like to briefly summarize the sketch of my classification of evaluative adjectives and apply it to the analysis of the term "tasty".

2. "Tasty" – an attempt at classification

In the present text I begin with the classical distinction between evaluative and descriptive terms. I hypothesize that the former, unlike the latter, generate faultless disagreement. That turns out to be problematic, however, since faultless disagreements arise sometimes when descriptive vague predicates are used to denote borderline cases. I adopt an independent division line, that is, between objective and subjective adjectives. I analyse Christopher Kennedy's proposal in which he distinguishes two kinds of subjectivity: the kind that has to do with vagueness and the kind connected with evaluativity. Assuming, after i.a. McNally and Stojanovic, that there is a separate group of adjectives which include an experiencer parameter, I postulate completing Kennedy's analysis by dividing the class of subjective expressions in the latter sense into proper evaluatives and experiential terms. The set of experiential terms would include those special senses of descriptive adjectives. I also propose a linguistic test useful in identifying these expressions.

Predicates of personal taste are probably the most often used expressions in discussions about faultless disagreement. Nevertheless, they have not received much detailed analysis in terms of descriptivity, evaluativity, thickness or thinness. In the last part of the present paper, I am going to draft a proposal of such an analysis in the light of the discussion above.

(31) A: Raspberry tomatoes are tastier than plum tomatoes.

B: No, plum tomatoes are tastier.

It looks like the disagreement in (31) can be called faultless and thus "tasty" should be considered subjective. It is clear that it is subjective$_{\mathrm{E}}$. Since it is used in the comparative form, it cannot be the case that it is being used to talk about a borderline case. The "seem" test is not conclusive because it is designed only to show that a given expression is not evaluative. "Tasty" is evaluative on first glimpse, however – it is difficult, if not impossible[13], to think of a context in which it does not carry positive evaluation of something. It is also experiential because this evaluation comes from the subjective experience of tasting something.

That "tasty" is a subjective$_{\mathrm{E}}$ predicate, does not mean that it is not subjective$_{\mathrm{V}}$ in a sense. It is a gradable adjective and therefore, it is vague. In my opinion, predicates of taste can have borderline cases only *intrapersonally* (not interpersonally) because of the presence of an experiencer. This is why cucumber soup might be definitely tasty to me, asparagus soup definitely not tasty and cauliflower soup "borderline tasty". There are no foods which would be borderline tasty to all language users.[14]

Finally, we should decide whether "tasty" is a thin or a thick term. I'm afraid that it is not an easy decision to make. On one hand, it seems that it cannot be thin since it says a little bit more than that something is good. On the other, it is not as thick as, for example, "generous". Someone is generous if she gives a lot to other people without expectation of reciprocity. Something is tasty – when? It seems that all the meanings that "tasty" encodes have two pieces of information: that it is good and that it can be used to talk about food (which is, plausibly, a metasemantic condition of its use). To say "x is tasty" is to say "x is good with respect to gustatory experience". Structurally, it is a counterpart of "x is beautiful" which would be equivalent to "x is good

[13]It should be stressed that this claim is limited to attributive uses of the predicate. It is possible that one can make a non-evaluative judgment, using "tasty" referentially, e.g., *A* asks: "Which cookies do you want me to buy?", *B* answers: "Buy me the tasty ones".

[14]The fact that the borderline zone on a scale of a predicate is not totally shared by all language speakers is common to all vague predicates to some extent. Speakers sometimes disagree about whether a given object is clearly P or borderline P. It is conceivable that there are such speakers who would say that someone who is 2.20 m tall is not clearly tall. The difference between these innocent gradable adjectives as "tall" and "tasty" is that one can always say that something that all others consider clearly tasty is disgusting. On the other hand, the person who calls a 2.20 m tall man not clearly tall, would be presumed as not grasping the meaning of "tall".

with respect to aesthetic experience". What remains to be explained is the mechanism by which this metasemantic information is communicated. Some patterns of behaviour of the term, including projection patterns, suggest it might be presupposition.

Providing a detailed lexical semantics of predicates of personal taste requires further deliberation. I hope that the classification proposed by me is a step in the right direction.

Bibliography

Barker, C. (2013). Negotiating Taste. *Inquiry*, *56*, 240–257.

Barker, C. (2002). The Dynamics of Vagueness. *Linguistics and Philosophy*, *25.1*, 1–36.

Bierwisch, M. (1989). The Semantics of Gradation. In M. Bierwisch, E. Lang (eds.), *Dimensional Adjectives* (pp. 71–261), Berlin: Springer.

Cepollaro, B., Stojanovic, I. (2016). Hybrid Evaluatives: In Defense of a Presuppositional Account. *Grazer Philosophische Studien*, *93*, 458–488.

Graff-Fara, D. (2000). Shifting Sands: An Interest-Relative Theory of Vagueness. *Philosophical Topics*, *28*(1), 45–81.

Grice, P. (1989). *Studies in the Ways of Words*. Oxford: Oxford University Press.

Hom, C. (2008). The Semantics of Racial Epithets. *Journal of Philosophy*, *105*, 416–440.

Karczewska, N. (2016). Disagreement about the Taste as Disagreement about the Discourse: Problems and Limitations. *Studies in Logic, Grammar and Rhetoric*, *46*(59), 103–117.

Kennedy, C. (2007). Vagueness and Grammar: The Semantics of Relative and Absolute Gradable Adjectives. *Linguistics and Philosophy*, *30*(1), 1–45.

Kennedy, C. (2016). Two Kinds of Subjectivity. In C. Meier, J. van Wijnbergen-Huitink (Eds.), *Subjective Meaning – Alternatives to Relativism* (pp. 105–126), Linguistische Arbeiten. Boston: De Gruyter Mouton.

Kolbel, M. (2003). Faultless Disagreement. *Proceedings of the Aristotelian Society*, *104*, 53–73.

Lasersohn, P. (2005). Context Dependence, Disagreement, and Predicates of Personal Taste. *Linguistics and Philosophy*, *28*, 643–68.

Lopez de Sa, D. (2015). Expressing Disagreement. *Erkenntnis*, *80*, 153–165.

Marques, T. (2016). Aesthetic Predicates: A Hybrid-Dispositional Account. *Inquiry*, *56*(6), 723–751.

McNally, L., Stojanovic, I. (2014). Aesthetic Adjectives. In J. Young (Ed.), *The Semantics of Aesthetic Judgments*. Oxford: Oxford University Press.

Odrowąż-Sypniewska, J. (2016). Faultless Disagreement, Predicates of Personal Taste and Vagueness. In P. Stalmaszczyk (Ed.), *Philosophical and Linguistic Analyses of Reference*. Bern: Peter Lang.

Richard, M. (2004). Contextualism and Relativism. *Philosophical Studies*, *119*, 215–242.

Sæbø, J. (2009). Judgment Ascriptions. *Linguistics and Philosophy*, *32*(4), 327–352.

Wright, C. (1997). *Truth and Objectivity*. Oxford: Oxford University Press.

Vayrynen, P. (2016) URL: `http://plato.stanford.edu/entries/thick-ethical-concepts/`, access: 20.11.2016.

Originally published as "Wyrażenia ocenne – próba klasyfikacji" *Studia Semiotyczne* 30/2 (2016), 129–153. Translated by the Author.

STUDIA SEMIOTYCZNE—ENGLISH SUPPLEMENT, vol. XXIX (2017), 201–247
ISSN 2451-2745
DOI: 10.26333/stsen.xxix.10

Adrian Ziółkowski[1]

The Reference of Proper Names, Semantic Intuitions and Experimental Philosophy[2]

Abstract This paper is a contribution to the debate concerning the kind of philosophical conclusions that can (or cannot) be derived from systematic empirical studies of intuitions about the reference of proper names. The focus of the paper is the famous study by Machery et al. (2004) in which intercultural differences in semantic intuitions between American and Chinese participants were observed. Machery *et al.* used the obtained results to question the usefulness of intuitions in philosophical discussions concerning the reference of proper names.

In this paper, I present the results of my own philosophical-experimental studies aimed at analyzing methods used in research dedicated to the problem of reference rather than semantic intuitions as such. These results indicate a significant instability of responses regarding the reference of proper names and their susceptibility to the impact of philosophically insignificant factors. Based on the collected data, I argue that methods used in experimental studies concerning the reference of proper names conducted to date do not guarantee the assessment of intuitions of the desired kind.

Keywords proper names, reference, intuitions, the descriptivist theory of names, the causal-historical theory of names, experimental philosophy

---

[1]Department of Philosophy and Sociology, University of Warsaw.
E-mail: adrian.a.ziolkowski@uw.edu.pl

[2]The project was financed with a grant from the National Science Centre, DEC-2012/05/HIS/03303.

Introduction

In recent years, philosophers have begun to advance arguments of a new kind. These arguments make significant references to empirical data gathered systematically for this purpose. Philosophers using such methods – experimental philosophers, as they call themselves – claim that many theses put forth in "classical" philosophical debates are empirical in nature and should be tested accordingly. At issue here is, first and foremost, the sometimes explicitly formulated expectation of uniformity of intuitions about popular philosophical thought experiments. As part of their research, experimental philosophers attempt to determine: (1) if the intuitions of nonphilosophers track those most often expressed in the literature (e.g. intuitions regarding the ascription of knowledge in Gettier scenarios); (2) what factors might contribute to the formation of such intuitions, and (3) if these intuitions can be systematically distorted.

Experimental philosophy is a young field: its history as an organized research enterprise in analytic philosophy goes back to the beginning of the 21st century[3]. Since then, experimental philosophers have sought to apply their methods in almost every philosophical discussion significantly shaped by thought experiments. This paper focuses on experimental methods used in the study of the reference of proper names, although this is not the only problem in the philosophy of language addressed by experimental philosophers to date – another example are intuitions about the contextual dependency of natural language expressions. Experimental philosophers have also participated in discussions pertaining to epistemology (of particular interest here has been the problem of knowledge ascription), moral philosophy (e.g. in relation to the famous trolley problem), action theory (in regard to questions such as intentionality or free will), philosophy of mind (focusing on the problem of consciousness, especially its types), and ontology (here, the intuitions of nonphilosophers about causation have been studied, among other issues). According to experimental philosophers, their methods can be applied to any problem addressable using thought experiments, as long as these can be presented in the form of scenarios comprehensible to nonphilosophers (an overview of issues addressed during the early stages of the development of experimental philosophy can be found in (Alexander, 2012)).

[3]In 2008, the leading figures of this emerging field published An Experimental Philosophy Manifesto, defining the field's main goals and research approach (Knobe & Nichols, 2008).

Experimental philosophy is not a homogenous field, a range of different conclusions are derived based on data obtained during research. Experimental philosophers often claim that data concerning the intuitions of nonphilosophers support or refute certain philosophical theories – for example, the fact that respondents are disinclined to ascribe knowledge to the protagonists of Gettier scenarios has been seen as an argument against the classical definition of knowledge. That said, researchers are sometimes more interested in undermining the overall usefulness of intuitions (or thought experiments) in a given field by demonstrating that responses elicited by thought experiments are systematically distorted – if it turns out that the intuitions being tested depend on some irrelevant factor, such as the respondent's ethnic background, they ought not to be trusted. It might be worth noting that these two ways of utilizing empirical data in experimental philosophy are almost antithetical. Therefore, philosophical-experimental research of the first kind is sometimes referred to as positive experimental philosophy, and research of the second kind, as negative experimental philosophy (Nadelhoffer, Nahmias 2007).

The legitimacy of the mode of philosophizing proposed by experimental philosophers is the subject of lively debate. Its beginning can be traced to the publication by Edouard Machery, Ron Mallon, Shaun Nichols, and Stephen Stich[4] (2004) of an article titled *Semantics, cross-cultural style*, the focus of this paper[5]. The article, written in a provocative style, garnered much scholarly interest as its authors used novel methods to formulate, in the spirit of negative experimental philosophy, strong conclusions that cast doubt on other, more "classical" philosophical approaches. The main research goal of MMNS was to gather empirical data which could be used to undermine the usefulness of semantic intuitions in debates concerning the reference of proper names. They searched for factors that seem philosophically insignificant but nonetheless impact semantic intuitions – their data, indicating cross-cultural differentiation in semantic intuitions about the reference of proper names, suggested that cultural background is one such factor.

[4]Hereafter referred to as MMNS.

[5]This is, to be precise, not the first publication considered to belong to the field of experimental philosophy. One earlier example is the report on the results of research on epistemological intuitions about Gettier scenarios and other related thought experiments by Weinberg, Nichols, and Stich (2001). The obtained data indicated that intuitions of this kind depend on the ethnic background and socioeconomic status of the respondents. However, the reliability of these results is doubtful – the experiment was conducted on very small respondent groups, and the more recent attempts to replicate the study turned out unsuccessful (Kim & Yuan, 2015; Seyedsayamdost, 2015).

In this paper, I present the results of my own studies. They cast doubt on the efficacy of research tools used to date to reveal the semantic intuitions of nonphilosophers, thus undermining the controversial thesis put forth by MMNS. According to MMNS, cultural differences in responses to the presented scenarios they observed testify to differences in semantic intuitions. The results of my experiments, methodologically modelled on the study conducted by MMNS, can be interpreted, as I am going to argue, to indicate that experimental methods used in this kind of research do not guarantee the acquisition of data reflecting intuitions of the desired kind. And since cultural differences noted by MMNS need not mean differences in semantic intuitions, the impact of their attack on the appeal to intuitions in debates concerning the reference of proper names is diminished.

In the first part of the paper, I briefly address controversies surrounding the notion of intuitions in philosophy. In the second part, I discuss the results of the study by Machery et al. indicating, in their opinion, the uselessness of semantic intuitions in philosophical debates. In the third part, I briefly recount the most important objections to their methods and interpretations to be found in the literature to date. In the fourth and fifth parts, I outline my own critique backed by the results of my experiments.

1. Intuitions in philosophy

The problem of intuitions has been the subject of considerable controversy in contemporary philosophical literature. It is often claimed that intuitions – particularly conceptual intuitions – are the factor responsible for our responses to philosophical thought experiments. On the other hand, no small number of philosophers have questioned the role of intuitions in philosophical debates. Timothy Williamson (2007), for example, claims that thought experiments are in fact modal arguments and that intuitions are irrelevant to the soundness of these arguments, thus playing no role in philosophical debates. Herman Cappelen (2013) has formulated an even more categorical thesis; according to him, the use of the term "intuition" in philosophy is so vague that it should be concluded that intuitions do not exist.

The crucial *locus* of conflict, regardless of the stance on the role of intuitions in philosophical considerations one might be inclined to adopt, is their nature – the only point philosophers agree on is that intuitions are mental states of a special kind. At least five individuation criteria for intuitions have been proposed in the scholarship so far, appealing to: (1) their phenomenal character; (2) their content; (3) their epistemic status; (4) their

origin; and (5) their functional role. The first approach suggests that the distinguishing characteristics of intuitions are their phenomenal properties – it is claimed, for instance, that intuitions are "seemings" accompanied by the impression that their content is necessarily true (e.g. Bealer, 1998). The content criterion may point to the abstract nature of the contents of intuitions (intuitions are supposed to not concern contingent objects) or to their modal nature (they are supposed to adjudicate questions of possibility or necessity). According to the third approach, the characteristic feature of intuitions is that they possess an *a priori* justification. The origin criterion refers to the cognitive competency considered to be the source of intuitions – usually linguistic competency and its related capacity to comprehend certain concepts (e.g. Ludwig, 2007 and 2010). The last approach to the problem of intuitions mentioned above is to determine some specific functional property shared by intuitions such as the fact that they are spontaneous reactions to philosophical thought experiments.

As can easily be seen, some of the aforementioned criteria are relatively strict and some relatively liberal – on some (e.g. the functional criterion), numerous mental states are going to be classed as intuitions; on others (e.g. the origin criterion), the set of mental states identifiable as intuitions is going to be considerably smaller. The solution of this conflict is beyond the purview of this paper. For the sake of the current discussion, I propose to assume a liberal, functional characterization of intuitions as spontaneous reactions to thought experiments – such characterization is presumed by the majority of experimental philosophers in their research (Weinberg & Alexander, 2014). This leaves open the question of whether a given spontaneous reaction to a scenario presented in an experimental study is an intuition of the desired kind, that is, if it is relevant to the philosophical issue raised in the study. The fundamental question I address in this paper is this: do methods hitherto employed in experiments concerning the problem of the reference of proper names provide data reflecting the sought-after intuitions, that is, intuitions expressing support for the particular theory of reference? As I have noted earlier, I am going to use the results of my research to argue for a negative answer to this question.

2. The study by Machery *et al.*

2.1. The area of dispute: two competing theories of reference

Let us discuss the results of the research to date. The departure point of the study conducted by MMNS (2004) was a debate between two historically

strongest traditions of thinking about the reference of proper names: the descriptivist and the causal-historical tradition. Neither of these traditions can be considered perfectly homogeneous; rather, they are groups of theories that share a common core but differ in detail. For simplicity's sake, I am going to speak of the descriptivist and the causal-historical theory of reference. Below, I present an outline of the key assumptions made in these approaches according to their characterization introduced into the literature by Machery et al. It is important to note that this outline paints a general, simplified and imprecise picture of two conceptions of reference neither of which deserves to be called a theory. Views on the problem of reference held by particular philosophers of language tend to be much more precise. However, even though the following reconstruction can be considered inadequate, it is important to present it in this form since it is this articulation that has shaped the approach to the problem of the reference of proper names dominant in experimental philosophy to date. Experimental philosophers seek to empirically distinguish between intuitions providing support precisely for such general, imprecisely characterized conceptions[6].

According to the descriptivist theory of reference, as it is reconstructed by MMNS, proper names are strictly tied to descriptions fulfilling two requirements: (i) the object referred to by a given name satisfies the description associated with that name; and (ii) this object is the only object in the *universum* that satisfies this description. Names "pick out" their reference from extralinguistic reality by means of such descriptions. MMNS go so far as to suggest that, according to the descriptivist, proper names simply are hidden descriptions, an opinion many proponents of descriptivism are likely to reject (e.g. Searle, 1985). The conception analyzed here is closest to the classical (and slightly archaic) standpoint of Frege (1977). Frege did tend to identify names with descriptions; he also permitted the possibility that the same name can be tied to different descriptions by different language users (that is, he permitted the instability of meaning for names) as long as all relevant descriptions unequivocally identified the same object. One language user, to give an example, could tie the name "Lech Wałęsa" to

[6]This fact can serve as the basis for another argumentation strategy against the study by MMNS: if their research does not concern intuitions relating to actual theories of reference, it can be considered philosophically irrelevant. In this text, I pursue a different strategy and claim that the methods proposed by MMNS do not guarantee empirical differentiation even between such generally characterized conceptions of the reference of proper names.

the description "the first head of Solidarity," and another to the description "the first Polish laureate of the Nobel Peace Prize."

The proponents of the causal-historical theory of reference, as it is reconstructed by MMNS, on the other hand, reject the notion that proper names should be perceived as mediated by other lexical units and hold them to be relatively independent carriers of the relation of reference. According to this theory, the relation of reference obtaining between a name and its designate is based on the history of the continuous use of that name. This history, in the case of each name, begins with the act of initial baptism whereby an individual language user (or a group of such users) introduces a convention for the use of the name in reference to some object. The name can be introduced into the language through ostension (the utterance "May this object bear the name N" accompanied by pointing to the "baptized" object) or by means of a determinate description. However, the name is autonomous in relation to the initial act of ostension or the description used in the course of the name's introduction into the language. The relation of reference between a proper name and its designate obtains thanks to the causal chain linking the current uses of the name to the act of initial baptism. It should be noted that this characterization of the causal-historical theory of reference is also a considerable simplification. It is hard to think of the outlined standpoint as a philosophical theory in the strict sense of the term; rather, it is a relatively general idea in need of precise articulation if it is to become a theory[7].

The goal of the study conducted by MMNS was to reveal preferences shared by members of different cultural groups in regard to the competing conceptions of the reference of proper names characterized above. It was carried out in reference to the famous thought experiments intended to counter the descriptivist theory of reference presented by Saul Kripke in the lecture series published as *Naming and Necessity* (1972).

In the first of these counterexamples, Kripke analyzes a situation in which the users of a certain name tie it to a description that is not satisfied by the object singled out as its designate by the history of linguistic practices related to the name in question. Kripke proposes the following hypothetical scenario. People generally associate the description "the author of the proof of the incompleteness theorem" with the name "Kurt Gödel." However, contrary to common belief, this proof is in fact due to a little known German mathematician named Schmidt. Schmidt died in unexplained circumstances

[7]For more precise and more adequate descriptions of the different philosophical theories of the reference of proper names, see Lycan (1999) and Muszyński (2000).

in Vienna and his friend, Gödel, got hold of the manuscript of the proof and published it under his own name. This raises the question: who does the user of the name "Kurt Gödel" associating this name with the description "the author of the proof of the incompleteness theorem" refer to here? To Schmidt or to the person who published the proof under their own name? The first option is supposed to obey the spirit of descriptivism; the second, the spirit of the causal-historical theory of reference. There is moderate consensus among philosophers that our semantic intuitions are bound to push us in the direction of the second option, undermining the descriptivist theory and providing support for the causal-historical conception of reference.

The second Kripkean counterexample concerns situations in which the description associated with a given name is not satisfied by any object in the universe of discourse. Kripke considers the example of the biblical prophet Jonah who was swallowed by a giant fish (or whale) for three days and three nights according to legend. Let us assume, after Kripke, that the story of the swallowing is an untrue tale, although the prophet whose life served as the basis for it really existed. This raises a question analogous to the one posed in the context of the Gödel case described above: who does the language user tying the name "Jonah" to the description "the prophet swallowed by a giant fish for three days and three nights" refer to? The causal-historical theory of reference permits the possibility of referring to the actual prophet. The same is not the case for descriptivism – since the key to the relation of reference in descriptivism is the description associated with the name, and the description under consideration is not satisfied by any object, the relation of reference simply does not obtain. In this case too, the majority of philosophers agree that our intuitions tend to support the causal-historical rather than the descriptivist theory of reference.

2.2. The experimental procedure and results of the study by Machery *et al.*

MMNS modelled the scenarios used in their study on the Kripkean thought experiments outlined above, dividing them into two types: Gödel cases and Jonah cases. Their interest focused on the so-called uniformity conjecture present in Kripke's argumentation, in their opinion. According to this conjecture, (a) there is going to be a far-reaching agreement among ordinary language users concerning the right responses to Gödel and Jonah cases, and (b) this agreement is going to favour the causal-historical theory. Therefore, in particular, no systematic differences in intuitions among groups

defined by philosophically insignificant factors, such as cultural background, should occur.

In order to find empirical evidence against the uniformity conjecture, MMNS decided to test semantic intuitions elicited by Gödel and Jonah cases in two distinct cultural groups expected to exhibit differences of the relevant sort given some more general psychological differences observed for them in previous research. Based on the results of cross-cultural studies conducted by Richard Nisbett (2003), in which the Western and East Asian traditions had been compared, MMNS formulated the expectation that there should occur systematic differences between these groups in regard to preferences related to theories of reference. The rationale was that, as Nisbett's research implied, members of East Asian culture are much less inclined to use causal categories to formulate descriptions of reality than are persons raised in the Western cultural sphere. Since the causal-historical theory is based on an explanation in causal terms, the researchers expected that intuitions supporting the descriptivist theory would be significantly more frequent among Asians than among members of Western culture – the cultural background of the majority of contemporary analytic philosophers so readily accepting Kripke's counterexamples.

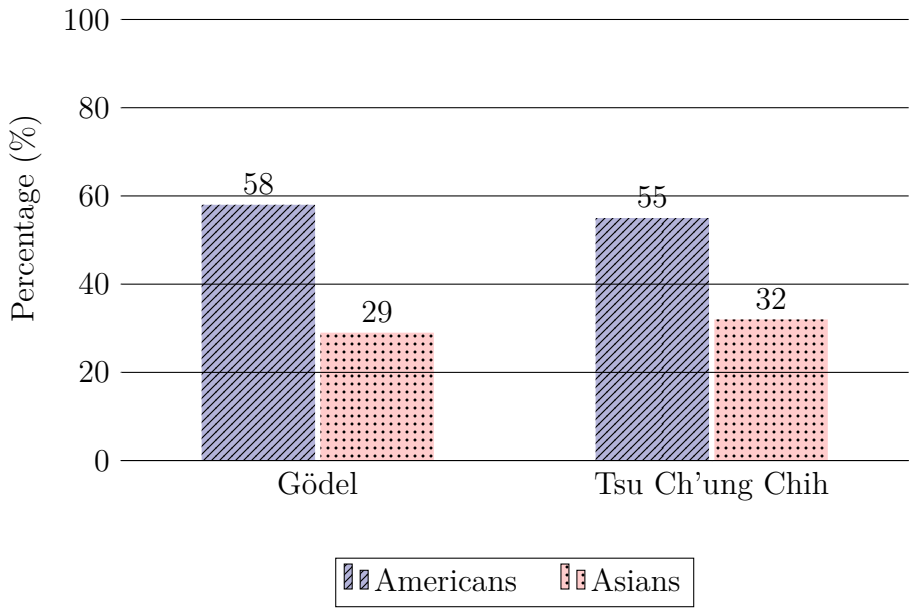
In order to verify this hypothesis, MMNS carried out an experiment involving members of Western culture (students of Rutgers University in the USA) and persons raised in the Asian cultural sphere (students of the University of Hong Kong)[8]. The groups comprised 40 persons each. The participants were presented with two scenarios of each type mentioned above: two Gödel cases and two Jonah cases. The scenarios presented to the two groups were almost identical, the main difference being that in one story of each type the protagonists bore names characteristic of Western culture (e.g. "Gödel"), and in the other, names typical of Asian culture (e.g. "Tsu Ch'ung Chih"). Importantly, the language of the experiment for both groups was English (students from Hong Kong were supposed to be fluent users of English).

The respondents revealed their preferences by choosing one of two response options. For Gödel cases, the descriptivist option was "[the protagonist] is talking about a person who actually satisfies [the description associated with N]," and the option consistent with Kripkean intuitions was "[the protagonist] is talking about a person commonly thought to satisfy [the description associated with N]." For Jonah cases, a respondent could express

[8]These were not seasoned philosophy students reasonably expected to be familiar with the problem discussed here.

their support for descriptivism by choosing the option "[the protagonist] is talking about a fictional person who does not really exist," and their support for the causal-historical conception by choosing the other option, stating that reference to an actual person who had inspired the false myths and legends obtains. Since the respondents had no opportunity to introduce an original response to the question posed, or to pick a third option such as "I do not know" or "none of the above is correct," one can speak of a forced choice here.

Whereas no statistically significant differences between persons raised in Western and Asian cultures were observed for Jonah cases (in both groups intuitions supporting the causal-historical theory predominated slightly), a clear difference between the groups occurred for Gödel cases: while responses supporting the descriptivist theory predominated among the Chinese, the Westerners were slightly more inclined to support the causal-historical theory. For details, see Graph 1.



Graph 1. Percentage of responses supporting the causal-historical theory for Gödel cases in the original study by Machery *et al.* (after: Machery, 2012).

It might be worth noting that although responses identified as supporting the causal-historical theory predominated among the Americans, almost half

of them reacted to Gödel cases by choosing a response considered to support the descriptivist conception. It could thus be concluded that the results for the members of Western culture alone undermine the uniformity conjecture since a considerable disagreement occurred in this group as to the correct response to Kripke's thought experiments.

The conclusion that the authors derived from the study is quite revolutionary as far as philosophical considerations regarding semantic issues are concerned. As has been mentioned earlier, MMNS claim that the data they obtained pose a serious challenge to the assumption, prevalent in philosophy, that the intuitions of philosophers about the reference of proper names (on which there is moderate consensus in philosophical circles) are universal. But this is not all. Citing the inclination on the part of Westerners, supposedly observed in their study, to support the causal-historical theory, Machery *et al.* suggest that a similar inclination among analytic philosophers, the majority of whom are Westerners, might be an expression of cultural conditioning or academic indoctrination adjusted to the demands of Western culture. And as they argue further, there is no conclusive argument for the idea that the semantic intuitions of philosophers from the West should be more accurate than, for example, those of Asians not trained in philosophy. In light of this, they propose a revision of the role assigned to semantic intuitions in contemporary philosophical discussions since the mode of doing philosophical semantics to date "smacks of narcissism in the extreme," in their opinion (Machery et al. 2004, p. B9).

3. Existing critique of the study by Machery *et al.*

The conclusions formulated by MMNS and their study have been subject to intense critique. Due to limitations of space, I cannot discuss all questions raised. In what follows, I provide a short description of the most important objections lodged by critics so far.

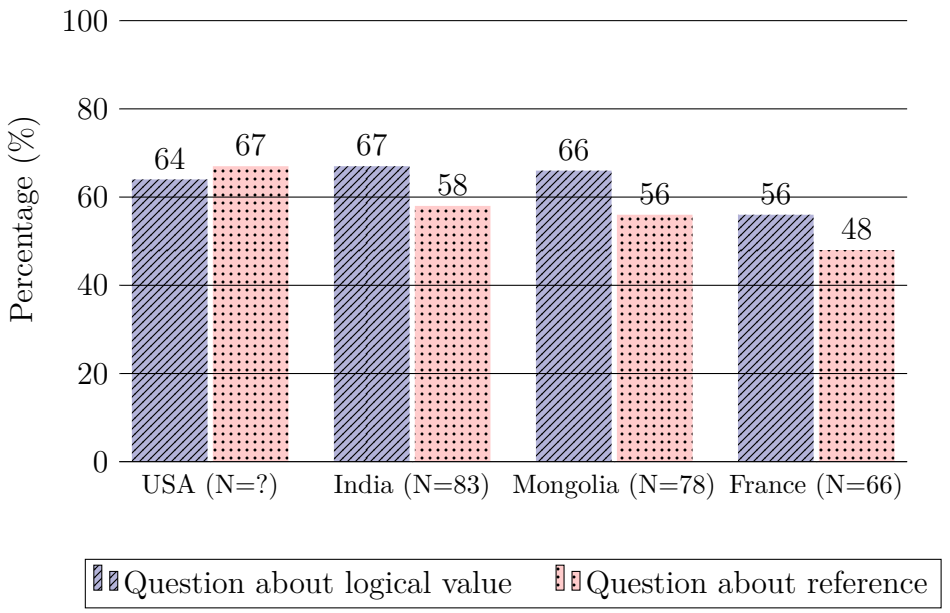
The study by MMNS, like all research in experimental philosophy, can be criticized from the position that philosophers are experts in regard to intuitions. Michael Devitt (2011) and Kirk Ludwig (2007 and 2010), for instance, hold that just like ordinary intuitions about issues in physics count less than those of experienced physicists, so too, ordinary intuitions about theories of reference count less than those of trained semanticians. Devitt (2011) also points to the excessive emphasis that MMNS put on the thought experiments they used – Gödel cases and Jonah cases – particularly in terms

of their role in Kripke's argumentation against the descriptivist theory of reference. As Devitt rightfully points out, even if MMNS's argumentation suffices to undermine the efficacy of these thought experiments, other elements of Kripke's refutation of descriptivism still retain their force and are enough to cast doubt on the criticized theory.

Since my paper focuses on methodological issues, I find such general objections much less interesting than more specific critical remarks – ones that do not undermine the legitimacy of MMNS's entire enterprise but point to elements in their experimental procedure that may have resulted in the acquisition of data related to a phenomenon altogether different from the intended one. Arguing in this spirit, Genoveva Martí (2009) doubts if the question posed to the respondents ("when the protagonist uses the name N, they are talking about…") – formulated such that the key name is mentioned rather than used – actually elicits the intended semantic intuitions. Intuitions of the desired kind should be about the mechanism of reference but, according to Martí, the question might have encouraged the respondents to instead focus on the theory that best describes this mechanism. As Martí argues further, insofar as competent language users not trained in philosophy can be seen as experts in the first matter, one ought not to expect their intuitions about the correct theory of reference to be authoritative.

In response to this objection, Machery *et al.* (2009) carried out an additional study of Gödel cases to determine if significant differences in response distribution would occur between the original question and its alternative formulation asking about the logical value of a sentence where the key name is used (e.g., "Gödel is the author of the incompleteness theorem"). If a participant's response was positive, it was treated as an expression of a semantic intuition supporting the descriptivist theory; if it was negative, it was seen as a token of support for the causal-historical theory. The participants responding to the two formulations of the question belonged to separate groups. Machery et al. (2009) asked members of four cultural groups: persons from India, Mongolia, France, and the USA.

Contrary to Martí's suspicion (Martí, 2009), the experimenters did not observe statistically significant differences in response distributions for the alternative formulations of the question for any of the tested groups (for results, see Graph 2). According to Machery *et al.*, this is enough to dismiss the objection that these formulations elicit intuitions of different kinds. Although I consider this conclusion to be premature, I am not going to explore this issue any further here. Something else should be noted: although the results of the second study by Machery *et al.* (2009) confirm the existence



Graph 2. Percentage of responses consistent with the causal-historical theory in the study by Machery *et al.* (2009).

of cross-cultural differences in reactions to Gödel cases, they also undermine the hypothesis which was supposed to explain them. As it turns out, the French express intuitions supporting the causal-historical theory significantly less often than not only Americans but also persons from Mongolia. At the same time, no difference for one of the formulations (the assessment of the logical value of the sentence in which the key name is used) between the Americans and the Mongols was observed. Therefore, if there are in fact cross-cultural differences in intuitions about the reference of proper names, they are not systematic East West divergences – some other phenomenon must be responsible for their occurrence than the one suggested by MMNS (2004) in their original study.

Another objection, one concerning the language of the original study, has been raised by Barry Lam (2010). The objection is that both the Americans and the respondents from Hong Kong assessed scenarios presented in English, the native language of the first group only. Lam thinks that, given this, differences in responses to Gödel cases chosen by the Americans and the Chinese need not testify to differences in intuitions about the reference of proper names between the two groups. In his opinion, an equally plausible

hypothesis is that the observed differences are due to varying degrees of linguistic competency.

In order to back his suspicion, Lam (2010) carried out an experiment in which he presented the users of English and Cantonese with appropriate native language translations of a scenario modelled on Gödel cases. The respondents learnt a story according to which a certain group does not know anything about Shakespeare other than the fact that he is the author of Romeo and Juliet. According to the scenario, the truth is that the play was not written by Shakespeare but by an unknown German writer by the name Spencer. The data obtained by Lam differed significantly from those acquired by MMNS – responses supporting the causal-historical theory predominated in both cultural groups.

This result provoked Machery et al. (2010) to attempt to replicate their original study in two languages. This time, they compared reactions of Americans to the original Gödel scenario (featuring the name "Gödel") to reactions of Chinese participants to a translation of this scenario. Contrary to the result obtained by Lam (2010), the cross-cultural differences observed in their original study were replicated: whereas responses supporting the causal-historical theory predominated among the Americans (62.2%), the Chinese were more inclined toward descriptivism (61%). The difference between data obtained by Lam on the one hand, and those acquired in the original experiment and the later study by Machery *et al.* (2010), on the other, might be the effect of using different scenarios or slightly different formulations of the key question regarding the name's reference. Options available to the respondents in the original study by MMNS were descriptions; those given in Lam's experiment were proper names – the respondent, when asked about the person referred to by the protagonist of the story using the name "Shakespeare," could choose between the option "Shakespeare" (in support of Kripke) and "Spencer" (in support of descriptivism). As has been rightfully noted by Beebe and Undercoffer (2016), the latter design does not allow one to successfully adjudicate between the two competing theories – for the descriptivist both responses are correct since, according to descriptivism, in the presented situation the names "Shakespeare" and "Spencer" co-refer. They both designate the actual author of *Romeo and Juliet*. The results of Lam's study should thus be approached with considerable caution.

An extremely interesting objection has been raised by Justin Sytsma and Jonathan Livengood (2011) who suggest that there is a dangerous ambiguity in the original study by MMNS. According to them, the formulation of the question regarding reference in this study did not sufficiently determine the

cognitive perspective that the respondent ought to assume while assessing the situation presented in the scenario. In particular, it is not clear if the desired perspective is one of the protagonist of the story (the user of the name ignorant of many facts, including the existence of Schmidt and his authorship of the incompleteness theorem) or one of the omniscient narrator (assuming access to information not possessed by the protagonist).

In order to test their hypothesis, Systma and Livengod (2011) decided to compare respondent reactions to three different formulations of the question regarding the reference of the name. The researchers used the original Gödel scenario (featuring the name "Gödel") from the study by MMNS. The only change concerned the formulation of the question – the one meant to encourage the respondents to assume the perspective of the protagonist was "when John uses the name 'Gödel,' John thinks he is talking about. . . "; on the other hand, to encourage the assumption of the perspective of the narrator, they used "when John uses the name 'Gödel,' he is in fact talking about. . . " The participants were divided into three groups: one answered the question in its original formulation, and the other two, questions unambiguously indicating the protagonist and the narrator perspective respectively.

Systma and Livengod (2011) observed significant differences in reactions between formulations encouraging respondents to assume the perspective of the protagonist and the narrator – in the first case the majority (78%) concluded that the protagonist thinks he is referring to the person who actually proved the incompleteness theorem; in the second case the majority (57.4%) stated that the protagonist in fact referred to the person who got hold of the manuscript and published it under their own name. Moreover, there also occurred a significant divergence in response distribution between the original version of the scenario (where less than 40% of the respondents chose the causal-historical response) and the other two versions. This means that even a slight change in the content of the question (the addition of "in fact") can translate into a significant change in response distribution. Based on this result, Systma and Livengood conclude that the method of measuring respondent opinion regarding the reference of proper names used in research to date does not provide decisive data as far as adjudicating between the competing theories of reference is concerned. It is simply not certain that all participants in the original experiment by MMNS responded to the same problem.

A possible ambiguity of the key question posed in the experiment by MMNS has also been pointed out by Kirk Ludwig (2007) and Max Deutsch (2009). In their opinion, the formulation of this question – the question about

the person the protagonist is talking about – does not distinguish between two significantly different senses of the term "reference": speaker's reference and semantic reference. The first concerns the person the user of a given name intends to refer to; the second concerns the actual reference of the name in the same use. The conflict between descriptivism and the causal-historical theory regards semantic reference, not speaker's reference. According to Ludwig and Deutsch, there is a risk that some of the respondents may have understood the question asked by the experimenters differently, and that while some expressed their intuitions in regard to semantic reference, others addressed the problem of speaker's reference. This could pose a serious challenge to MMNS since, granted the objection, interpreting all or at least some (which?) of the responses as supporting either of the competing theories might be invalid.

The discussion between Machery and Deustch ultimately resulted in a cooperative project – they decided to join forces to experimentally test the aforementioned objection. The experiment by Machery, Deutsch and Systma (2015) used the Gödel scenario (featuring the name "Gödel") from earlier research, but featuring a clear formulation of the question regarding reference: "when the protagonist of the story uses the name 'Gödel,' regardless of his intention, he is in fact talking about…". Just like in the original study by MMNS, the scenario was assessed by American and Chinese participants. It turned out that the results for the clear formulation did not significantly diverge from the original study. 59.9% of the Americans and only 38.8% of the Chinese chose the response linked to the causal-historical theory. Differences between the groups once again turned out to be statistically significant, apparently dismissing the objection raised by Ludwig and Deutsch.

As can be seen from the discussion surrounding the results of the numerous studies based on methods similar to that initially employed by MMNS, there is a considerable instability in response distribution even if the material is largely similar or identical to that used in the original study. In the most recent attempt to replicate the experiment by MMNS carried out by Beebe and Undercoffer (2016) and involving a sufficiently large respondent group, effects similar to those observed by MMNS occured: the Chinese were less inclined than the Americans to support the causal-historical conception for Gödel cases, and no cross-cultural differences occurred for Jonah cases. That said, the distributions were different than in the original study. For Gödel cases, the differences were small but significant (53% of the Americans and 43% of the Chinese chose the causal-historical option); for intuitions elicited by Jonah cases, the majority of which (ca. 2/3) in the study by

MMNS coincided with Kripke's intuitions, clearly supported descriptivism in the replication study (also ca. 2/3).

The discussion surrounding the controversial study by MMNS and their conclusions briefly outlined above does not yield an unequivocal picture. The remainder of this text is devoted to a detailed description of my own studies concerning intuitions elicited by scenarios modelled on Kripke's counterexamples. I am going to use the results obtained during these studies to argue in favour of the positions of the critics. My argumentation pertains to methodological issues – I am going to claim that the method used by MMNS does not guarantee that the assessment of participant responses reflects their semantic intuitions.

4. The author's research

4.1. Basic premises and main goals

So far, the discussion of the results of philosophical experiments concerning the reference of proper names has focused predominantly on Gödel cases, for which an interesting disproportion has been noted, and much less on Jonah cases. However, Jonah cases could provide data crucial for a deeper understanding of how ordinary intuitions about reference are shaped. According to Devitt (2011), Jonah cases can tell us more about the ordinary notion of reference because they are closer to problems typically encountered by everyday language users[9].

My studies focused on Jonah cases. One reason for that was to fill the gap in the existing literature. The second and more important reason was my suspicion that Jonah cases had been adapted for philosophical-experimental research in a methodologically flawed way. Namely, one of the options that the respondents could choose from in the original study by MMNS (2004) – the option interpreted as expressing support for the descriptivist theory of proper names – did not constitute a response the proponent of this theory should prefer in a Jonah context. The option identified as descriptivist was: "[the protagonist] is talking about a fictional person who does not really exist." This suggests that the name in question refers to some fictional object. It is difficult to tell exactly how the expression "fictional object" should be understood here but, regardless of its interpretation, there is no doubt that, as far as Jonah cases are concerned, the classical descriptivist theory

[9]It should be noted that, in general, Devitt has considerable reservations concerning the possibility of extracting accurate and competent semantic intuitions out of nonphilosophers using techniques proposed by experimental philosophers.

of proper names does not entail reference to a particular object. Jonah cases are situations in which a name is tied to a description that does not unequivocally designate any object in the world. The descriptivist response in a Jonah situation should thus indicate reference failure. The first person to point this out was Henry Jackman (2009). In light of this, it is doubtful that choosing the aforecited option can be considered an expression of support for descriptivism.

The main goal of my experiments was to determine if formulating a more adequate option expressing support for descriptivism would spur different reactions from those elicited by the original response option. Inspired by suggestions made by Jackman (2009) and Deutsch (2009), I also decided to check if different Jonah scenarios constructed based on the same strategy as the story presented by MMNS would result in different responses. I thus set out to determine, first, if differences in response distribution for a given scenario would occur depending on the formulation of one of the response options, and secondly, if response distributions for different Jonah cases would turn out the same. The latter effect is to be expected if the respondents express support for one consistent theory of reference. The second venture was purely exploratory but it ended up providing data which I found the most interesting.

4.2. Design and procedure

Three Jonah type scenarios were prepared for the purpose of the study. They were similar in structure to the stories presented in the original study by MMNS. Each scenario described a language user belonging to a larger linguistic community tying a given proper name to a description. The titles of these scenarios – *Mapemba*, *Homer*, and *Einstein* – stem from the proper names used in each scenario. All three stories contain the information that, contrary to the opinion widely shared by the appropriate linguistic community, a single person satisfying the description associated with the key name never existed. This is described in the scenarios as resulting from a "mythologization" of an actual historical person (*Mapemba*, *Homer*) or a simple mistake (*Einstein*). Importantly, although in no case is the description true of some one individual and that individual only, the descriptions featuring in *Homer* and *Einstein* could be treated as nonempty general names – they could be truthfully predicated of every member of a group of authors/inventors indicated in the scenarios.

Each participant in the experiment learnt about all three scenarios presented in random order. Each respondent was randomly assigned to one

of three versions of the scenarios – *Fictional Person*, *No One in Particular*, and *No Reference* – corresponding to different formulations of the response option aspiring to reflect intuitions supporting the descriptivist theory. In the *Fictional Person* version, the formulation from the original study by MMNS was used: "[the protagonist] is talking about a fictional person who does not really exist"[10]; the formulation in the *No One in Particular* version was: "[the protagonist] is not talking about anyone in particular"; and the formulation in the *No Reference* version was "[the protagonist] does not refer to anyone." The last formulation is assumed to come closest to the response that the proponent of the descriptivist theory should issue in each situation. Significantly, the response assumed to express support for the causal-historical conception was the same across all three versions of each scenario. Hence, the only manipulation consisted in changing one response option available to the respondent. The expected divergence in response distribution for the particular versions (formulations of the second response option) was thus assumed to provide an argument in favour of the thesis that the three alternative formulations of the descriptivist option[11] do not in fact express the same ordinary intuitions about the reference of proper names.

All participants in the study were native speakers of English. The scenarios were presented in English in the form recounted below.

4.3. The material: three Jonah cases

Below I present the content of the three scenarios prepared for the purpose of the study. Differences between their particular versions have

[10] As has been rightfully noted by an anonymous referee, this formulation is troublesome for an additional reason – it is a pleonasm, that is, the same thought is stated twice. The expression can be seen as unfortunate and thus should not be used since such elements lead to interpretive problems. In light of Grice's theory of conversational maxims (e.g. 1975), tautological utterances usually constitute breaches of the maxim of quantity (they convey superfluous information) and thus suggest an occurrence of a conversational implicature (indirect communication). MMNS certainly did not intend to use this formulation to encourage the respondents to search for conversational implicatures. Unfortunately, since the goal of my study is to test respondent reactions to scenarios designed based on the same method as the one used in the experiment conducted by MMNS, I could not avoid "inheriting" this flaw from their original study.

[11] I use the expression "descriptivist option" as shorthand – of course, if the respondents' preferences in the case of the three supposedly descriptivist responses are not identical, one cannot say that they all support descriptivism. Strictly speaking, at least two of them are not descriptivist responses but responses initially intended to express descriptivist intuitions, although they do not actually do so.

been indicated. These differences concern questions only; the content of each scenario is the same across all versions.

MAPEMBA

Thenga lives in a small African town called Kwende. Like most of Kwende inhabitants, he believes that Kwende was founded by a shaman called Mapemba. Moreover, Thenga believes that people who lived in the area before Mapemba's reign had been affected by recurring plagues and catastrophes. In fact, Mapemba is believed to have stopped those plagues with his magical powers and thus to have given the people of Kwende a peaceful life.

The truth is different, however. There never was any shaman who fought the plagues with magical powers. The origins of Kwende are connected with the activity of an inventive tribal leader called Ndembo, who had an idea how to use a nearby river to irrigate the cultivations and increase their efficiency. This solution improved the living standards of Kwende's people so much that it gave rise to a legend. The story of the inventive leader was passed from generation to generation. In the process it was gradually altered so that in the end it became a story about a shaman with magical powers. Those changes were accompanied by alterations in the name of Kwende's founder, which in the end became 'Mapemba'.

Assuming that the above story is true, answer the following question: when Thenga uses the name 'Mapemba', is he actually talking about the inventive leader Ndembo, who is the original source of the Mapemba legend, or is he talking about a fictional person, someone who does not really exist [the *Fictional Person* version] */ or is he talking about no one in particular* [the *No One In Particular* version] */ or maybe he is not referring to anyone* [the *No Reference version*]*?*

HOMER

Jacques is an inhabitant of 16th-century France. Like most of his well-educated contemporaries, Jacques believes that Homer, a nomadic blind poet living in the 7th century BC, was the author of The Odyssey, a famous ancient Greek epic. Jacques acquired this belief while studying at leading medieval universities in Europe. But the truth is different.

The Odyssey is a piece of work that has no single author. The inspirations for the story depicted in The Odyssey can be found in the tales told by Callicrates, a story-teller living in ancient Greece in the 10th century BC. Callicrates was neither blind nor did he travel much in his life. His stories

were so popular that they spread around Greece and people passed them on from generation to generation.

As the time passed, the stories were altered, some elements were replaced by new ones, some of them disappeared. Many people contributed to the final version of The Odyssey. In the end, the story does not have much in common with Callicrates' original tales. Along with the changes of the content and of the form of the story, people were changing their beliefs about the author, his life and his name. They started to believe that The Odyssey – which is the title that appeared with all the other modifications – was written by a nomadic blind poet called Homer.

Assuming that the above story is true, answer the following question: when Jacques uses the name 'Homer' is he actually talking about the story-teller Callicrates, whose tales were the inspiration for The Odyssey and who is the original source of the Homer legend, or is he talking about a fictional person, someone who does not really exist [the *Fictional Person* version] */ or is he talking about no one in particular* [the *No One In Particular* version] */ or maybe he is not referring to anyone* [the *No Reference* version]*?*

EINSTEIN

James is a high-school student living in Tinsbury, a small town in the south of England. Like most of Tinsbury inhabitants who attended high-school in their hometown, James believes that Albert Einstein was a physicist who invented the atomic bomb. Like most of residents of Tinsbury, James hasn't got any other beliefs concerning Albert Einstein. The truth is different, however.

The atomic bomb was not invented by Albert Einstein. In fact it was not invented by any single person but by a large group of scientists who participated in the Manhattan Project in the USA during World War II. Among others, Robert Oppenheimer, Ernest Lawrence and Harold Urey were involved in this project. Albert Einstein, who is famous mostly for his contribution to the Theory of Relativity, had never worked on the atomic bomb.

James' belief concerning Einstein is due to a mistake of an aged Physics teacher. Thinking that the atomic bomb is one of the greatest inventions of the 20th century and believing that Albert Einstein was the most eminent physicist of that century, the teacher ascribed this discovery to Einstein by mistake.

Assuming that the above story is true, answer the following question: when James uses the name 'Albert Einstein' is he actually talking about Albert Einstein, the author of the Theory of Relativity, who was the source

of the teacher's mistake, or is he talking about a fictional person, someone who does not really exist [the *Fictional Person* version] / *or is he talking about no one in particular* [the *No One In Particular* version] / *or maybe he is not referring to anyone* [the *No reference* version]*?*

4.4. Subjects

The experiment took the form of an electronic survey published online. As has been noted earlier, the language of the study was English and the participants were native English speakers, mainly from the United States of America, but also from Great Britain. The respondents received an invitation to participate in the study with a hyperlink to the survey via email (in the email, they were also encouraged to pass the invitation on and ask others to participate).

The respondents were volunteers and did not receive any remuneration for their participation in the study. 136 persons responded to the survey; 22 submissions were eliminated from further analysis because the persons in question were not native speakers of English or because they reported having undergone philosophical education at the level of Bachelor's degree or higher. The data presented below are from a group of 114 respondents.

39 persons were assigned to the Fictional Person version, 38 to the No One In Particular version, and the remaining 37 persons, to the No Reference version. 56.1% of the sample were women, 43.9% were men. The youngest participant was 18, and the oldest 71 years old – the average age of the sample was 34.4 years old with the standard deviation of 12.2. The majority of the respondents, 62.5% to be precise, were persons not older than 35.

4.5. Results

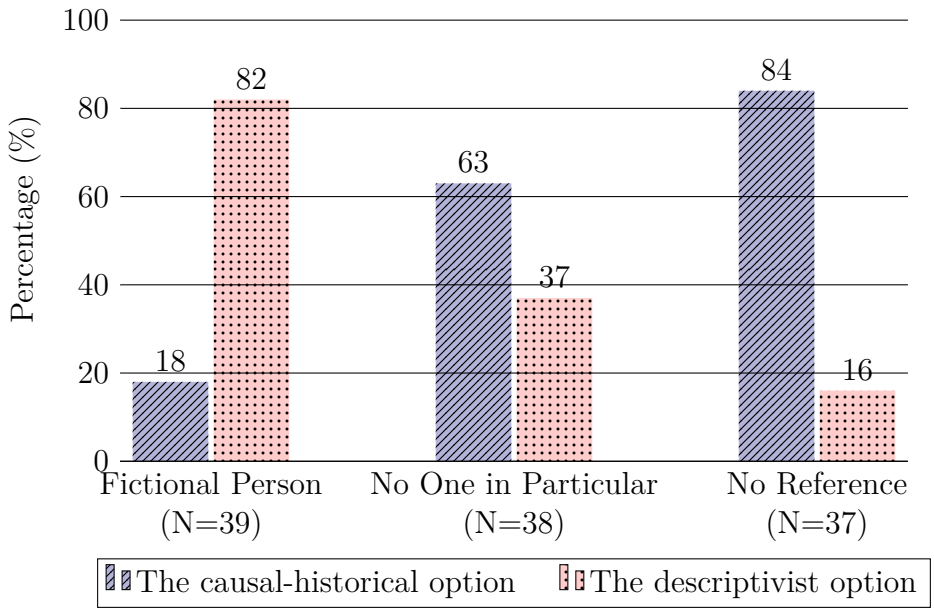
4.5.1. Comparison of the alternative formulations of the descriptivist option

I begin the presentation of the results by comparing respondent reactions to the scenarios depending on the particular version, that is, the formulation of one of the response options available to the respondents[12].

For the *Mapemba* scenario, clear and statistically significant differences in response distribution between the particular versions of the scenario

[12]Since the method used to gather the data (forced choice from several options, also referred to as closed-ended multiple choice) only allows for the measurement of the dependent variable on a nominal scale, the statistical tests used in the analysis were based on the comparison of frequency for the appropriate categories (response options). χ-square and Z tests were used.

were observed[13]. In the case of the Fictional Person version, a significant majority of respondents chose the option considered to express support for descriptivism in the original study by Machery *et al.* Namely, they concluded that the protagonist of the story referred to the fictional person who does not really exist. However, the situation was different in the case of the *No One In Particular* and *No reference* versions – persons expressing support for the causal-historical theory constituted the majority in each group. The percentage of responses supporting the causal-historical theory in these cases was significantly higher than in the case of the *Fictional Person* version[14]. One can thus speak of a "reversal" in respondent intuitions between the *Fictional Person* version on the one hand, and the *No One In Particular* and *No Reference* versions, on the other. The described result is illustrated on Graph 3.

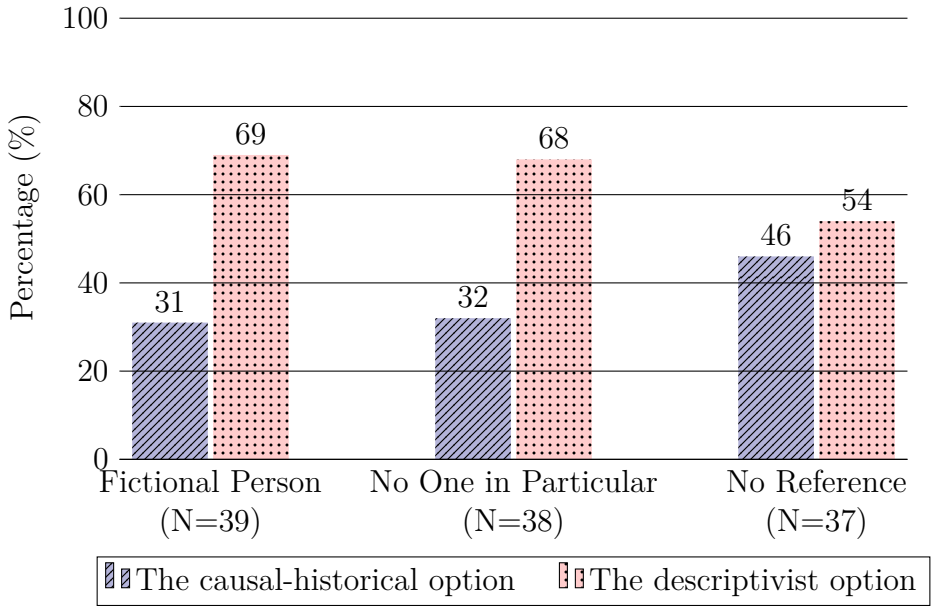


Graph 3. Response distribution for the *Mapemba* scenario depending on the version.

Unlike in the case of *Mapemba*, in the case of *Homer*, χ-square and Z tests did not indicate any significant differences in response distribution between the particular versions of the scenario. In the case of the *Fictional Person*

[13] $\chi^2(2) = 34.94$; $p < 0.001$.
[14] Z tests (significance level $p = 0.05$).

and *No One In Particular* versions, persons selecting the option supporting the causal-historical theory constituted a clear minority, around 1/3 of the respondents. A slightly different situation occurred in the case of the *No Reference* version – here, the preferences of the participants were distributed almost equally, with slightly more than half choosing the descriptivist option stating reference failure. However, statistically adequate testing does not in fact permit the conclusion that this option was more popular than the competing one[15]. In the case of the other two versions, the predomination of the descriptivist options was statistically significant[16]. Therefore, despite the fact that the tests cited at the beginning of the paragraph did not show any significant differences in response distribution between the particular versions, it seems justified to speak of a clear tendency. The results for the *Homer* scenario are illustrated on Graph 4.



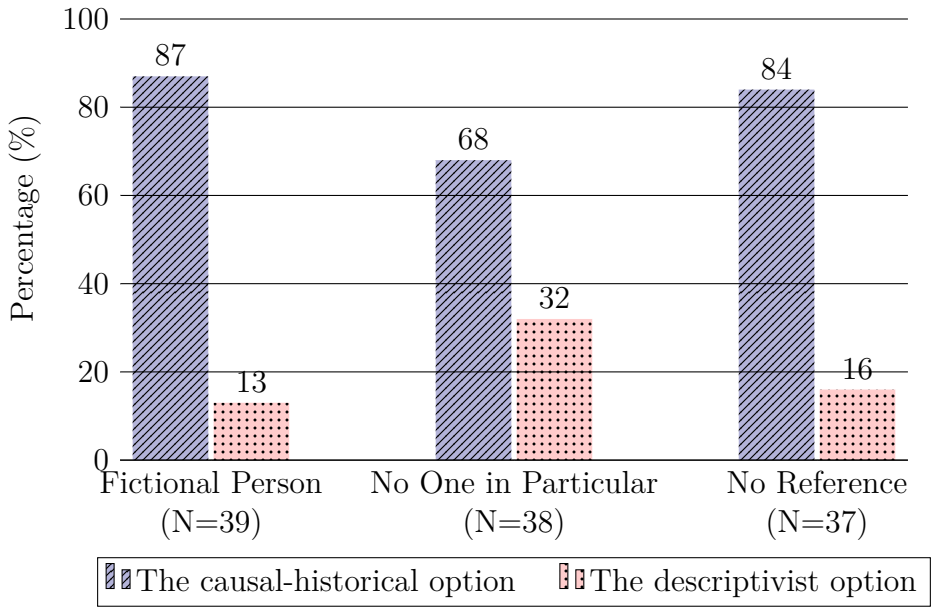
Graph 4. Response distribution for the *Homer* scenario depending on the version.

The formulation of the descriptivist option also had an impact on respondent preferences as far as the *Einstein* scenario is concerned[17]. In

[15] $\chi^2(1) = 0.24$; not significant.
[16] Fictional Person $\chi^2(1) = 5.77$; $p = 0.016$. No One In Particular: $X^2(1) = 5.16$; $p = 0.023$.
[17] $\chi^2(2) = 4.94$; $p < 0.047$.

the case of the *Fictional Person* and the *No Reference* versions, responses supporting the causal-historical theory, that is, those stating reference to Albert Einstein, predominated decisively. However, in the case of the *No One in Particular* version, the predomination of the causal-historical option over the descriptivist one was slightly smaller – here, close to 1/3 of the respondents concluded that the protagonist did not refer to anyone in particular. The results of the experiment for the *Einstein* scenario are illustrated on Graph 5.

To summarize, the results of the study in terms of the comparison of the alternative formulations of the descriptivist option permit the conclusion that the impact of this factor was observed for all three tested scenarios, although in some cases it was weaker than in others.



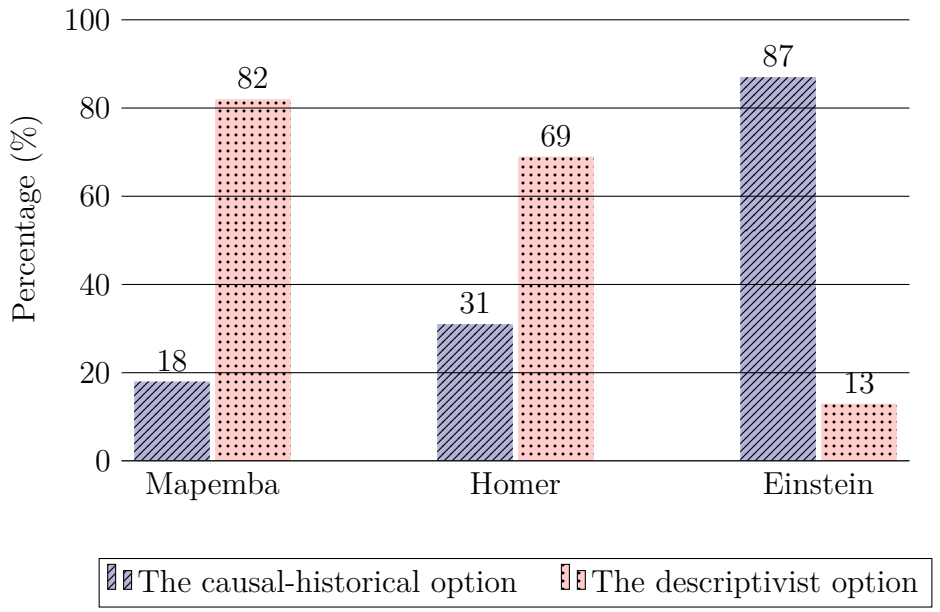
Graph 5. Response distribution for the *Einstein* scenario depending on the version.

4.5.2. Comparison of respondent reactions depending on the particular Jonah type scenario

.

The second goal of my first methodological study concerning the reference of proper names was to determine if respondent preferences for the analyzed conceptions of reference would be stable across different Jonah case scenarios.

For the formulation stemming from the original study by MMNS, that is, the one including the option that the protagonist of the story referred to a fictional person, significant differences in response distribution were observed between *Einstein* on the one hand, and *Mapemba* and Homer, on the other. In the first case, the majority of respondents concluded that the user of the name "Einstein" in fact referred to Albert Einstein; in *Mapemba* and *Homer*, the majority opined that the protagonist referred to a fictional person[18]. It thus turns out that even in the case of the standard approach to Jonah cases proposed by MMNS the preferences of nonphilosophers are not stable across different scenarios of this type. This result is summarized on Graph 6.

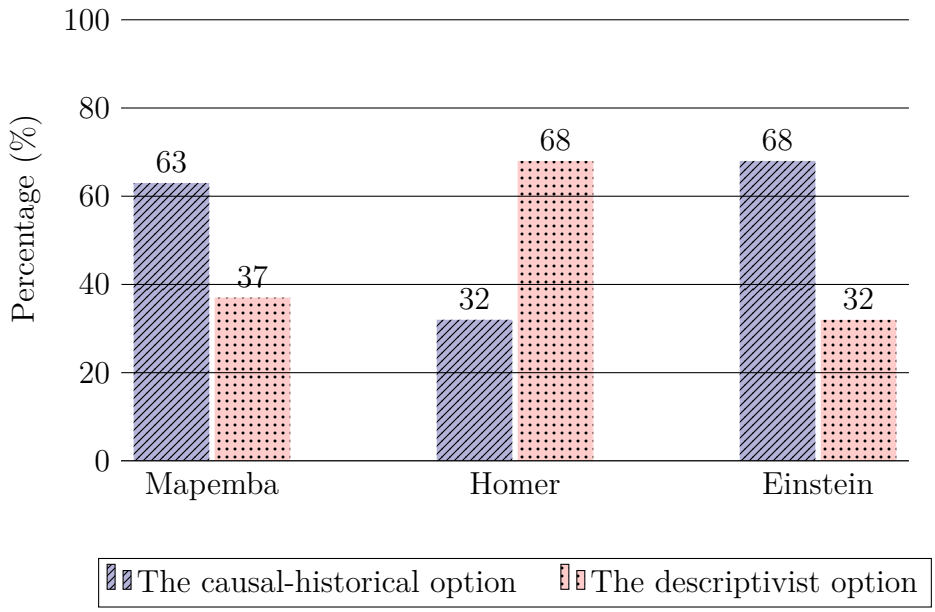


Graph 6. Response distribution for the particular scenarios in the case of the *Fictional Person* version (N = 39).

Similarly significant, albeit differently distributed, differences were observed in the case of the *No One In Particular* version. The respondents

[18] $\chi^2(2) = 42.7$; $p < 0.001$ and adequate comparisons using the Z test indicated that these differences are statistically significant.

assigned to this version assessed the *Mapemba* and *Einstein* scenarios in a very similar manner – here, responses supporting the causal-historical theory predominated. However, in response to *Homer*, the majority of respondents stated that the protagonist did not speak of anyone in particular[19]. This result is shown on Graph 7.



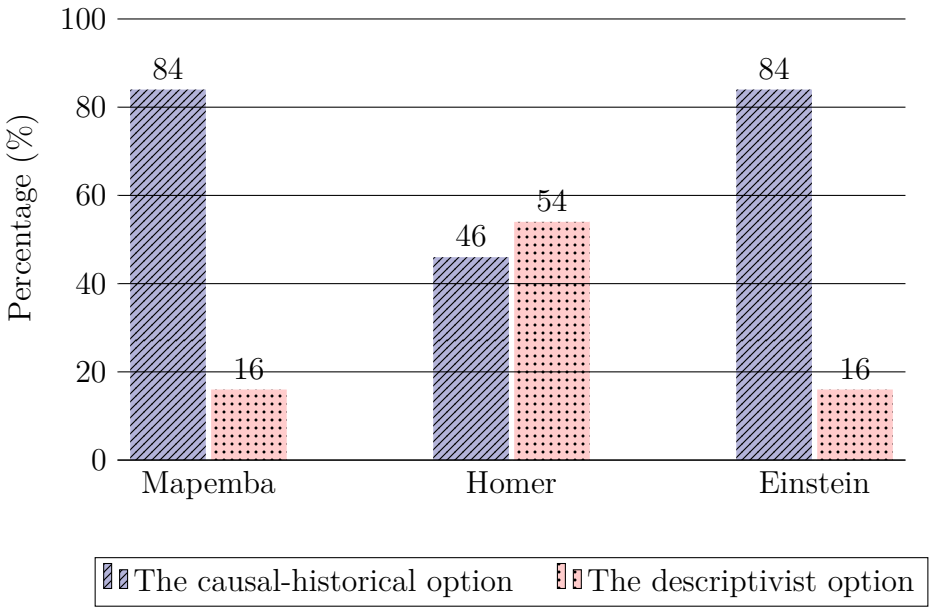
Graph 7. Response distribution for the particular scenarios in the case of the *No One in Particular* version (N = 38).

A clear divergence in respondent reactions to the particular scenarios also occurred in the case of the *No Reference* version. Just like in the case of *No One in Particular*, here too, respondent reactions to *Mapemba* and *Einstein* were very similar – in both cases the response stating reference failure was highly unpopular. In the case of Homer, on the other hand, none of the options predominated – respondent preferences were distributed almost evenly between the descriptivist and the causal-historical options[20].

[19]justifying The statistical significance of these differences has been confirmed by both the χ-square test, $\chi^2(2) = 12.16$; $p < 0.002$, and appropriate comparisons using the Z test.

[20]Differences in respondent reactions to *Homer* relative to *Mapemba* and *Einstein* turned out to be statistically significant according to both the Z tests ($p = 0.05$) and the χ-square test $\chi^2(2) = 17.21$; $p < 0.001$.

Detailed information regarding the response distribution for this version are shown on Graph 8.



Graph 8. Response distribution for the particular scenarios in the case of the *No Reference* version (N = 37).

A summary of the comparison of the results for the particular scenarios should emphasize the fact that in the case of each formulation of the descriptivist option, the respondents reacted to one of the scenarios differently than to the other two. Interestingly, it was not the same scenario – in the case of the *No One in Particular* and *No Reference* versions, the outlying scenario was *Homer*; in the case of the original formulation from the study by Machery *et al.*, it was *Einstein*. It is worth noting at this point that the picture implied by the data gathered during my experiment is far more complex than has been the case in previous research dedicated to the problem of the reference of proper names. However, before I proceed to interpret these results, I would like to discuss the results of an additional study I conducted since they can help cast more light on the data presented so far.

4.6. The additional experiment

4.6.1. General characterization

There are several significant differences between the Jonah type scenarios used in my first experiment. I address them later in the paper. However, one difference between the *Einstein* scenario and the *Mapemba–Homer* pair is so fundamental that it calls for separate treatment. Namely, in the case of *Einstein*, the proper name borne by the referent intended by the protagonist and the name used by the protagonist have the same shape; the protagonists of *Mapemba* and *Homer*, in contrast, use names different in shape from those borne by the possible referents of these names ("Ndembo" vs. "Mapemba" and "Callicrates" and "Homer"). As is implied by the results of the first experiment, respondent reactions to the Einstein scenario were only slightly sensitive to the formulation of the descriptivist option – responses stating that the protagonist actually referred to Albert Einstein predominated in the case of each version. As regards *Mapemba* and *Homer*, the descriptivist response predominated in the case of at least one version. Perhaps this difference in the shape of the name had a significant impact on respondent reactions? The main goal of my second study was to empirically test this supposition.

In light of the above, alternative *Mapemba* and *Homer* scenarios were used in the second study in which the names of the persons whose actions inspired the legends responsible for the false beliefs of the described linguistic communities had the same shape as the names used by the protagonists. This was the only difference between the original scenarios and those analyzed in the additional study.

Each participant in the second experiment learnt two scenarios – *Mapemba'* and *Homer'* – presented in random order. The respondents were randomly assigned to one of the three versions of each scenario: *Fictional Person*, *No One in Particular*, or *No Reference*. The characteristics of these versions were analogous to the first study – the versions only differed in the formulation of the descriptivist option.

4.6.2. Subjects

Just like the initial study, the second survey was carried out over the Internet. The participants were recruited via the online portal Amazon Mechanical Turk (`www.mturk.com`) – registered users had access to a hyperlink to the survey and could commit to taking it in exchange for a fee of 0.30 USD. This time, then, the participants were not volunteers. 156 persons

filled out the survey; 21 respondents were not taken into account in later analysis because they reported having an academic degree in philosophy or were not native speakers of English. The statistics presented below are based on the sample size of 135 persons.

55.6% of the respondents were men, 44.4% were women. The youngest participant was 19 and the oldest one was 73 years old, the average age being 37.4 years old, with the standard deviation of 12.2. 60% of respondents were not older than 36. The distribution was thus slightly skewed toward younger persons.

4.6.3. Results

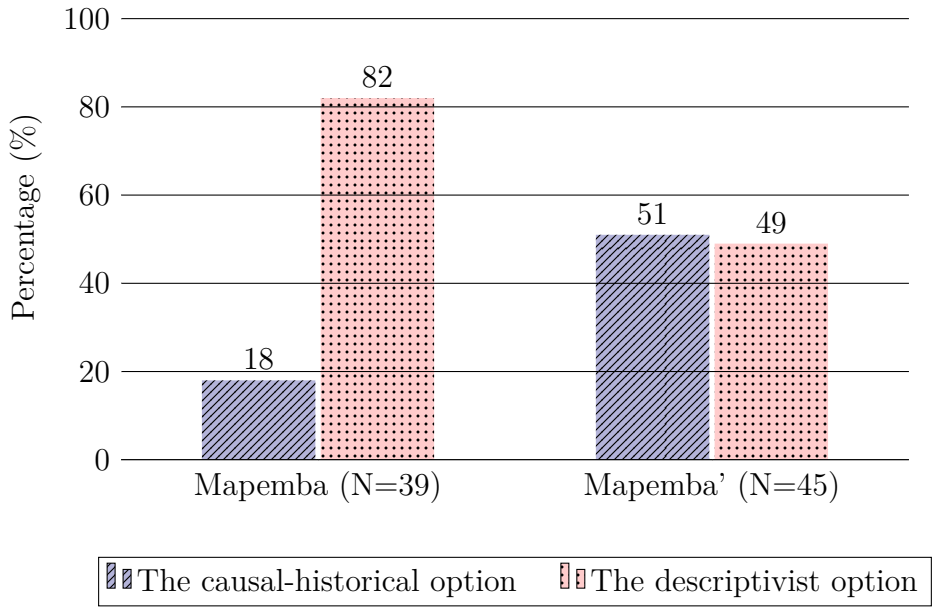
4.6.3.1. Respondent reactions to *Mapemba* and *Mapemba'* depending on the particular version

As part of the following statistical analysis, I compare respondent reactions to the *Mapemba* scenario (where the name used by the protagonist differs in shape from the name borne by the person who inspired the legend, here called Ndembo) and the alternative *Mapemba'* scenario (where the name used by the protagonist has the same shape as the name borne by the person constituting the possible referent of that name). A separate comparison was carried out for each version, that is, for each formulation of the descriptivist option.

Graph 9 illustrates the response distribution for the alternative scenarios in the case of the *Fictional Person* version. Let us remember that in the case of the original scenario responses stating reference to a fictional person predominated. In the case of the alternative scenario, where the leader was in fact called *Mpemba*, in contrast, the response stating reference to a fictional person was almost as popular as the one pointing to the ingenious leader as the referent. Differences in response distribution for the alternative scenarios are statistically significant[21].

Sameness of shape of the name used by the protagonist and the name borne by the person who inspired later false beliefs shared by the protagonist had a similar impact in the case of the *No One in Particular* version. Although here responses supporting the causal-historical theory predominated both for *Mapemba* and *Mapemba'*, in the first case this predomination was clearly smaller than in the latter case, where the two names had the same

[21] $\chi^2(1) = 10.01$; $p = 0.002$.



Graph 9. Response distribution for *Mapemba* and *Mapemba'* in the case of the *Fictional Person* version.

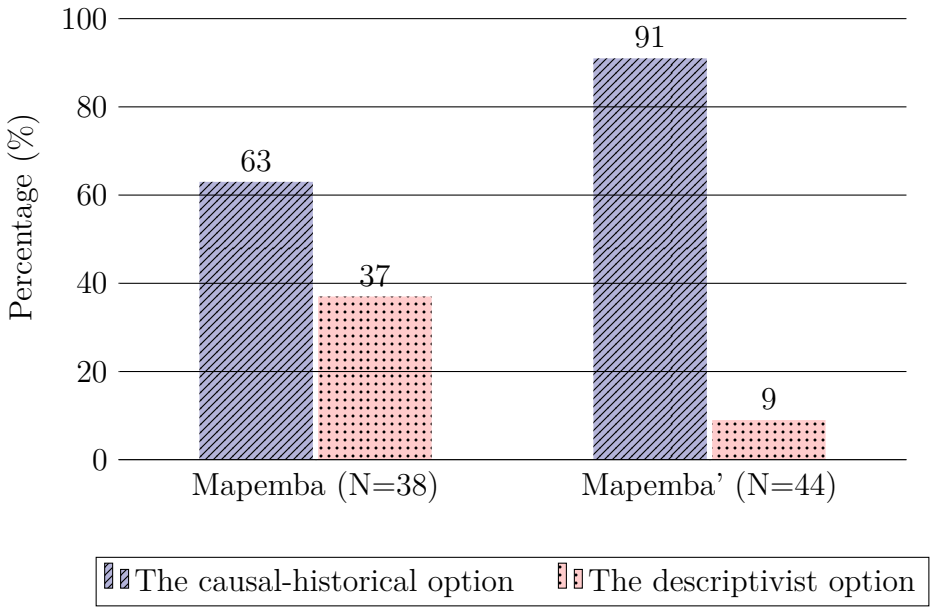
shape. Just like in the case of the *Fictional Person* version, here too, the difference is statistically significant[22]. The result is illustrated on Graph 10.

At the same time, no divergence in respondent preferences occurred in the case of the *No Reference* version[23]. Here, a clear majority of respondents concluded that the protagonist of the story referred to the ingenious leader whose achievements inspired legend regardless of whether the shape of the name used by the protagonist and the shape of the name borne by the leader were the same or not. This result is shown on Graph 11.

4.6.3.2 Respondent reactions to *Homer* and *Homer'* depending on the particular version

The distribution of respondent preferences in regard to *Homer'*, where the name borne by the person who inspired later generations of poets to create the *Odyssey* was in fact *Homer*, in the case of the particular versions was similar to that observed for *Mapemba'*.

[22] $\chi^2(2) = 9.17$; $p < 0.002$.
[23] $\chi^2(1) = 0.02$; not significant.

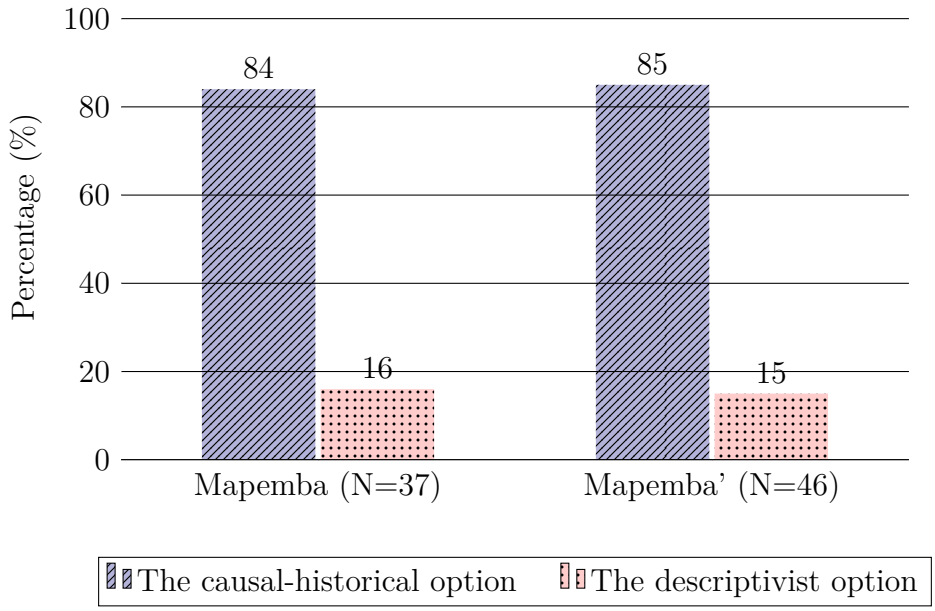


Graph 10. Response distribution for *Mapemba* and *Mapemba'* in the case of the *No One in Particular* version.

In the *Fictional Person* version, participant responses were divided almost in half. To recall, the original scenario, where the story teller who inspired the creation of the *Odyssey* was called Callicrates, led the majority of respondents to conclude that the protagonist referred to a fictional person. Differences between the two alternative scenarios were at the level of a statistical trend[24]. The relevant result is shown on Graph 12.

Graph 13 shows the distribution of results for the *No One in Particular* version. Here, a slight difference between *Homer* and *Homer'* led to the complete reversal of respondent preferences. In response to the original articulation, where the names differed in shape, the majority of respondents concluded that the user did not refer to anyone in particular. The altered scenario, where the story teller who inspired the creation of the *Odyssey* was in fact called *Homer*, in contrast, led the majority of respondents to conclude that the protagonist referred to this person. These differences turned out to be statistically significant[25].

[24] $\chi^2(1) = 3.56$; $p = 0.059$.
[25] $\chi^2(1) = 17.29$; $p = 0.001$.

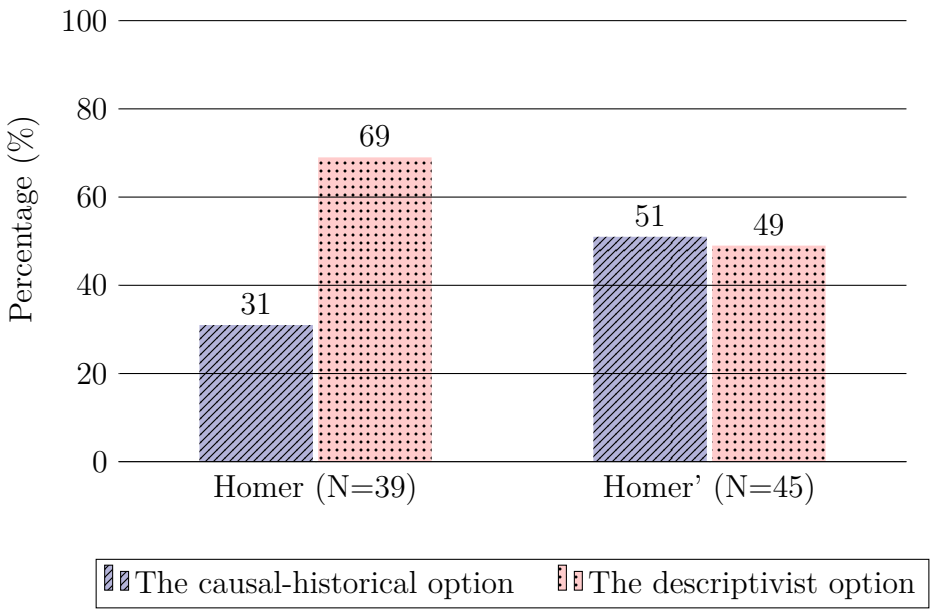


Graph 11. Response distribution for *Mapemba* and *Mapemba'* in the case of the *No Reference* version.

The change in *Homer'* relative to *Homer* also affected participant responses in the case of the *No Reference* version – a statistically significant divergence depending on the shape of the name of the person constituting the possible referent was observed[26]. When this person bore the name "Callicrates," no response predominated (slightly more than a half of respondents concluded that the protagonist did not refer to anyone); however, when his name had the same shape as that used by the protagonist, the opinion that the protagonist referred to that person clearly predominated. This result is shown on Graph 14.

To summarize the results of my second methodological experiment, one could say that, according to predictions, sameness of shape of the name used by the protagonist and that belonging to the person who inspired legend led to a significant increase in the percentage of responses supporting the causal-historical theory relative to the scenarios used in the first experiment, where these names differed in shape. In the case of *Homer* this tendency was observed for all three versions. In the case of *Mapemba*, it was observed for the *Fictional Person* and *No One in Particular* versions. It should perhaps

[26] $\chi^2(1) = 10.73$; $p = 0.001$.



Graph 12. Response distribution for *Homer* and *Homer'* in the case of the *Fictional Person* version.

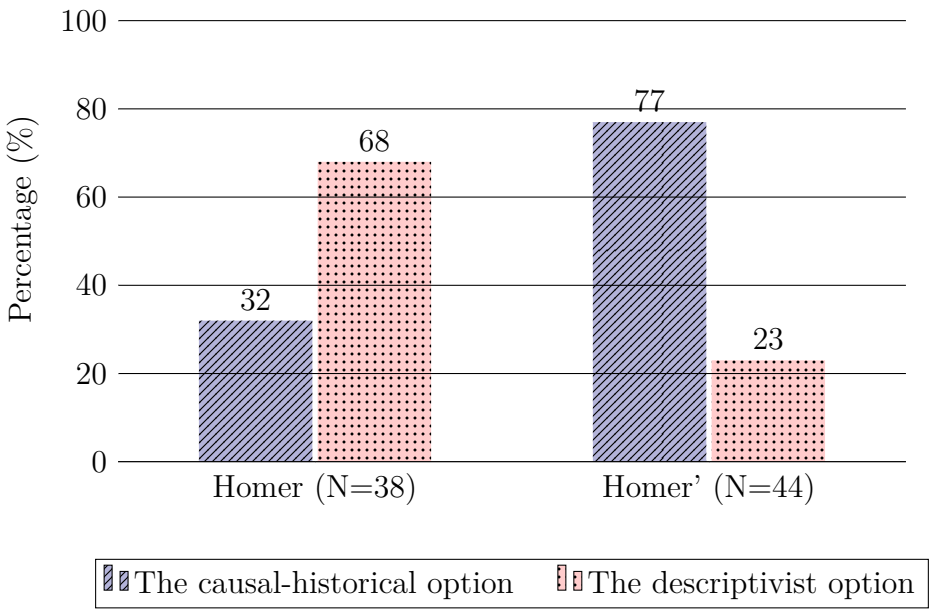
be emphasized that in some cases (*Homer*, the *No One in Particular* version) the manipulation of the content of the scenario had such an immense impact on participant responses that it led to the complete reversal of preferences.

5. Results

The picture emerging from the data gathered during my studies is fairly complex, especially in light of the results obtained in previous research concerning the reference of proper names. In what follows, I present two different interpretive strategies leading to different conclusions. At the same time, I should point out that these two approaches certainly do not exhaust the set of all consistent explanations of the obtained data. I should also stress that the discussed results ought to be approached cautiously since the conclusion is based on limited empirical material.

5.1 The first interpretive strategy: the data support the position of Machery *et al.*

One possible interpretation of the data presented here is that they in fact support the main thesis put forth by Machery *et al.* (2004). To

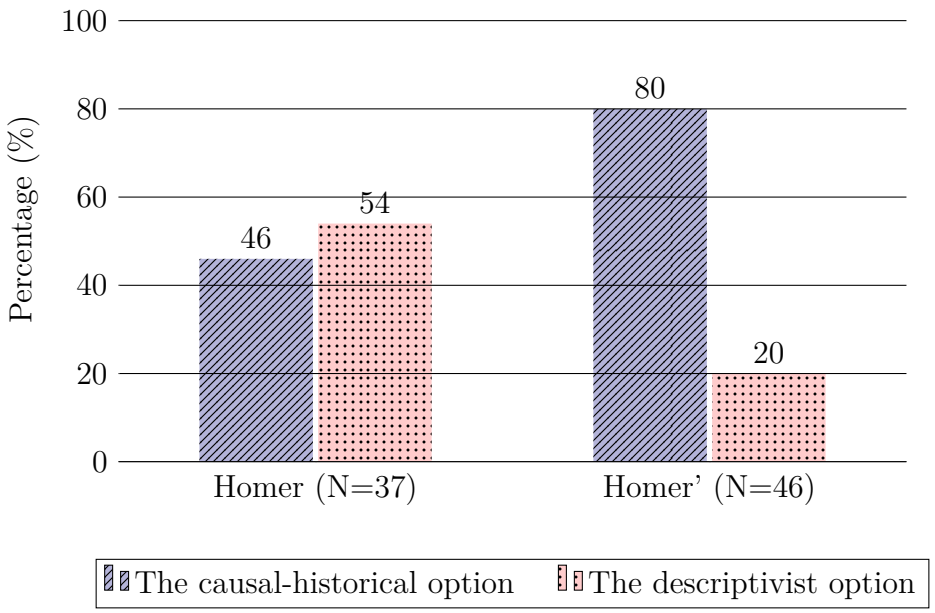


Graph 13. Response distribution for *Homer* and *Homer'* in the case of the *No One in Particular* version.

recall, MMNS claim that intuitions about the reference of proper names are susceptible to the impact of philosophically insignificant factors. In the case of their research, this factor was the cultural background of persons expressing opinions regarding reference. My research is not cross-cultural and thus does not constitute further evidence regarding the impact of this factor on semantic intuitions. However, the results of my study, targeting predominantly members of one nation (citizens of the USA), demonstrate an intracultural and even intrapersonal variation in intuitions about the reference of proper names. This variation seems to be due to factors that should not impact such intuitions[27]. In particular, no divergence should have occurred between respondent preferences in regard to the different Jonah type scenarios used in the study since, as far as philosophically significant aspects are concerned, these scenarios are similar. Such divergence was nonetheless noted.

Before I develop and assess the idea that the obtained data could be interpreted as supporting MMNS, I must point to one aspect that constitutes

[27]For more detail regarding these factors and their impact, see Section 5.2. dedicated to the discussion of the second interpretive strategy.



Graph 14. Response distribution for *Homer* and *Homer'* in the case of the *No Reference* version.

a serious problem for them. Namely, it turned out, according to predictions, that changes in the formulation of the descriptivist option relative to the formulation used in the original MMNS study led to a variety of different participant responses. It is thus not possible to claim that by choosing alternative formulations of the descriptivist option the respondents expressed the same intuitions – not every formulation permitted the respondents to express their support for descriptivism. As I have argued earlier, there are independent reasons to maintain that the option stating reference to a fictional person used in the study by MMNS is in fact inadequate. Therefore, it is justified to claim that their experiment – at least in the part focusing on Jonah cases – did not measure the preference of nonphilosophers for the causal-historical *versus* the descriptivist theory but their preference for the first versus some other conception of the reference of proper names (one allowing for the possibility of referring to a fictional object).

Moreover, even assuming that I have used the right operationalization of descriptivism in Jonah type situations is either the *No One in Particular* or the *No Reference* version, respondent preferences in these cases are still ambiguous and unstable across the tested scenarios. In the *No Reference*

version, where the formulation of the descriptivist option is perhaps closest to the spirit of descriptivism, preferences for the causal-historical theory clearly predominated in the case of *Mapemba* and *Einstein*; at the same time, no response predominated in the case of Homer. Therefore, if this is considered to be the right approach to determining which responses support the descriptivist theory and which accord with the causal-historical conception in Jonah cases, it does not eliminate the instability characterizing the responses of nonphilosophers – not so much on the cross-cultural level as within one culture and even between the individual assessments made by the same person. Hence, the dose of suspicion MMNS propose to apply to semantic intuitions seems even more justified in light of my results.

The results of my research can thus be read as supporting the main thesis put forth by Machery *et al.*: that semantic intuitions about the reference of proper names are unstable and uncertain and therefore useless in philosophical debates. However, let us note that their conclusion is based on a hitherto unquestioned assumption that, in philosophical-experimental research of the type similar to theirs, by choosing one of the available response options the respondents do in fact express support for certain theories of reference. Their explanation for the cross-cultural differences observed in much research is consistent with this assumption. The reason for an increased tendency among the Chinese, relative to the Americans, to express intuitions supporting the descriptivist theory is supposed to be a tendency prevalent among the latter (and characteristic of all members of Western culture) to perceive reality in causal terms. The preference for the causal-historical theory among members of Western culture is thus presumably an effect of a more general phenomenon, namely, a preference for a particular cognitive strategy characteristic of that culture.

The key feature of the argument presented by Machery *et al.* is the observation that the majority of analytic philosophers, who tend to support the causal-historical conception, are persons raised in Western culture. Since according to MMNS there is no basis for the claim that philosophers are not susceptible to the influence of culture, it is safe to assume that their preferences too are culturally conditioned, especially since the majority have been raised in the Western cultural sphere. And here lies the crucial problem – the results of my experiments (and research results showing relevant differences within Western culture in general) do not warrant such an easy and fluid passage from tendencies observed in nonphilosophers to

alleged tendencies in the semantic intuitions of trained philosophers[28]. To wit, I observed clear intrapersonal differences in the responses of nonphilosophers to different Jonah type scenarios. I am of the opinion that a similar effect would not occur in philosophers – the commitment to avoid contradiction and to maintain the consistency of expressed views is one of the top priorities of the philosophical academy (at least in so-called analytic philosophy). Moreover, it does not seem plausible that the impact of factors affecting participant responses observed in my experiments could be explained in a manner tracking one of the discussed theories of reference as closely as the cross-cultural explanation proposed by MMNS. Let us consider the strong impact of the shape of the name used by the protagonist on participant responses observed in my experiment. Whereas it can be expected that the tendency to support the causal-historical theory of the reference of proper names should correlate positively with the tendency to perceive reality in causal terms, the impact of the shape of the name on preferences for either the causal-historical or the descriptivist conception is unexpected. As we have seen, sameness of shape of the name used by the protagonist and that borne by the person constituting the possible referent of that name translates into a significant increase in the percentage of responses interpreted as supporting the causal-historical theory. Meanwhile, this factor should not significantly impact preferences for this theory since a change in the shape of the name does not change the causal chain linking its use to the referent (if such an object exist) – the most important feature of that conception. The case of descriptivism is similar. Here, the description tied to the name is crucial since it is the description that determines the referent; the shape of the name does not play any role. The impact of the shape of the name on preferences guiding the selection of either the causal-historical or the descriptivist theory is thus difficult to explain. A doubt therefore arises: do the participants in experiments based on methodologies similar to that used by MMNS (the choice of determinate response options) in fact express support for either descriptivism or the causal-historical theory, if only in a trivialized version reconstructed by MMNS?

If the above line of reasoning is correct, then another strategy of interpreting the data gathered during my research must be pursued. According

[28]To recall, another problem connected to this explanation is the fact that its correctness is doubtful. As has turned out in the course of the experiment by Machery *et al.* (2009), the French, that is, members of Western culture, supported the causal-historical theory in the case of Gödel situations less willingly than did Mongols, members of Asian culture.

to the alternative interpretation presented below, methodologies modelled on the approach adopted in the study by Machery *et al.* (2004 and 2009) are not effective as far as revealing the semantic intuitions of nonphilosophers is concerned. This is because many participants in studies of this kind express intuitions that are not semantic.

5.2 The second interpretive strategy: the data do no reflect semantic intuitions

According to the second interpretive approach, the data gathered during my methodological experiments concerning the reference of proper names can be used to undermine the key assumption shared by the authors of research to date: that by selecting a particular response in Jonah type situations the respondents in fact express their support for either descriptivism or the causal-historical theory (this conclusion could perhaps be generalized to include Gödel cases; this would require further empirical research). The reason for this interpretation is the instability of participant responses observed in the course of my experiments. In what follows, I explain in detail which aspects of the scenarios used in my studies translated into differences in respondent preferences and what their impact might consist in. However, before I proceed to this detailed interpretation, I would like to characterize the general mechanism I consider to be responsible for participant reactions to experiments based on methodologies modelled on the study by MMNS.

According to the second interpretive strategy, at least some of the persons participating in my studies selected the responses they did not due to their semantic intuitions about the reference of proper names but due to heuristics based on simple associations: they focused on certain "superficial" verbal tips embedded in the individual scenarios. Both Gödel and Jonah cases are complex and exceptional situations manageable by philosophers but not by persons lacking philosophical training. It can thus be expected that the semantic intuitions of the latter are not well grounded and simply break down when such uncommon situations are considered, giving no basis for unambiguous and certain responses. It is more than likely that many nonphilosophers, when confronted with stories they found troublesome and lacking clear guidance stemming from their linguistic competency, tried to imbue these stories with some sense by associating them with other scenarios, previously encountered in daily life, and by adopting simplified coping strategies.

There are many reasons for claiming that the responses of at least some respondents participating in my experiments resulted from superficial

information processing and simplified techniques of analysis. In my opinion, almost all systematic differences in respondent preferences between the particular Jonah scenarios and the particular versions of these scenarios are of this sort. This is especially striking once the most salient and commonsensical explanation of these differences is taken into account.

Let us consider my first experiment and differences in participant responses to each scenario in the *Fictional Person* version. How can we explain, in simple terms, the stark difference in participant responses between *Einstein* on the one hand, and *Mapemba* and *Homer*, on the other? The construal of the content of *Mapemba* and *Homer* activated in the respondents a cognitive schema for processing fictional and legendary persons. The description tied to the name in the first story mentioned the possession of magical powers, a feature considered by most to be fictional in and of itself; the second story referenced the hypothesis still being explored in the history of literature that Greek epics were not created by single author and that Homer is a legendary figure (many of the respondents may have learnt about this hypothesis in the course of their education). The Einstein scenario, in contrast, did not contain such elements (here, an erroneous belief of the linguistic community did not result from a long process of collective legend making but a single person's mistake) – the potential referent here is a person known to the majority of the respondents (including from photographs), a person whose existence was thus most likely not in doubt. The impact of this element was in fact so strong that the response stating reference to Albert Einstein predominated regardless of the formulation of the competing response (an effect which did not occur for the other two scenarios). It is thus possible that the basis for the respondents' reactions to the three scenarios consisted in superficial associations triggered by these scenarios. As my second experiment demonstrated, another factor responsible for significant differences in responses to these scenarios (albeit not for their entirety) was the fact that in *Mapemba* and *Homer*, but not in *Einstein*, the name used by the protagonist differed in shape from the name borne by the possible referent. Here too, it seems justified to conclude that the impact of this manipulation was based on a simple mechanism of association.

Respondent preferences in the case of the *No One in Particular* version can be explained in a similar manner. The response option stating reference to no one in particular was preferred by the majority of respondents for *Homer*; in the case of *Einstein* a minority preferred it, although it was still the most popular of all three formulations of the descriptivist option. This is likely related to the fact that both these scenarios, unlike *Mapemba*, clearly

indicated that the description tied to the key name by the protagonist can be truthfully predicated of all members of a group even if it did not unequivocally point to any one individual. It seems that an association between the relevant situation and linguistic practices involving the expression "no one in particular," especially pronounced in the English language, was decisive here.

The marginal popularity of the descriptivist option in the No Reference version is not surprising in this context either. If the majority of respondents perceived the scenarios in light of typical situations – situations featuring successful communication between interlocutors – it is possible that they rejected the response stating reference failure "in advance," without further analysis, regardless of the user's explicit intention. This preference may have been additionally strengthened by a familiar pragmatic phenomenon consisting in the tendency on the part of the recipient of an utterance to interpret it so that it possesses the desired semantic value (most often truth) – this can be linked to such philosophical concepts as Lewis's principle of accommodation (e.g. Lewis, 1979) and Davidson's principle of charity (e.g. Davidson, 1973).

On the other hand, one must admit that the response distribution for the *Homer* scenario in the *No Reference* version turned out to be slightly surprising. Here, the response stating reference failure was as popular as the one indicating reference to the story teller who inspired the creation of the *Odyssey*. However, the *Homer* scenario differs from the other two stories in a way which could provide a basis for the observed differences in respondent reactions. The specificity of the *Einstein* scenario relative to *Homer* (and *Mapemba*) consists in the fact that respondent reactions (the majority of respondents supported the causal-historical theory for all three versions of the scenario) could be shaped, to a large degree, by the conviction that all uses of the name "Einstein" refer to Albert Einstein – it is possible that they did not see any other option as attractive. The difference between *Homer* and *Mapemba*, on the other hand, could stem from the fact that whereas the protagonist of *Mapemba* shares certain adequate beliefs concerning the possible referent of the name he uses, the protagonist of Homer does not share any such beliefs. According to the latter scenario, Jacques believes that Homer is the author of the *Odyssey* and he does not have any other beliefs regarding this person – the description is false since, as it turns out, the *Odyssey* simply does not have an individual author. The protagonist of the *Mapemba* scenario, in contrast, not only believes that Mapemba was a shaman imbued with magical powers but also that

he was the founder of Kwende. As it turns out, even though the founder of Kwende was not a magic yielding shaman and his name was not Mapemba, he nonetheless existed. It is thus possible that the respondents assumed that the protagonist of the story successfully referred to the leader Ndembo since one description associated by him with the name "Mapemba" – the description "the founder of Kwende" – does identify Ndembo. The situation in *Homer* is different; here, the protagonist cannot be considered to share any belief correctly identifying Callicrates. It must be stressed at this point that this explanation of the differences between *Homer* and *Mapemba* points to questions of substance, relevant from the semantic perspective – it is thus not the sought-after pragmatic explanation of superficial differences between the scenarios. Despite this minor exception, the explanation of the observed differences presented here provides strong evidence for the conclusion that a crucial role in shaping participant responses in my experiments was played not so much (or not entirely) by the respondents' semantic intuitions as by (broadly construed) pragmatic phenomena.

I am of the opinion that there is another reason to trust the explanation according to which participant responses are shaped based on superficial associations. Namely, the method of measuring respondent opinions used in research to date makes it difficult for the respondents to express their semantic intuitions elicited by the particular scenarios. Both the participants of the pioneering study by MMNS (2004) and those of many later experiments dedicated to the same issue (including the participants of my experiments) were forced to choose between two response options. The differences between the particular versions, differing in the formulation of one of the options, observed in my experiments demonstrate that many alternative and potentially attractive responses to Jonah scenarios can be indicated. The measurement of semantic intuitions, as this is the intention here – based on forced choice of one of two options limits the free expression of respondent opinion[29]. On the one hand, this fact could be demotivating, discouraging the respondents from an in-depth consideration of the presented problems. As I have noted earlier, considering problems such as Jonah type situations, lying outside the scope of non-philosophical experience, is demanding enough – limiting the freedom of expression certainly made the task even more daunting. On the other hand, if in the opinion of the respondent none of the proposed response options correctly characterized the reference of the name in the described

[29]I received several emails from the participants of my experiments registering the complaint that none of the proposed options reflected their intuitions, making it impossible for them to express their actual intuitions.

situation, but the respondent was forced to choose one option, it should be expected that they sought some kind of justification for their choice. But this justification need not be related to the respondent's semantic intuitions; it could refer to some other beliefs shared by the respondent or to aspects of the scenarios that are not necessarily relevant to the reference of proper names. Some candidates for such justifications have been indicated in the above attempt to explain the source of the differences observed in the course of my experiments.

It should be emphasized once more that my studies were methodologically similar to earlier research dedicated to the problem of the reference of proper names. The impact on respondent preferences was achieved by introducing minor differences into the content of the presented scenarios or the formulation of the response options available to the respondent – differences that should not influence, at least in theory and certainly not as strongly, semantic intuitions. Therefore, I am of the opinion that the key conclusion I have argued for could be generalized onto other experiments aimed at measuring nonphilosophers' opinions concerning the reference of proper names. Most probably not all participants of the studies carried out by Machery et al. (2004 and 2009) based their responses, including responses to Gödel cases, on their semantic intuitions, at least because the measuring tool employed in these studies was not sensitive enough to allow them to fully express such intuitions. Even on the assumption that, despite these difficulties, each respondent made sure that their responses are in fact close to their semantic intuitions, the interpretation of these choices as supporting either descriptivism or the causal-historical theory is still uncalled for. The spectrum of possible and philosophically consistent responses to Gödel and Jonah cases is undoubtedly far broader than the small range of options proposed by Machery *et al.* Forcing the respondents to choose one of them is unlikely to tell us anything precise about the semantic intuitions of nonphilosophers.

Conclusion

I am of the opinion that the data gathered during my studies and their discussion presented above constitute a relatively strong argument for doubting the methodology employed in research concerning the reference of proper names proposed by MMNS. In particular, the discussion demonstrates that more caution is needed in regard to the assumption, prevalent among

experimental philosophers, that responses to their surveys can be directly linked to philosophical intuitions of a certain kind.

The acceptance of the above argumentation should perhaps lead to the formulation of some additional methodological conclusions. Namely, since the structure of the studies dedicated to the problem of the reference of proper names carried out to date – one based on forced choice as the method of measuring semantic intuitions – limits the participants' freedom to express their opinions, a change in the way respondent preferences are measured should be considered. A natural alternative to forced choice is to give the respondents an opportunity to freely formulate their opinions or to choose from a broader range of options. The first solution is not popular among researchers – not only among experimental philosophers but in every social science reliant on survey methods, because data gathered this way is much more demanding and tedious to analyze than forced choice from a limited number of options. In order to carry out a quantitative analysis of such data it is necessary to group and categorize the free responses of the participants (coding). This task, if bias is to be avoided, should not be carried out by the authors of the experiment but by so-called competent judges – persons not familiar with the goals of the study and the hypotheses posed but competent enough to classify participant responses in a manner useful to the researchers. This further complicates research procedure. Another and perhaps more significant problem related to the analysis of free respondent feedback is the high percentage, noted for studies employing this method of measurement, of responses that are ambiguous, difficult to classify, or simply irrelevant to the question.

In light of the results of my experiments, formulating an open question which would encourage all participants to interpret the presented problem in the same way and to base their response on semantic intuitions (as opposed to some other kind) would be a particularly demanding task. Here, the use of a range of previously prepared responses can help curb ambiguity – in a sense, limiting the spectrum of responses makes the presented problem more determinate and precise, suggesting a perception desirable from the perspective of the experimenters. Therefore, a potentially interesting strategy could be to use surveys allowing the participants to freely express their opinion regarding the reference of proper names to determine a range of attractive and relevant responses which could be presented to respondents in multiple choice format in subsequent research. Such an approach would likely lead to obtaining relatively precise knowledge regarding the semantic intuitions of the respondents based on their reactions even to one scenario.

There is no doubt that further philosophical-experimental studies concerning the reference of proper names require an exploration of novel methodological avenues.

Bibliography

Alexander, J. (2012). *Experimental Philosophy. An Introduction*. Cambridge: Polity Press.

Bealer, G. (1998). Intuition and the Autonomy of Philosophy. In M. DePaul, W. Ramsey (Eds.), *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry* pp. 201–240. New York: Rowman & Littlefield.

Beebe, J. R., Undercoffer, R. J. (2016). Individual and Cross-Cultural Differences in Semantic Intuitions: New Experimental Findings. *Journal of Cognition and Culture*, *16*(3–4), 322–357.

Cappelen, H, (2013). *Philosophy Without Intuitions*. Oxford: Oxford University Press.

Davidson, D., (1973). Radical Interpretation. *Dialectica*, *27*, 314–328.

Deutsch, M. (2009). Experimental Philosophy and the Theory of Reference. *Mind and Language*, *24*(4), 445–466.

Devitt, M. (2011). Experimental Semantics. *Philosophy and Phenomenological Research*, *82*(2), 418–435.

Frege, G. (1997). Sens i znaczenie. In *Pisma semantyczne*, trans. B. Wolniewicz (pp. 60–88). Warszawa: Wydawnictwo Naukowe PWN.

Grice, H. P. (1975). Logic and Conversation. In P. Cole, J. Morgan (Eds.), *Syntax and Semantics*, vol. 3 (pp. 41–58). New York: Academic Press.

Jackman, H. (2009). Semantic Intuitions, Conceptual Analysis, and Cross-Cultural Variation. *Philosophical Studies*, *146*(2), 159–177.

Kim, M., Yuan, Y. (2015). No Cross-Cultural Differences in the Gettier Car Case Intuition: A Replication Study of Weinberg et al. 2001, *Episteme*, *12*(3), 355–361.

Knobe, J., Nichols, S. (2008). An Experimental Philosophy Manifesto. In J. Knobe, S., Nichols (Eds.), *Experimental Philosophy*, vol. 1 (pp. 3–14). Oxford: Oxford University Press.

Kripke, S. (1972). Naming and Necessity. In D. Davidson, G. Harman (Eds.), *Semantics of Natural Language* pp. (253–355). Dordrecht: Reidel.

Lam, B. (2010). Are Cantonese-Speakers Really Descriptivists? Revisiting cross-cultural semantics. *Cognition*, *115*, 320–329.

Lewis D. (1979). Scorekeeping in a Language Game. In: R. Bauerle, U. Egli, A. von Stechow (Eds.), *Semantics from a Different Point of View* (pp. 172–187). Berlin: Springer.

Ludwig, K. (2007). The Epistemology of Thought Experiments: First Person versus Third Person Approaches. *Midwest Studies in Philosophy*, *31*, 128–159.

Ludwig, K. (2010). Intuitions and Relativity. *Philosophical Psychology*, *23*(4), 427-445.

Lycan, W. G. (1999). *Philosophy of Language: A Contemporary Introduction.* London: Routledge.

Machery, E. (2012). Expertise and Intuitions about Reference. *Theoria*, *27*(1), 37-54.

Machery, E., Deutsch, M., Mallon, R., Nichols, S., Sytsma, J., Stich, S. (2010). Semantic Intuitions: Reply to Lam. *Cognition*, *117*, 361–366.

Machery, E., Deutsch, M., Sytsma, J. (2015). Speaker's Reference and Cross-Cultural Semantics. In A. Bianchi (Ed.), *On reference* (pp. 62–76). Oxford: Oxford University Press.

Machery, E., Mallon, R., Nichols, S., Stich, S. (2004). Semantics, Cross-Cultural Style. *Cognition*, *92*, B1–B12.

Machery, E., Olivola, C. Y., De Blanc, M. (2009). Linguistic and Metalinguistic Intuitions in the Philosophy of Language. *Analysis*, *69*(4), 689–694.

Marti, G. (2009). Against Semantic Multi-Culturalism. *Analysis*, *69*(1), 42–48.

Muszyński, Z. (2000). *Komunikacja i znaczenie: semantyczny aspekt komunikacji.* Lublin: Wydawnictwo UMCS.

Nadelhoffer T., Nahmias E. (2007). The Past and Future of Experimental Philosophy. *Philosophical Explorations*, *10*, 123–149.

Searle, J. (1958). Proper Names. *Mind*, *67*, 166–173.

Seyedsayamdost, H. (2015). On Normativity and Epistemic Intuitions: Failure of Replication. *Episteme*, *12*(1), 95–116.

Sytsma, J., Livengood, J. (2011). A New Perspective Concerning Experiments on Semantic Intuitions. *Australasian Journal of Philosophy*, *89*(2), 315–332.

Weinberg, J., Alexander, J. (2014). The Challenge of Sticking with Intuitions through Thick and Thin. In A. Booth, D. Rowbottom (Eds.), *Intuitions* (pp. 187-212). Oxford: Oxford University Press.

Weinberg, J. M., Nichols, S., Stich, S. (2001). Normativity and Epistemic Intuitions. *Philosophical Topics*, *29*(1–2), 429–460.

Williamson, T. (2007). *The Philosophy of Philosophy*. Malden, MA: Blackwell.

Originally published as "Odniesienie nazw własnych, intuicje semantyczne i filozofia eksperymentalna" *Studia Semiotyczne* 30/2 (2016), 45–97. Translated by Joanna Guzowska.

STUDIA SEMIOTYCZNE—ENGLISH SUPPLEMENT, vol. XXIX (2017), 248–271
ISSN 2451-2745
DOI: 10.26333/stsen.xxix.11

Joanna Odrowąż-Sypniewska[1]

Are There Any Subsentential Speech Acts?[2]

Abstract In this paper, I critically examine the major philosophical standpoints regarding (apparent) subsentential speech acts such as "Nice dress", "Under the table", or "Where?". The opponents of this category (e.g. Stanley, Merchant) argue either that apparent subsentential speech acts are ellipses (i.e. sentential) or that they are not full-fledged speech acts. The defenders of subsentential speech acts (e.g. Stainton, Corazza) argue that even though they are not sentences in the syntactic or the semantic sense, they can be used to perform a speech act. I argue in defence of subsentential speech acts and propose to analyze them using Recanati's moderate relativism.[3]

Keywords speech acts, ellipsis, subsentential speech acts, moderate relativism

1. Introduction

In his book titled *Frege: Philosophy of Language*, Michael Dummett holds that "a sentence (...) is the smallest unit of language with which a linguistic act can be accomplished, with which a 'move can be made in a language-game'" (Dummett, 1973, p. 194).[4] Kent Bach considers this thesis to be an idealization (Bach, 2008, p. 739), and takes the fact that people state, propose, ask etc. using bare phrases, even single words, to

[1]Department of Philosophy and Sociology, University of Warsaw.
E-mail: j.odrowaz@uw.edu.pl

[2]Research preceding the writing of this article was financed with a grant from the National Science Centre (UMO-2014/15/B/HS1/00171).

[3]I would like to thank the anonymous referees for their critical remarks which helped make this paper better.

[4]Dummett is a "villain" in the eyes of the proponents of subsentential speech acts. They make frequent references to the quoted passage.

be obvious. He takes Dummett's thesis to be clearly false, to the point that he finds it strange that anyone should try to demonstrate its falsity. Meanwhile, Robert Stainton dedicated most of his extensive monograph *Words and Thoughts* to arguing that apparent subsentential speech acts are indeed speech acts performed using non-sentential utterances. Stainton thinks that we encounter such speech acts on a daily basis and lists the following examples (Stainton, 2006):

1. "Sanjay and Silvia are loading up a van. Silvia is looking for a missing table leg. Sanjay says: "On the stoop"". (Stainton, 2006, p. 5)

2. "Benigno gets into a taxi and says: "To Segovia. To the jail"". (Stainton, 2006, p. 5)

3. John demonstrates a letter he is holding in his hand and says: "From Spain".

4. During a conference, a linguist says to a colleague in order to identify a person entering the room: "Barbara Partee" (cf. Stainton, 2006, p. 6)

5. "Meera is putting jam on her toast. As she scoops out the jam, she says 'Chunks of strawberries'. Anita nods, and says 'Rob's mom'" (Stainton, 2006, p. 115).

6. "I'm at a linguistic meeting. (. . .) There are some empty seats around a table. I point at one and say, 'An editor of *Natural Language Semantics*'" (Stainton 2006, p. 209).

7. I walk into a pub and say to the bartender: "Three pints of lager".

Although in each of these cases the speaker utters a non-sentential expression, there is no doubt, according to Stainton, that a speech act has been performed. Sanjay said that the table leg is on the stoop; Benigno asked the taxi driver to take him to Segovia, to the jail; John reported that the letter came from Spain; the linguist said that the person entering the room is Barbara Partee etc. Stainton claims that users of language frequently utter words and phrases that are not complete sentences but whose utterance constitutes the performance of a full-fledged speech act (Stainton, 2006, p. 3). In his opinion, these words and phrases are not sentences in the syntactic sense (they do not have sentential syntax) or the semantic sense (they do not express propositions) but are sentences in

the pragmatic sense (they can be used to perform a speech act) (Stainton, 2006, p. 32). A subsentential assertion, for example, is the utterance of a non-sentential expression ("in isolation": not embedded in any larger syntactic structure (Stainton, 2006, p. 11)) with a determinate illocutionary force and a determinate truth-conditional content. A subsentential assertion does not express a proposition but can be used to assert a proposition. This proposition is supposedly what is said: the literal, not implied or suggested, content. Jason Stanley, who does not believe in the existence of subsentential speech acts, defines a subsentential assertion as "an unembedded utterance that is a successful linguistic assertion" (Stanley, 2000, p. 402).[5]

The examples given by Stainton stem from "real life". I am convinced that the reader has encountered thousands of utterances of this sort. There is thus no doubt that, from the point of view of successful communication, it is often sufficient to utter a fragment of a sentence. One can nonetheless doubt, first of all, if such fragments are indeed full-fledged speech acts, and secondly, if they are not hidden sentences.

2. Semantics-oriented standpoints regarding subsentential utterances

2.1. Jason Stanley's principle of "divide and conquer"

Standpoints regarding apparent subsentential utterances can be divided into semantics-oriented and pragmatics-oriented ones (see Stainton, 2006). The proponents of the semantics-oriented standpoints, such as Jason Stanley, Jason Merchant, and Michael Dummett, hold that there are no subsentential speech acts. They claim that since, in the absence of clear semantic rules, context cannot supply constituents directly to the contents of the expressed propositions, utterances that are not complete sentences cannot be speech acts. According to them, the examples given above are either not full-fledged speech acts or are, contrary to appearances, not subsentential. Accordingly, Stanley employs the principle of "divide and conquer" (cf. Elugardo & Stainton, 2004, p. 446), claiming that these examples can be divided into three groups and that none of these groups can serve as a counterexample to Dummett's position quoted at the beginning. In his opinion, the alleged

[5] "An utterance is unembedded if and only if it is an utterance of a non-sentential expression, and it is not part of an utterance of a sentence in which that expression occurs as a constituent" (Stanley, 2000, p. 402).

subsentential assertions are: (1) elliptical sentences; (2) not full-fledged speech acts; or (3) shorthand for sentential utterances.

Stanley thus claims that some of Stainton's examples are ellipses, even though they are not accompanied by any linguistic context. It is usually thought that ellipses cannot occur at the beginning of a conversation – an ellipsis must be preceded by some other utterance serving as an antecedent providing content for its supplementation. But according to Stanley:

> (. . .) explicitly providing a linguistic antecedent by mentioning it is only the simplest way to provide it. There are other methods of raising linguistic expressions to salience in a conversation without explicitly using them. (Stanley, 2000, p. 404).

For instance, an apparent subsentential assertion can be an answer to a question that has not been asked but is obvious given extralinguistic context. Stanley considers the following example.

> Suppose that Bill walks into a room in which a woman in the corner is attracting an undue amount of attention. Turning quizzically to John, he arches his eyebrow and gestures towards the woman. John replies "A world famous topologist" (Stanley, 2000, p. 404).

Even though no question has been uttered, the described extralinguistic context makes it obvious, according to Stanley, that Bill's gesture and quizzical look express the question "Who is she?". John's utterance is thus an ellipsis replacing the sentence: "She is a world famous topologist" (see Stanley, 2000, p. 406).

Here is another example:

> Suppose that a group of friends, including John and Bill, has gone bungee jumping. Every member of the group is watching Bill, who is the first to muster the courage to bungee jump. As Bill is standing eight stories above the water on the platform of a crane, ready to plummet into the water below, Sarah, aware of John's terror of heights, turns to one of the other friends and utters ["John won't"], shaking her head (Stanley, 2000, p. 405).

According to Stanley, in this situation one should not claim that Sara's utterance has no linguistic antecedent. This is because context brings attention to the expression "bungee jump" which can serve as an antecedent for the ellipsis ("John won't bungee jump").

According to Stanley, some other apparent subsentential speech acts are not speech acts at all because they do not have a sufficiently determinate illocutionary force or a sufficiently determinate content. Here, Stanley gives an example of a thirsty man staggering up to a street vendor and saying "Water!". Stanley holds that no determinate illocutionary force can be ascribed to an utterance like that since it is not clear if it is supposed to be an assertion, a request, or an imperative. No determinate content can be ascribed to it either since it is not clear if the proposition expressed by the speaker is the proposition that the speaker wants to drink some water or the proposition that the vendor should give the speaker some water (Stanley, 2000, p. 407). Therefore, an utterance like that is not a speech act.

The examples that cannot be classed under either of these two groups are treated as shorthand by Stanley. For example, if someone utters the words "nice dress" to a women met in the street, it is "fairly clear that an assertion has been made, whose content is a singular proposition about the object in question, to the effect that it is a nice dress" (Stanley, 2000, p. 409). Hence, the expression "nice dress" uttered in such a context is simply shorthand for the sentence "This is a nice dress".[6]

Staley would most probably see examples 1, 2, 4, 5, 6 and 7 as ellipses whose linguistic antecedents are provided by implicit questions, and example 3 – as shorthand.

Stanley's standpoint has attracted serious criticism. Regarding the postulate to treat subsentential utterances accompanied by extralinguistic context as ellipses, it has been objected that extralinguistic context can bring attention to objects but not to expressions referring to these objects. Stainton claims that "non-linguistic context cannot determine a linguistic item" (Stainton, 2006, p. 169). As has already been mentioned, it is usually assumed that an ellipsis must have a linguistic antecedent. Stanley does not reject this assumption but argues that such an antecedent can be provided by extralinguistic context. If, then, Stainton is right that context cannot make salient a linguistic item, the alleged ellipsis does not have an antecedent after all. According to Alison Hall, on the other hand, the fact that subsentential utterances require extralinguistic context in order to be correct does not entail that they are ellipses. There are sentential utterances that also require extralinguistic context (for example, during a book signing, the host might

[6]Compare below.

point to a book and say: "The author's going to be signing copies later.")[7] (Hall, 2009, p. 240).

Regarding the exclusion of subsentential utterances from the category of speech acts due to the lack of a determinate force or content, it has been noted that this strategy is too restrictive – its adoption is bound to preclude many uncontroversial sentential utterances from being counted among speech acts. In particular, it seems doubtful that uncertainty as to whether an utterance is an assertion, a request or an imperative should suffice to exclude it from being a speech act. It seems that uncertainty of this sort also arises in situations featuring sentential utterances, and in such cases there is usually no suspicion of the speech act being defective. There is no reason for the conditions imposed on subsentential speech acts to be more severe than those imposed on sentential utterances. Elugardo and Stainton (2004, p. 466) give the example of a sentence uttered by Mary to Susan, her subordinate and good friend at the same time: "You must turn in your final report before you leave in the afternoon". Given that Susan and Mary are bound by professional and personal relations, it may be unclear if the sentence uttered by Mary is a request, a command, or a description of rules in place at the office. However, we do not want to assume that Mary did not manage to perform a speech act.

Regarding the argument that in contexts featuring subsentential utterances there is always an implicit question, Stainton notes that even if this were the case, one should not assume that such a question can serve as a linguistic antecedent for an ellipsis. This is because questions are not linguistic items, only interrogative sentences are. Even if there is an implicit question in a given context, but no interrogative sentence is uttered, one should not assume that the linguistic form of the implicit question is determinate enough for it to serve as an antecedent. In order to speak of an omission of a fragment one must know its shape, not just its content. If I ask "Who bought the bread?", the answer can be "John" but not "by John", even though the full answer "The bread was bought by John" is correct. This is because a shorthand answer is assumed to "inherit" the structure of the question and not just its content (cf. Merchant, 2010, p. 18). In the example about the topologist described by Stanley, for instance, the asker could have meant the question "What does she do?" rather than "Who is this?". The utterance "a world famous topologist" does provide an answer to the first question, but it cannot be treated as an ellipsis supplemented

[7]The example is controversial since it could be considered to constitute an ellipsis: "The author of this book. . . ," where relevant book is determined by context.

by the structure of this (silent) question since such supplementation would lead to the formation of an ungrammatical utterance: "She [does] a world famous topologist".[8]

Regarding treating some subsentential utterances as shorthand, Elugardo and Stainton argue in their article *Shorthand, syntactic ellipsis, and the pragmatic determinants of what is said* that this strategy is bound to fail.[9] They provide four possible interpretations of the thesis that subsentential utterances are mere shorthand and demonstrate that none of these interpretations can be used by the proponents of the semantics-oriented approach. In their opinion, the fact that one expression is shorthand for another can mean that: (a) the latter could be used instead of the former to achieve the same effect; (b) one is a synonym of the other (on some reading); (c) one is conventionally tied to the other; and (d) the two are not conventionally paired but the speaker intended the hearer to read the first as the latter and to use the latter expression to interpret what was meant (Elugardo & Stainton, 2004, pp. 449–454). Interpretation (a) does not exclude the possibility that the first expression can be a subsentential speech act; interpretation (b) leads to treating many expressions as ambiguous – one must admit that the expression "nice dress", for example, could, depending on context, express a proposition or a property. Interpretation (c) results in the need to postulate numerous linguistic conventions and, moreover, just like interpretation (d), does not deny that speakers use subsentential utterances to perform speech acts (Elugardo & Stainton, 2004, pp. 449–454). Hence, none of these interpretations permits the conclusion that the fact that a subsentential utterance is shorthand for a sentential one means that the former is not a subsentential speech act.

Stanley thinks that the truth-conditional role of context is limited to the resolution of indexicality, broadly construed (Stanley 2000: 401). For this reason, he rejects the idea that, in the case of subsentential utterances, context supplies constituents directly to the content of the proposition

[8]It must be admitted that in many ellipses the omitted fragment is of a different form than its linguistic antecedent. One example of this is the sentence "John plays the piano, and Barb and Zoe, the triangle". However, it seems that in the case of short direct answers to questions the requirements imposed on ellipses are more restrictive (see above). Subsentential speech acts would clearly be closest to ellipses of this sort (if they were to be considered ellipses at all).

[9]Stainton notes that Stanley himself has admitted as much. Incidentally, it might be worth noting that, as has been mentioned earlier, Stanley considers the expression "nice dress" to be shorthand for "This is a nice dress"; Hall, on the other hand, points out that it could be shorthand for "You are wearing a nice dress" (Hall, 2009, p. 237).

expressed (Stanley, 2000, p. 402). This is why he attempts to demonstrate that all apparent subsentential speech acts are in fact either not subsentential or not speech acts. However, his argumentation for the elliptical character of apparent subsentential speech acts requires that context precisely indicates linguistic antecedents for such ellipses. But it is implausible to assume that in such cases context operates strictly within the boundaries determined by the rules of language. Therefore, one may doubt if Stanley's attempt to avoid an undue extension of the truth-conditional role of context does not lead him to ascribe another role to it, one equally problematic for an opponent of contextualism.

2.2. Jason Merchant's limited ellipsis and scripts

Jason Merchant holds that most apparent subsentential speech acts are ellipses, and thus, in fact, complete sentences. Some of them are syntactic ellipses and others semantic ones. An ellipsis is syntactic if the uttered expression is a part of a larger unuttered syntactic structure. As has been mentioned earlier, short answers to questions are one kind of syntactic ellipses. For example, if Beatrice asks "Did you buy the tickets?" and John answers "I did", then his utterance can be treated as an ellipsis. The utterance "I did" is a part of the structure "I did buy the tickets" omitted by John. Merchant claims that many examples of subsentential speech acts can be analyzed using a "limited ellipsis" strategy (Merchant, 2010, p. 25)[10]. According to this strategy, expressions such as *it*, *this*, *that*, *he*, or *she*, accompanied by an appropriate form of the verb *be*, can be omitted as long as their reference is obvious. This strategy can also be applied to examples in which the expression uttered denotes a property belonging to a salient object, and to examples in which it denotes an individual bearer of manifest properties (Merchant, 2010). Hence, "on the stoop" is generated by omitting "it is" from "It is on the stoop" since context makes the matter clear. A similar omission occurs in the case of "This is Barbara Partee" or "This came from Spain".[11]

[10] It is called "*limited* ellipsis hypothesis" because it concerns two cases: one mentioned in the text above and another in which "do it" is elided.

[11] Merchant is aware of certain technical difficulties facing his position (e.g. the fact that apparent elliptical fragments, unlike other ellipses, cannot be embedded in larger structures) but believes that these difficulties are not insurmountable (Merchant, 2010, p. 28 ff.).

Alternatively, these examples can be treated as semantic ellipses. Here, one must assume that apparently subsentential utterances contain hidden free variables to which one must assign an appropriate value based on context. For example, the semantic value of "on the stoop" is "on the stoop (x)", where x is a free variable whose value is determined by context. Merchant emphasizes that:

> no extraordinary appeal to pragmatics is necessary [here] beyond what we already assume: namely that the assignment function is set by the context, not the semantics, but is used to determine the semantic value of an expression in a context. (Merchant, 2010, p. 41).

The role of context here is the same as in the determination of the reference of indexical and demonstrative expressions. The postulate of hidden variables makes it possible to treat the majority of apparently subsentential utterances such as "on the stoop" as expressing propositions. The following example, however, cannot be treated as either a syntactic or a semantic ellipsis:[12]

> A father is worried that his daughter will spill her chocolate milk. The glass is very full, and she is quite young, and prone to accidents. He says, "Both hands" (Stainton, 2006, p. 5).

Stainton considers the "both hands" example to be a particularly good illustration of the use of a subsentential speech act. Since it is difficult to treat the utterance "both hands" as a syntactic or a semantic ellipsis (the assumption that the father is saying "These are both hands" does not make much sense), the example is promising for Stainton. However, it should be noted that this is due to the lack of cases in English. A Polish father would have to say "obiema rękoma" ("both hands" in the instrumental case) instead of "obie ręce" ("both hands" in the nominative case) in this situation. In light of this, one might ask, after Merchant, "Where does the case come from?" (Merchant, 2010, p. 42). This example, therefore, contrary to the author's intentions, is in fact an argument in favour of treating subsentential utterances as fragments of longer sentential utterances. Those

[12]For examples such as "an editor of *Natural Language Semantics*" Merchant proposes a separate analysis, based on the idea of labelling. A label can be the name of the labelled object or the name of another object bound to the first by some pragmatic relation. In the editor example, this relation is "being a chair reserved for" (Merchant, 2010, p. 27).

understanding the father's utterance as a fragment of "Use both hands" ("Trzymaj kubek obiema rękami") have no problem explaining why the father says "obiema rękoma" instead of "obie ręce". Those taking it to be a subsentential utterance, on the other hand, will have a much harder time explaining why the instrumental case is used in this situation instead of the nominative.

Many more examples where the uttered fragments are in cases other than the nominative can be given. Merchant thinks that one can appeal to scripts in order to explain the other case forms occurring in these examples (Merchant, 2010, p. 41). According to him, in everyday conversations we often use scripts, and since these scripts are well known we can use fragments of longer utterances – the speaker uses a script and their utterance is a fragment of a larger whole featuring in the script. Possibly the utterance "Both hands" is an ellipsis which might be supplemented by anyone familiar with the appropriate script. Similarly, a person saying "Water!" utters only a part of a script which in its entirety says "Give me some water!" or "I'll have some water, please". Merchant thus assumes that the speaker utters the appropriate fragment in a grammatical form that fits the script deemed by them as befitting the situation. The speaker assumes that their interlocutor knows this script and will be able to supplement the utterance:

> In following a script, the participants know and can anticipate the actions (including the utterances) of the others following the same script, and can plan accordingly (. . .). In such a context, certain particular linguistic phrases can be expected: they are 'given', though not by the immediate actually spoken linguistic precedents, but rather by mutual knowledge of the script being followed (Merchant, 2010, p. 44).

The conception of scripts thus treats subsentential utterances as fragments of sentences whose other parts remain unspoken but are available for all participants in a given conversation because they all follow the same script.

Hall points out that Merchant's analysis proves inadequate in the case of some examples, especially examples featuring names, such as "Rob's Mom" or "Nova Scotia" (see below). Let us remember that Anita, who says "Robert's Mom", means that it was Rob's mother who made the jam in which Meera found chunks of strawberries. Anita only uttered the nominal phrase. This phrase cannot be treated as a syntactic ellipsis since such an

assumption would yield the utterance "This is Rob's Mom". It also seems that there is no ready script here for Anita to use.

Hall emphasizes that in many situations we are forced to appeal to pragmatic inferences not only to ascribe values to hidden indexical expressions featured in the logical form but also to choose the logical form itself (see Hall, 2009, p. 249). This can be seen in the case of the call "Water!", for instance. In this situation, no one expression is "given" as undoubtedly intended by the speaker. It is often the case, especially in situations featuring a subsentential utterance at the beginning of a conversation, that context does not point to unequivocally determined linguistic material. Multiple supplementations are possible. For example, in a situation where someone says "John's father" and points to a man on the other side of the room, the following supplementations are possible: "this is,' "this person is," "this man is," "the person I am pointing to is," 'the person that has just entered the room is," "... has just entered" etc. (cf. Hall, 2009, p. 243). In Hall's opinion, this suggests that the utterance was not a fragmentary sentence: it was not generated by excluding certain expressions and its supplementation is not a matter of reconstructing its true logical form.

3. Pragmatics-oriented standpoints regarding subsentential utterances

3.1. Robert Stainton's Neo-Russellian propositions

Stainton is a proponent of a pragmatic analysis of subsentential utterances. He thinks that such utterances are indeed subsentential and that they can be speech acts. Subsentential assertions have semantic truth-conditional content that is asserted, not just implied, by the person performing the speech act. Stainton holds that:

> The propositional content of subsentential speech acts is arrived at by grasping (a) a content from language, and (b) a content from elsewhere, which is never translated into natural language format (Stainton, 2006, p. 156).

Let us consider the example of Sanjay uttering the expression "on the stoop". According to Stainton, Sanjay states a *de re* proposition about a table leg that it is on the stoop. The assumption that a proposition has been stated is due to our intuition that Sanjay could be right or not; he could also be

lying. Stainton claims that the proposition is stated through the utterance of an expression which, syntactically speaking, is a bare prepositional expression not embedded in any larger syntactic structure. Its meaning is a property. Semantically speaking, this expression is thus incomplete and must be assigned an argument. This argument is provided by context: it is an object salient in this context. It is therefore the object itself and not its name that serves as the argument here. The argument and the function are combined in Mentalese – it is a combination of two mental representations stemming from different sources: the representation of the property stems from the decoding of an appropriate linguistic signal; the representation of the object comes from a source other than language (memory, sight, understanding intentions of agents etc.) (see Merchant, 2010, p. 9). Propositions stated by speakers using subsentential utterances are thus Neo-Russellian propositions featuring extralinguistic objects as their constituents.

The author of *Words and Thoughts* distinguishes two cases: (1) the speaker utters an expression whose content is a propositional function the argument for which is provided by context (as in the case described above), and (2) the speaker utters an expression whose content is an argument the function for which is provided by context. This function is not reducible to a demonstrative function (such as Merchant's "this is x"). Stainton gives the following example:

> After two weeks of cold and rainy weather in mid-summer, in a part of Canada that is usually hot and sunny, Brenda ran into Stan. Brenda looked up at the sky and said "Nova Scotia" (Stainton, 2006, p. 6).

The function provided by context in this situation is "the weather here is similar to…".

Here again the translation of Stainton's example into Polish turns out problematic since the function "the weather here is similar to the weather in…" requires an argument in the appropriate case (the locative), and the expression uttered by Brenda is in the nominative form. However, this situation is different from that of the father saying "Both hands" since here, even though the most probable function provided by context is indeed "the weather here is similar to…", Brenda would say "Nova Scotia" in the nominative rather than the locative even if she spoke Polish.

Since the example featuring the expression "Both hands" has been treated as an argument in favour of ellipses and implicit scripts, this example should be treated as an argument in favour of the pragmatics-oriented

standpoint. Even though there is an implicit linguistic context here which could be seen as a ready script, Brenda's utterance does not fit this context and yet it does not seem incorrect. Stainton could claim that Brenda's utterance provides an argument for a function determined by context. A grammatical discrepancy is not a problem here since – to recall – Stainton thinks that the content provided by context is not formulated in natural language. It is a *de re* content grasped in Mentalese; it need not be a concretely articulable function. Stainton himself gives a similar argument in favour of his conception: Hans and Franz play by exhibiting different objects and saying who these objects remind them of. They part and meet again several days later. Hans points to an old beer-stained table and says "My father". Although it is clear to Franz that Hans said that the table reminds him of his father, the utterance is in the nominative and not in the accusative (Stainton, 2006, p. 107).

Examples such as "Nova Scotia" or "My father" pose a serious problem for the proponent of scripts since she must find a script accommodating a nominative form. Otherwise, the utterances made by Brenda and Hans are bound to be seen as counterexamples undermining her position.

3.2. Eros Corazza's situated unenriched illocutions

Eros Corazza refers to a conception by John Perry, in particular, to the latter's distinction into objects that a proposition is about and objects that a proposition concerns. In the classic text *Thought without Representation* Perry notes that thoughts and propositions can concern objects that correspond to no constituents of the sentences expressing them. For example, if I look out the window and utter the sentence "It is raining", the proposition I express will concern the place I am at while uttering that sentence even though the place is not a part of the content of the relevant proposition (it is not its constituent). Similarly, if a child says "It is three", the proposition she expresses concerns a particular time zone even though the child may be unaware of the existence of time zones. The appropriate parameters (e.g. PM, *Central European Time*) are provided by the situation in which the sentence "It is three" is uttered (Corazza, 2011, p. 566).

Corazza also makes use of the distinction between reflexive and incremental truth conditions introduced by Perry, and the latter's thesis about the multiplicity of propositions semantically related to sentences. Reflexive truth conditions are conditions based on linguistic conventions. For sentence (1) "Jane smokes cigars", these conditions are the following:

(1) A. There is an individual x and a convention C such that:

(i) C is exploited by (1);

(ii) C permits one to designate x with "Jane";

(iii) x smokes cigars (Corazza, 2011, p. 563).

Incremental truth conditions are what is said: the semantically expressed proposition, in this case the proposition that Jane smokes cigars. In order to grasp the reflexive conditions of an utterance, it is enough to know the language – it is not necessary to know the context of the utterance or to perform any pragmatic inferences. In particular, it is not necessary to grasp the proposition expressed by that utterance (that is, its incremental conditions).

Corazza follows Perry in adopting the principle of cognitive economy and claims that "since in many cases the situation fixes all that needs to be fixed, the speaker and her audience need not represent what their discourse concerns" (2011, p. 567). Here, one should distinguish the contextual dependence of sentences from their situational dependence. A sentence depends on the situation if its logical value depends on it, but the speaker need not have a representation of this situation. If a sentence is indexical, on the other hand, and depends on context, the speaker must have a representation of the latter in order to be able to determine the reference of the relevant indexicals.

A lot of information can be "stored" in situations or in our long-term memory. This information allows us to act successfully without us having to articulate it in our thought (Corazza, 2011, p. 567). Subsentential speech acts are a very good illustration of this phenomenon, according to Corazza. Let us consider the following situation:

> John, a well-known anti-Fregean, has been told that Jane is desperately looking for Dummett's *Frege: Philosophy of Language*. Jane walks into John's office. John suddenly utters: [2] Hidden on top of the shelf (Corazza, 2011, p. 570).

In this situation, Stainton would assume that John made an assertion whose content is the proposition that Dummett's book is hidden on top of the shelf. However, as is noted by Corazza, Jane need not have been thinking about the book as she walked into John's office. She may have even forgotten that she had been looking for it. Since the book is hidden and

Jane is unable to see it, it seems that the relevant proposition cannot be *de re* about this book. What Jane can do, according to Corazza, is grasp the reflexive truth conditions of John's utterance, which are:

There is an x such that:

(i) (2) concerns x;

(ii) x is hidden on top of the shelf (Corazza, 2011, p. 575).

It is enough that such truth conditions are conveyed and grasped in order for the communicative goal to be achieved. Even if Jane does not remember that she had been looking for Dummett's book, if she grasps the reflexive conditions of John's utterance, she will be able to reach out to the top of the shelf and find the book there. Corazza holds that a speech act can be successful even if it is not accompanied by pragmatically enriched thoughts possessing truth conditions. The situatedness of the utterance is enough for it to have truth conditions. The utterance "hidden on top of the shelf" is not an ellipsis, according to Corazza, nor does it require pragmatic enrichment. All the necessary information is stored in the situation and need not "enter into" the utterance. Corazza's conception thus differs significantly from Stainton's position since the latter holds that propositions stated by the speakers of subsentential assertions have contents stemming from two sources: a linguistic utterance and a context. According to Stainton, content unenriched by context does not have truth conditions and does not constitute a proposition. A successful subsentential speech act must have a pragmatically enriched content. Meanwhile, Corazza claims that:

> in distinguishing between reflexive truth conditions and incremental truth conditions we can deal with successful communication involving subsentential speech without appealing to ellipsis and/or enrichment. And we can do so only by considering our thoughts and utterances as situated (Corazza, 2011, p. 577).

It should be emphasized here that Corazza is more interested in successful communication than in the assertoric content of utterances used in such communication. He considers the reflexive truth conditions described above to be sufficient for successful communication and writes explicitly that it does not matter if Jane grasped a *de re* thought about Dummett's book or a general thought about something being hidden on top of the shelf (Corazza, 2011, p. 578). The only factor important for successful communication in

this case is whether the hearer's action resulting from the conversation is in accord with the speaker's intention. Since we can engage in action based on general thoughts – i.e. on reflexive truth conditions – successful communication does not require that *de re* thoughts be conveyed. As we have seen, in Corazza's opinion, reflexive truth conditions are sufficient for successful communication, and it is not necessary that incremental truth conditions be conveyed or grasped. If, then, one assumes, as is usually done, that the asserted content of a speech act must be a proposition expressed (i.e. incremental truth conditions), then many successful subsentential utterances will not be speech acts after all.

3.3. François Recanti's strong moderate relativism and subsentential speech acts

François Recanati is a proponent of a standpoint he calls strong moderate relativism. He does not discuss subsentential speech acts specifically,[13] but in my opinion his conception is perfectly suited for their analysis. According to the position in question, sentences have two kinds of content: explicit content and complete content. Explicit content (the *lekton*) may not possess absolute truth conditions and may be true only relative to some particular circumstance of evaluation. Complete content (the Austinian proposition) is explicit content plus the appropriate circumstance of evaluation. For example, the explicit content of the utterance "It is raining" is ⟨it is raining⟩; whereas the time and place of the rain are constituents of the situation determined by the context of the utterance. Recanati makes the following assumptions:

- duality: both a content and a circumstance of evaluation are necessary to determine logical value;
- distribution: the determinants of logical value (such as time) are given either as ingredients of the content or as aspects of the circumstance of evaluation (Recanati, 2008, p. 42).

Recanati also assumes the principle of economy according to which the elements necessary to determine logical value are either ingredients of the content or aspects of the circumstance of evaluation but never both.[14] In

[13]Recanati at one point cites the utterance "Very handsome!" and considers it to express a proposition dependent on the person. However, he does not explore this issue much further (see Recanati, 2007, p. 252).

[14]Compare Corazza's principle of cognitive economy cited above.

other words, the richer the content, the poorer the circumstance necessary for its evaluation can be and *vice versa*. The *lekton* of the sentence "It is raining" differs from the *lekton* of the sentence "It is raining here" because the content of the first sentence does not include the place of the rain. The *lekton* is the explicit content of a sentence in a context: the indexical sentence "It is Friday today" is thus going to have different explicit content in different contexts. Context determines the reference of the relevant indexicals and the situation in which the *lekton* is to be evaluated. The situation of evaluation need not be the same as the situation of context: a person saying "It is raining" in Warsaw may mean that it is raining in Cracow (if he has just spoken on the phone to someone in Cracow and is reporting the conversation, for example). The content of a sentence is the function from a situation to logical value. The truth of a sentence is thus relative: the same sentence can be true in one situation and false in another. Recanati defends a strong version of moderate relativism according to which even sentential utterances semantically expressing propositions (such as "It is raining here and now") have two levels of content: explicit content and complete content. The explicit content of a sentential utterance is a classic proposition (e.g. "It is raining in Warsaw on Jan 11, 2017 at 12:00"), and in order to assign determinate logical value to this proposition one only needs a possible world, not a constituent-rich situation. According to the weak version of moderate relativism, the *lekton* in this situation is simply its complete content. According to strong moderate relativism, the complete content of such an utterance will also include the appropriate situation: "what the utterance [of such a sentence – J.O.-S.] ⟨says⟩ is that *the situation in question supports the proposition in question*" (Recanati, 2007, p. 49).

In the case of sentences whose content is semantically complete, two kinds of evaluation are thus possible: one can evaluate the sentence itself (the proposition in regard to the actual world) or the utterance (the proposition in regard to the situation featured in the Austinian proposition) (see Recanati, 2007, p. 50). Let us imagine the following situation (Recanati, 2007, p. 50). I am looking at a group of people playing poker. It seems to me that among them is Claire. I see her cards and say: "Claire has a good hand now". It so happens that Claire is not present among the players I am looking at but she is at the same time playing poker somewhere else and indeed has a good hand there. Is the sentence uttered by me true? Our intuitions are contradictory here: on the one hand, the sentence is false about the situation I am looking at, but on the other, accidentally true about some other situation of which I am ignorant. Recanati claims that both intuitions can be grasped from

within his position for one can say that the *sentence* "Claire has a good hand now" is true (because it is made true by the situation of which the speaker is ignorant) but the *utterance* "Claire has a good hand now" is false because it is not true in regard to the situation featured in the Austinian proposition (Recanati, 2007, p. 50).

As I have already mentioned, I think that Recanati's conception is ideal for the analysis of subsentential speech acts. From our point of view, it is irrelevant which version of moderate relativism – weak or strong – is assumed since subsentential utterances must always be completed by a situation. According to the principle of distribution, some elements necessary for the determination of the proposition expressed and its logical value can be located in the circumstance of evaluation. Subsentential speech acts thus seem to represent a limiting case illustrating the functioning of the principle of economy: their content is very poor but the other necessary elements can all be found in the situation of the utterance. Let us assume that in the situation described above I only say "good hand". If it is clear to everyone whose cards I am looking at, the utterance will be understood as an assertion that the person I am looking at has (at the time of the utterance) a good hand. Here, only the denotation of the expression "good hand" belongs to the explicit content; all the other elements are unarticulated (the person in possession of a good hand, the relation of possession, time and place).

Moderate relativism, just like Corazza's position, refers to Perry's conception, in particular, to the idea of situatedness and unarticulated constituents. However, an analysis of subsentential speech acts in light of moderate relativism is more promising than their treatment as situated illocutions. This is because it allows us to speak of the explicit and the complete content of these acts rather than just their reflexive truth conditions.

Unfortunately, this application of Recanati's conception is not free from problems. On the one hand, it permits the ascription of content to all communicative acts, not just speech acts. On the other hand, it is not clear if it can distinguish asserted content from content that is implied or communicated otherwise. Both these difficulties blur the category of speech acts, especially the category of assertion. Let us imagine that instead of saying "good hand" I only say "good" and make a head gesture to indicate what I mean, or that I make the same gesture and give a thumbs-up. If my gestures are sufficiently clear and precise, I will have managed to successfully communicate that the person I am looking at has a good hand. But were any of these behaviours a speech act? Did any of these situations feature an assertion? The second situation certainly cannot be deemed a speech act

because I did not say anything. What about the first? It seems that as long as one can claim that I asserted a proposition, and that at least a part of this proposition stems from a linguistic source, this was indeed a speech act.

Merchant has made similar remarks regarding Stainton's position (Merchant, 2010, pp. 10–11). The advantage of this position, according to him, is that it separates the act of assertion from the particular kinds of linguistic utterances. This means that one can make assertions using utterances that semantically do not express propositions. Merchant entertains the idea that, in light of this, assertions need not be speech acts. If in an answer to the question "How many children do you have?" a person shows three fingers, they assert having three children, according to Merchant, despite having said nothing (in the sense of Grice) (Merchant, 2010, pp. 10–11). Regarding the last example, one could maintain that although this is not a speech act, it is a communicative act. The question is: can one take such acts to have asserted content as opposed to content communicated otherwise? Separating assertions from what is said certainly makes the boundaries of the former category extremely fuzzy.

3.3.1. What is said in subsentential assertions

Even if we assume that an assertion must be a speech act, we still need to introduce a principle allowing us to distinguish asserted content from otherwise communicated content. Stainton writes that an asserted proposition is a proposition generated through a minimal enrichment of the content of the expression uttered (Stainton, 2006, p. 161). Minimal enrichment is enrichment necessary for the content to have truth conditions.[15] However, it seems that Stainton's "Nova Scotia" example violates this rule. To recall, Stainton would like the proposition asserted in this example to be "the weather here is similar to..."; but it is difficult to take such an enrichment to be minimal.[16] According to Stainton, the criterion allowing one to distinguish asserted content from implied content is whether the speaker could be accused of lying as opposed to merely misleading the hearer. The

[15] Merchant notes that it is usually assumed that a proposition p is minimal relative to all the other propositions q in a contextually determined set P if for all q, q entails p. The problem is that there will be many propositions in a context that are not bound by the relation of entailment (Merchant, 2010, p. 15).

[16] One might try to save the situation by claiming that at issue here is a minimal function *manifest* in the situation (cf. Merchant, 2010, p. 26).

speaker can only be accused of lying if they asserted something (as opposed to suggesting it, implying it etc.) (Stainton, 2006, p. 58).

Stainton used the lie criterion to distinguish asserted content from communicated content, and Jennifer Saul (2012) uses asserted content (what is said) to define lying. When characterizing what is said, she appeals to the minimal enrichment criterion proposed by Stainton and puts forth the following definition:

> A putative contextual contribution to what is said is a part of what is said only if without this contextually supplied material [the sentence] S would not have a truth-evaluable semantic content in [the context] C (Saul, 2012, p. 57).

Saul claims that this contextual supplementation need not be grasped by the hearer, although she does not determine if it must be obvious in a given context or intended by the speaker. She also constructs examples aimed at illustrating this (see Saul, 2012, p. 60 ff.). Since her examples concern sentential utterances, I will not cite them here. Instead, I would like to propose similar examples (modifications of scenarios proposed by Corazza and Stainton) featuring subsentential utterances:

A. John, a well-known anti-Fregean has been told that Jane is looking for Dummett's *Frege: Philosophy of Language*. John knows that the book is lying in his desk drawer. Jane walks into John's office. John says: (1) Hidden on top of the shelf. Jane has bad hearing and did not hear a thing.

B. Sanjay and Silvia are loading furniture onto a van. Sanjay thinks that Silvia is looking for a misplaced table leg. He is mean and wants to mislead Silvia. Since he thinks that the leg is on the hutch, he says "on the stoop". It so happens that Silvia is looking for a desk drawer which is in fact on the stoop.

In scenario A, the intended content of the utterance was not grasped by the hearer. Corazza would consider John's utterance to be unsuccessful and would probably forego the analysis of its content. For Saul, it is irrelevant whether the hearer grasped the intended content or not. It seems that also Stainton would conclude that the Neo-Russellian proposition constituting the content of John's utterance comprises a copy of Dummett's book and the property of being on top of the shelf. For both Saul and Stainton,

the decisive factor would be the fact that there is no doubt in this case that John lied (even though the lie was unsuccessful). A similar solution could be proposed within moderate relativism. In scenario A, there is a situation sufficiently determinate to provide all the elements necessary for the utterance's complete content. As regards scenario B, Corazza would simply conclude that a successful communicative act occurred. There is no room in his conception for the charge of lying. Saul, on the other hand, has written about an analogous situation that it is unclear if this is a lie or not. The speaker (who is wrong as to the context in which he is situated) intended the table leg to enter the expressed proposition, but the context decided otherwise. Assuming that the situation is a part of the actual world, and that it is not determined by the speaker's intention, moderate relativism must conclude in this case that the content of Sanjay's utterance is different from the content he intended. These examples show that appealing to minimal contextual enrichment can be controversial, and that it might not suffice to univocally determine if a given linguistic item is a speech act. This need not be a problem since it might be argued that in cases like these it is just not clear if a speech act has been performed or not.

4. Conclusion

The main motivation for semantics-oriented standpoints denying the existence of subsentential speech acts is an unwillingness to expand the role of context. According to these standpoints, context does not play the controversial truth-conditional role consisting in providing constituents directly to the asserted content of the proposition expressed (as opposed to assigning semantic values to the constituents of the expression uttered) (see Stanley, 2000, p. 402). The truth-conditional function of context is limited to disambiguation and the resolution of indexicality.

Semantics-oriented standpoints can be criticized for their inability to offer a convincing analysis of all pertinent examples (e.g. "Rob's Mom"), for the fact that they make an unjustified appeal – unjustified given their own assumptions – to pragmatics (a pragmatic inference is often supposed to determine the logical value of the utterance), and the fact that they postulate, at least in some cases, counterintuitive contents for assertions.

The departure point of pragmatics-oriented standpoints is the assumption that the manner in which the content of a speech act is enriched is determined to a large extent by context, not by linguistic rules. The pro-

ponents of these standpoints assume that the role of context consists not only in disambiguation and the resolution of indexicality but also in the provision of constituents directly to the content of propositions uttered in context. The advantage of pragmatics-oriented standpoints is their appeal to familiar and otherwise required pragmatic mechanisms, and the fact that they do not interfere with syntax and semantics (they do not postulate introducing unarticulated structures or hidden variables) (compare Merchant, 2010, p. 10). Their weakness lies in the fact that they make it difficult to maintain the distinction between asserted content and content that is suggested or implied.

How to answer the titular question then? I have no doubt that we perform speech acts using at least apparently subsentential speech acts. It is true that their content or illocutionary force may not be fully determinate, but this is also the case for many sentential speech acts. The only thing that could be questioned is whether these utterances are indeed subsentential. I consider arguments offered by the proponents of semantic standpoints unconvincing. Many different supplementations are permissible in the majority of contexts, not just one select supplementation, which is why these utterances cannot be treated as ellipses, in my opinion. Appealing to widely known scripts does not help much either. The situation where a person is getting in a taxi can indeed be considered standard, that is, one for which some kind of script is in place. However, even here it is still unclear if the driver would ask "Where to?",[17] "Where are we going?", or "Where shall I take you?". Hence, Benigno's utterance "To Segovia. To the jail" cannot be treated as an answer to a particular question.

For these reasons, I consider pragmatics-oriented standpoints affirming the existence of subsentential speech acts to be more adequate. I have proposed to analyze such acts within the framework of Recanati's moderate relativism. Moderate relativism allows us to grasp the intuitions behind Stainton's standpoint in a more organized manner. In the proposed analysis, we can consider the contents of speech acts to be distributed between what was said on the one hand, and the situation in which the utterance was made, on the other. Since the explicit (semantic) content of subsentential speech acts is usually very limited, the situation plays an immense role in its completion. In Stainton's conception, emphasis is put on the controversial assumption that the content of a speech act is not formulated in natural

[17] The question "Where to?" is, of course, another example of an apparent subsentential speech act. The opponent of subsentential speech acts would have to appeal to another script in order to address it.

language; propositions stated by the speakers performing such acts are Neo-Russellian and their constituents include extralinguistic objects. All such propositions are thus *de re*. This assumption is not necessary in moderate relativism. What a subsentential speech act expresses is a propositional function relativized to the situation. Moderate relativism introduces the additional level of explicit content – explicit content need not be complete and can only have truth conditions in a sufficiently rich situation. That said, before moderate relativism can be considered a fully adequate analysis of subsentential speech acts, it is necessary to propose a satisfactory method of distinguishing asserted content from implied content – I think that the distinction into lying and misleading is on the right track, although it must be supplemented by an appropriate metaphysics of situation, among other components. The proponent of a pragmatics-oriented conception must also convincingly explain why in some cases subsentential utterances occur in cases other than the nominative.

Bibliography

Bach, K. (2008). Review: Robert J. Stainton: Words and Thoughts: Subsentences, Ellipsis, and the Philosophy of Language. *Mind*, *117*(467), 739–742.

Corazza, E. (2011). Unenriched Substentential Illocutions. *Philosophy and Phenomenological Research*, *83*, 560–580.

Dummett, M. (1973). *Frege: Philosophy of Language*. London: Duckworth.

Elugardo R., Stainton, R. (2004). Shorthand, Syntactic Ellipsis, and the Pragmatic Determinants of What is Said. *Mind & Language*, *19*(4), 442–471.

Hall, A. (2009). Subsentential Utterances, Ellipsis, and Pragmatic Enrichment. *Pragmatics & Cognition*, *17*(2), 222–250.

Lewis, D. (1969). *Convention: A Philosophical Study*. Cambridge Mass.: Harvard University Press.

Merchant, J. (2010). Three Kinds of Ellipsis: Syntactic, Semantic, Pragmatic? In F. Recanati, I. Stojanovic, N. Villanueva (Eds.), *Context-Dependence, Perspective, and Relativity* (pp. 141–192). Berlin – New York: De Gruyter Mouton. Page numbers based on the Internet version: `http://home.uchicago.edu/merchant/pubs/merchant.3kinds.pdf`.

Recanati, F. (2008). Moderate Relativism. In M. Kolbel, M. Garcia-Carpintero (Eds.), *Relative Truth*. Oxford: Oxford University Press.

Recanati, F. (2007). *Perspectival Thought, A Plea for (Moderate) Relativism*. Oxford: Oxford University Press.

Saul, J. (2012). *Lying, Misleading and What is Said*. Oxford: Oxford University Press.

Stanley, J. (2000). Context and Logical Form. *Linguistics and Philosophy*, *23*, 391–434.

Stainton, R. (2006). *Words and Thoughts: Subsentences, Ellipsis, and the Philosophy of Language*. Oxford: Oxford University Press.

Originally published as "Czy istnieją niezdaniowe akty mowy?" *Studia Semiotyczne* 30/2 (2016), 155–183. Translated by Joanna Guzowska.

STUDIA SEMIOTYCZNE—ENGLISH SUPPLEMENT, vol. XXIX (2017), 272–292
ISSN 2451-2745
DOI: 10.26333/stsen.xxix.12

Piotr Konderak[1]

The Conscious Semiotic Mind

Abstract The paper discusses possible roles of consciousness in a semiotic (meaning-making) activity of a cognitive agent. The discussion, we claim, is based on two related approaches to consciousness: on Chalmers' theory of phenomenal and psychological consciousness and on Damasio's neural theory, which draws a distinction between core and extended consciousness.

Two stages of cognitive-semiotic processing are discussed: the moment of perception of a sign as a meaningful entity and the metasemiotic processes understood as the human capacity to reflect on signs and their usage, analyse and control processes of recognition, interpretation of signs and to detect and correct errors in semiotic activity.

In the case of the first stage, it is argued that signs as meaningful entities have a distincly experiential character. The feeling of meaningfulness is a result of phenomenal consciousness, in particular a result of the so-called valuation features of phenomenal experience. I claim that this aspect of cognitive-semiotic activity is possible owing to a special neural mechanism called a semiotic marker.

It is argued that semiotic systems have to be able to use signs as signs, i.e. they should display some metacognitive capacities, in particular an ability to analyse semiosis at a metalevel. It is argued that such metasemiosis is dependent on psychological consciousness (in Chalmers' terms: awareness) and is realized at the neural level in the form of extended consciousness.

The paper is based on a particular understanding of cognitive semiotics as a discipline involving analyses of cognitive processes as semiotic processes, i.e. processes requiring usage of signs.

Keywords cognitive-semiotic system, phenomenal consciousness, awareness, core consciousness, extended consciousness, valuation features, somatic marker, metasemiosis

[1]Maria Curie-Sklodowska University in Lublin, Faculty of Philosophy and Sociology.
E-mail: kondorp@bacon.umcs.lublin.pl

1. A view on cognitive semiotics

As Sonesson (2012, p. 208) notes, "cognitive semiotics has been invented many times over during the past few decades". In the context of the statement it is not surprising that different researchers take different perspectives towards such a marriage of cognitive, linguistic and semiotic studies. Sonesson himself characterizes cognitive semiotics as an approach which aims to wed cognitive science and semiotics. In a similar vein, Zlatev (2012, p. 2) defines cognitive semiotics as a discipline whose "ultimate goal is to provide new insights into the nature and culture of human beings, as well as other meaning-making creatures [...]". Cognitive Semiotics (henceforth CS) can be defined as an interdisciplinary matrix of (sub- parts of) disciplines and methods, focused on the multifaceted phenomenon of meaning. This article presents some aspects of cognitive semiotics as seen from the perspective of standard cognitive science[2] – from the Chalmersian (1996) conception of mind. The approach presented in this paper highlights the role of signs, in particular linguistic signs, in the explanations of cognitive functioning of a cognitive agent. Consequently, my understanding of cognitive semiotics is that it encourages us to study cognitive systems, either natural ones like animals or human beings, or, possibly, artificial ones such as sign-using and meaning-making systems. In other words, I assume that at least some cognitive processes involve, in a nontrivial way[3], the use of signs. This means that there are sign-using (semiotic) processes which are in fact cognitive processes. For now I leave open the question as to whether all semiotic processes or only some of them are cognitive. Even if just some of them happen to be cognitive, we can still gain some knowledge about the nature of semiosis by studying selected cognitive activities.

My cognitive reading of semiosis (and semiotic reading of cognition) is motivated by the Peircean theory of signs, semiosis as well as his epistemology. The Peircean notion of a sign states that, "a sign, or representamen is something that stands to somebody for something in some respect or

[2]The qualification "standard" or "cognitivist" seems to be necessary here, as cognitive semiotics highlights the role of non-standard: enactive and embodied cognitive science. My approach is grounded in the "old-fashioned" or Cartesian cognitive science based on the notion of representation and cognitive modeling (either symbolic or connectionist) as a primary method. See e.g. Harnish (2002) for a systematic presentation of standard cognitive science and Rowlands (2010) for discussion on the relationship between standard cognitive science and "4e" approaches (enacted, embodied, embedded, extended).

[3]"Nontrivial" here means that neither one can describe, nor explain, such cognitive systems without mentioning the notions of 'sign' and 'meaning'.

capacity. [. . .] The sign stands for something, its object. It stands not in all respects, but in reference to a sort of idea which I have sometimes called the ground of the representamen" (CP 2.228). In consequence, the notions of a representamen, its interpretation, and a respect are of special importance. The meaning of a sign, in turn, arises in the (dynamic) process of interpretation of the sign. Instead of a dyadic relationship between a sign and its object, we have here a triadic relationship involving interpretation as the third element. I interpret the notion of a sign in terms of the Peircean definition. In addition, the Peircean theory of signs is understood here not in a narrow sense, as a description of actual and possible signs and sign systems, but is interpreted in a broader sense, as a theory addressing fundamental questions of cognition, its relation to reality as well as a logical analysis of knowledge[4].

Any human being, or – more generally – any meaning-making creature (to use Zlatev's formulation[5]) uses signs in his or her everyday cognitive activity. Sign-using agents recognize smoke as a sign of fire, photos as signs of real people and a red traffic light as a sign of an obligation to stop. In addition, natural-language using creatures recognize certain sounds as sounds of language and (at least sometimes) they are able to understand and interpret these sounds as signs, i.e sounds standing for something else in some respect. One can interpret in a similar way more complex (or: high-level) cognitive abilities. One of the basic cognitive activities discussed within cognitive science, namely the problem-solving activity, calls for the use of signs (in the broad sense: indexes, icons or symbols) as clues or premises and usually requires an interpretation of such signs' indications[6]. (The reader

[4]It is worth mentioning that Peirce also proposed a classification of consciousness in connection with his triadic definition of a sign. According to Peircean classification, (pure) feeling is consciousness of the Firstness (CP 7.551); experience (CP 8.266) or Altersense (CP 7.551) is consciousness of "otherness or secondness", and – finally – Medisense (CP 7.544) is awareness of the Thirdness. The latter may be divided further into abstraction, suggestion and association (CP 7.544–548). In addition, consciousness has a bodily (neural) and social dimension (CP 575). See also (Houser, 1983).

[5]It is necessary to notice that Zlatev uses the formulation – in line with Thompson's (2007) approach – only in reference to natural, autonomous systems (see also: Zlatev, 2009, Thompson & Stapleton, 2009). In other words, Zlatev excludes possible artificial systems from the scope of meaning-making creatures; in his view artificial systems cannot be truly "cognitive".

[6]As one of the reviewers of the paper noted, this statement concerning a problem-solving activity may be interpreted as a statement about the dependence of problem-solving on its capacity to use symbols. The remark is justified in light of enactive approaches to cognition (e.g. Noë, 2004), supported by some robotic experiments (Brooks, 1991, Beer,

can imagine here Sherlock Holmes solving one of his cases.) In a similar vein, one can emphasize the role of signs (and semiosis) in other activities like decision making, planning, etc. I would like to stress that analyses in terms of signs and meanings are not only a fancy way of describing these activities, but I am convinced that sign using and meaning making are unavoidable elements of these activities[7].

The main point of the paper is that all these (and similar) cognitive-semiotic activities require some form of consciousness. This is in line with the general phenomenological orientation of cognitive semiotics. Cognitive semiotics highlights the importance of the first-person perspective by stressing the role of consciousness understood as a subjective, qualitative experience. Phenomenology, in turn, is considered to be an approach which provides the right kind of method for studying the structure and content of consciousness (Zlatev, 2012, p. 2).

In contradistinction to the above assumption, however, I am convinced that phenomenological experience goes well beyond phenomena involved in semiotic activity which are merely labelled as "conscious"; indeed, I suggest that we should broaden our perspective by including analyses of instances of so-called psychological consciousness (or awareness, cf. Chalmers 2004, pp. 618–619).

In what follows I present the conceptual background of the paper – the notion of a cognitive-semiotic system and a notion of consciousness (section 2). Section 3 presents two distinctions concerning the notion of consciousness: a philosophical one, based on the Chalmersian approach, and a neuroscientific stance based on Damasio's theory. Sections 4 and 5 are devoted to the two stages in the cognitive processing of signs. Section 4 presents the initial stage, that of sign perception. I highlight here the role of phenomenal experience of meaningfulness and its role in the cognitive activity of a sign-using agent. In section 5 I propose a metasemiotic level of analysis of semiosis.

1995), where not only symbols, but representations in general are rejected. To clarify my viewpoint, I take problem-solving to be one of the higher-level, "representation-hungry" activities (cf. Clark & Toribio, 1994) and – as such – involving (at least partially) usage of signs. It does not imply the necessity to use symbols.

[7]In other words, I am taking here a realist stance towards semiotic phenomena rather than an instrumentalist one.

2. Two basic notions

Before taking a look at the role of consciousness in a semiotic and cognitive activity, let me elaborate the two key notions of the paper, namely the notion of a cognitive-semiotic system and the notion of consciousness.

2.1 A cognitive-semiotic system

The notion of a mind as a cognitive system is one of the basic notions used within standard, cognitivist (Thompson, 2007) cognitive science. Keeping in mind the multidimensional character of a mind (phenomenal, emotional, subjective, cognitive, even computational), I will discuss – in line with the cognitivist approach – only cognitive aspects of mind. The focus on cognitive aspects is motivated by the initial assumption of the paper, i.e. the claim about a relationship that holds between cognition and semiosis.

The term "cognitive system", as I understand and use it, describes a complex, structured entity which is a subject of processes such as perception, action, reasoning, planning, problem soving, and natural language understanding. It is understood as a dynamic structure which receives environmental and bodily information, processes it according to its internal organization, stores the information and finally acts on the basis of this information (cf. Nęcka *et al.*, 2006).

As mentioned above, I am interested in a subclass of cognitive systems, namely cognitive-semiotic systems. Such systems are understood here as systems which use signs in their cognitive activity, i.e. they are able to create, distinguish, interpret signs as well as use them in directing their behavior. However, to avoid the temptation of behavioristic (cf. Fetzer, 1997) interpretations of the statement[8], one must take into account one more condition: any cognitive system should use signs as signs, i.e. the system should treat signs as something that stands for something else in some respect or other. In other words, "the behavior of the system is causally affected by the presence of a sign because that sign stands for something else iconically, indexically, or symbolically, for that system" (Fetzer, 1997, p. 358). As a consequence, the system is, or at least should be, aware that its mental activity and physical behavior is influenced by semiotic processes.

[8]I am not going to justify such avoidance – behaviorist approaches to mind have been severly criticized by philosophers representing different stances and repetition of all the arguments seems to be pointless. Consult e.g. (Kim, 2011) for an overview.

2.2 Consciousness (and awareness)

As David Chalmers notes, "consciousness is an ambiguous term, referring to many different phenomena" (Chalmers, 2004, p. 617). Contemporary literature on consciousness abounds with differing approaches to the phenomenon and various attempts to define it (cf. Jackendoff, 2007, pp. 77–80 for an overview). The spectrum embraces, among others, eliminativist approaches, which treat consciousness as a useful fiction at best (Churchland, 1981), reductionist theories (Place, 1956, Smart, 1959), functionalist approaches (Armstrong, 1980, Putnam, 1975) as well as theories highlighting the subjective character of conscious experience (Nagel, 1974, Searle, 1992). One can hardly disagree with Damasio, who claims that "the conflation of so many meanings around the word consciousness renders it almost unusable without qualification, and this conflation is probably responsible for the supreme status to which consciousness has been elevated" (Damasio, 1999, p. 309). To avoid the danger of conflation of this kind, I would like to put my philosophical cards on the table: I understand the phenomenon of consciousness in the sense of Chalmers' (1996) naturalistic and nonreductive theory of consciousness. Chalmers distinguishes between phenomenal and psychological consciousness, stressing both an experiential character of consciousness and a role of consciousness in a mental activity (functional aspect). Phenomenal consciousness is – in the context of this paper – an answer to the question: How is it like to experience signs or meaningful entities? Awareness (or psychological consciousness), in turn, answers the question of what the role of conscious states (processes) in recognition, comprehension and usage of signs is?

Even if one rejects materialistic approaches, which reduce consciousness to a brain activity (*pace* Place), one nowadays can hardly deny that it is impossible to discuss consciousness independently from the achievements of neuroscience[9]. This is the reason why I wish to suggest a kind of interpretation of Chalmers' distinction in terms of a neuroscientific approach to consciousness – from the point of view of Damasio's distinction between core and extended consciousness (Damasio, 1999).

[9]The need of neuroscientific grounding is appreciated also within phenomenological tradition. Neurophenomenology (Varela, 1996) is seen as an important project integrating phenomenological research on consciousness and results of sciences. In the case of standard, functionalist cognitive science the connection between consciousness studies and neuroscience is evident.

3. What is consciousness?

To sum up the above terminological considerations, I treat consciousness as a heterogenous phenomenon, which involve two distinctions: a philosophical distinction between phenomenal and psychological notions of consciousness and a neuroscientific distinction between core and extended consciousness. The two approaches are presented below.

3.1 Chalmers' approach to consciousness

In his nonreductive theory of consciousness, Chalmers attempts to explain a wide spectrum of phenomena called in commonsense language "conscious phenomena". The phenomena include, *inter alia*, perceptual experiences (experience of redness, auditory experience of loud sound or tactile experience of a sheer surface), experience of pain, reportability of mental states ("I see red"), belief formation and revision ("I believe I should stop"), decision making ("I deliberately choose not to obey the rules and proceed despite the red light"), problem solving ("How to explain it to the policeman?"), planning, etc. All these phenomena may be treated as conscious ones. Analysing such and similar examples of mental activities, commonly acknowledged as "conscious", Chalmers claims that these phenomena should be grouped into two classes: phenomenal and psychological, reflecting in this way the two ways of thinking or talking about consciousness[10].

Certain cognitive subjects, particularly human beings[11], sense the world and have feelings or experiences connected with sensory data. They experience – subjectively and privately – their world and their bodies. In that sense, cognitive agents are sentient. On the other hand, in the context of standard cognitive science and studies on cognitive systems, cognitivists highlight the sensitivity of an agent to information and they stress the role of information in controling agents' actions. In this sense cognitive agents are conscious as to whether they are able to adjust their mental or physical activity to incoming stimuli, state of knowledge, data in memory, etc. In other words, cognitive agents are *sapient*.

[10] The distinction is somehow grounded in (and motivated by) Ned Block's (1995) distinction between access consciousness and phenomenal consciousness.

[11] This formulation raises the question about a class of sentient creatures. Zlatev (2009, p. 1981) notes that a subject should be a "minimal self" in the sense of Gallagher (2005) and enumerates – on the basis of first-, second- and third-person arguments – monkeys, dogs, cats, rats as possible sentient creatures. My argumentation concerns primarily human beings.

Both aspects of conscious phenomena: sentience and sapience have been stressed in the philosophy of mind: the first one has been elaborated and discussed in the phenomenological approaches as well as in the "subjectivist" theories (Nagel, 1974, Searle, 1992); the second one can be traced back to the behaviorist descriptions of cognitive systems[12] and is present in contemporary materialist and functionalist theories (e.g. Kim, 2000). Accordingly, one can associate, as Chalmers does, the above-mentioned two kinds of mental phenomena with the following types of consciousness.

Psychological consciousness (awareness) is a state in which a cognitive system has access to information which he or she uses in controling and directing their cognition. Most typical examples include reportability of mental states, belief formation and revision, discrimination and categorization as well as decision-making, problem solving, planning, etc. One is conscious psychologically when one is aware of the environment and its pariticular state ("There is red light") as well as of his/her own bodily state ("I am cold") or mental state ("I am too stressed"). In addition, the agent is able to report these states, draw conclusions and use the knowledge in directing his/her behavior. To apply the above characteristics to semiotic activity such as the detection or recognition of a red light (as distinguished from a green light) may result in awareness of an obligation to stop, to stop at an intersection (i.e. the sign here influences one's behavior) or break the law (with an awareness of the consequences of such behavior). All these mental activities: distinguishing, reacting, reasoning about consequences are examples of awareness. As Chalmers notices (1996, p. 28), in everyday settings we use the word "consciousness" in reference to such a situation[13].

Phenomenal consciousness is, in turn, a state in which a cognitive agent experiences subjectively the perceptual stimuli. In other words, there is something it is like to be a cognitive agent; in particular, there is something it is like to be a conscious creature. (Nagel, 1974, , p. 619) When an agent is suffering pain, if he or she is enjoying experienced sounds of someone's speech, if a cognitive system is experiencing redness (or roundness) of a signal on a traffic light or coldness of the day, all this is a manifestation of phenomenal consciousness[14]. The reader may have noticed some correlations

[12]I do not mean here eliminativist behaviorism.

[13]From the functional point of view, it is the only aspect of consciousness that is explainable. As Putnam notices (1981) even if there is something more, it cannot be explained in a functionalist framework.

[14]The reference to Nagel is somehow misleading: subjectivity – according to his approach – consists of two aspects: phenomenal content and particular individual point of view (perspective).

between both psychological and phenomenal examples. On the one hand, I am aware of a red light: I can report it, I can react in the presence of it; on the other hand, I experience subjectively redness. This is no coincidence: it may be the case that the two types of consciousness are closely related[15].

I would like to stress that the difference between phenomenal and psychological consciousness presented above is not only terminological ("two senses of the word") or conceptual ("two concepts of consciousness". The distinction may reflect the ontology of the world containing cognitive agents.

3.2 Damasio's approach

The second distinction has been suggested by Antonio Damasio, who has drawn a distinction between core and extended consciousness, based on the results reported in neuroscience. According to Damasio, core consciousness is a very basic process which enables a sense of self: a sense of the here and now. That is, an agent is aware of feelings occuring at the moment when his or her internal state changes. Core consciousness is a result of the interaction between a mind and an external entity (Damasio uses the term "object"). To quote Damasio: "the brain of the organism creates an image of its internal state, an image of the object, and an image of the internal state as it is modified by its interaction with the object. In addition, it creates a second-order image that includes all of these and may result in the feeling of the core self experiencing the resulting qualia" (Damasio, 1999; my emphasis).

To highlight the basic features of core consciousness: it is a simple, biological phenomenon and it is stable across an agent's lifetime. Damasio claims that we, human beings, share this type of consciousness with some other species.

Such a characterization suggests that core consiousness may be closely correlated with phenomenal consciousness (in Chalmers' sense). I would say that the philosophical notion of phenomenal consciousness is implemented at the neural level in the form of core consciousness[16]. I would like to stress that, according to Damasio, core consciousness does not depend on higher cognitive processes like planning, reasoning or language (Damasio, 1999, p. 16).

[15] Chalmers claims that it is a fact about our world (Chalmers, 1997, p. 18) that psychological processes of awareness are accompanied by experiences.

[16] However, I do not suggest that core consciousness is limited to phenomenal consciousness; the experiential aspect of consicousnes goes beyond core consciousness.

On the other hand, Damasio distinguishes extended consciousness, which goes beyond an agent's "here and now" and beyond his or her basic feelings. It enables "an elaborate sense of self" (Damasio, 1999, p. 16), i.e. an agent's awarenes of location in space and time (including memories and predictions of the future), an explicit distintion between "me and other" and between a subject and his or her environment. It enables deliberations on possible causes and results of actions as well as on failures and successes. Finally, it provides explicit metaknowledge in that it allows one to access consciousness. According to Damasio (1999, p. 16), extended consciousness is a complex biological phenomenon; it requires both long-term memory and working memory and it evolves during the lifetime of an agent. Because extended consciousness in its highest form is partially a result of language, it is supposedly present only in human beings. Characterized in such a way, extended consciousness may be treated as a neural realization of psychological consciousness. As stated by Damasio (1999, p. 201), "Extended consciousness is a bigger subject than core consciousness, and yet it is easier to address scientifically. We understand fairly well what it consists of cognitively and we also understand the corresponding behavioral features". The quotation matches Chalmers' characteristics of awareness.

4. Perception of a sign

With the above distinctions and clarifications made, I can now present putative roles of consciousness in the semiotic activity of a cognitive agent.

One of the basic methods used within cognitive science is cognitive modeling. This method focuses on computational – either symbolic, connectionist or hybrid – simulation of cognition. Cognitive models are based on the initial set of facts (initial knowledge) and certain control structures specifying how to cope with the data. Cognitive models are supposed not only to produce the same or similar behavior as human beings; they should also predict behavior as well as learn task-specific knowledge (cf. Taatgen & Anderson, 2008). As I have argued elsewhere (Konderak, 2015), it is possible and fruitful to model in this way a process of semiosis and, in particular, language comprehension, interpretation and production (cf. Konderak, 2007). To create a model of cognitive ability one usually analyses the processes modeled into a number of stages or steps. In the present chapter I will follow the procedure, indicating steps in cognitive processing important from the point of view of semiosis. In my opinion, there are at least three areas of

activity involving a semiotic (sign-using, meaning-making) mind in which the role of consciousness is indispensable, namely:

– an initial step: perception (proprioception) of an object as possibly meaningful (e.g. I see someone waving her or his hands; it may be the case that I do not understand what that waving is supposed to mean; I may even wrongly treat it as meaningful);

– establishment or recognition of the relationship between that perceived phenomenon (*representamen* in Peircean terms) and its Object (once again, understood in the framework of Peircean theory of signs) (e.g. I interpret such waving as a sign of a windy area);

– metasemiotic processes (explicit analysis of a sign as a sign), including recognition of a ground of the relationship (indexical, iconic or symbolic), discovery of an an error and ability to re-interpret a sign (e.g. I try to justify my interpretation on the basis of iconicity; I may also realize that I misinterpreted the gesture – the person observed just wanted to get rid of a fly).

All these three areas require some kind of consciousness[17]. In the sections to follow I analyse two stages of semiotic processing: the initial stage involves the perception of a sign as a meaningful entity and the metasemiotic processes, i.e. awareness of the semiotic activity of an agent.

4.1 Experiencing a sign as a meaningful entity

We are surrounded by signs. At first blush, the statement seems to be false: we are surrounded by objects with certain properties, processes or events. If Peirce is right, anything from our environment may be considered as a sign (once again: a red light at an intersection as a sign to stop, someone's statement: "It's red", a person stopping before a pedestrian crossing as signs of the same; pain in my stomach as a sign of e.g. stomach ulcer; doctor's words being a sign of the same disease, etc.). Everything may be a sign, but, certainly, it is not the case that everything is a sign. What is important here is the subject's perspective: stomach pain can be taken as a sign (an index)

[17] As Zlatev noticed, the argument may be related to the one from phenonology (Zlatev, 2010): consciousness is needed to have a world (of reference); the differentiation between the expression and the referent is based on consciousness, as well as the asymmetrical relation between the two.

of – let us say – an ulcer, but one need not take this viewpoint on the pain in question. We, as cognitive agents, have the ability to pick up some elements of our environment and treat them as signs (meaning-inducing entities). In other words, it is a fact about our cognitive activity, that we perceive only selected subsets of surrounding objects, situations and processes, first as meaningful entities, then as signs[18].

The process is often very fast and does not require much reflection. It is often the case that we perceive signs without explicit consideration or awareness of them as entities standing for something else. To illustrate, a perception of a preceeding car slowing down suddenly may evoke a relevant behavior of a person driving behind (applying the brakes) or induce some emotions (fear) or beliefs ("Something happened"). That is, slowing down may function as a sign of a danger or as an unexpected event for another driver. In a similar way, the utterance "Stop!" may catch the driver's attention and cause him or her to stop the car or at least make them pay more attention to the surroundings. In both cases the reaction is immediate and neither reflection nor consideration is involved; if an analysis takes place, it follows the initial phase of the perception process.

I suggest that the initial moment of the semiotic activity of a cognitive agent is an experience of meaningfulness, a feeling that some perceived entity or event is meaningful. In other words, to paraphrase Jackendoff (2007, p. 81), a meaningful entity has a distinguished experiential character. Such a feeling may cause an immediate decision as to whether the entity mentioned is meaningful or not (cf. the discussion of a somatic marker below) and then may trigger further analysis of the experienced phenomenon.

4.2 An experiential basis: qualia

The above stipulation about the experience of meaningfulness requires special features of our perceptual experience – features allowing for distinguishing potentially meaningful entities from meaningless ones. "Traditional" approaches to qualia (e.g. Lewis, 1929, Jackson, 1982) are not sufficient to explain a special mechanism detecting "meaningfulness". It means that – contrary to some of the researchers treating qualia as basic, unanalysable elements of our, human conscious experience – one should take a closer look at subjective experience and its features. It is possible (and probably necessary) to study the structure of qualia.

[18] A clarification is necessary here: experience of a phenomenon as meaningful does not make it a sign. It is the initial step in the process of semiosis. In other words, it might be necessary for a sign to be experienced as meaningful, but not sufficient.

Jackendoff (2007, chapter 3) analyses phenomenal experience (in reference to perception of natural-language utterances). According to his approach, consciousness has at least two dimensions: phenomenal content (or in Jackendoffian terms, content features which are traditionally discussed in philosophy of mind) and valuation features (Jackendoff, 2007, p. 87). Jackendoff enumerates a number of candidate pairs of the latter, e.g.: external (or not), self-initiated (or not), familiar (or not), affective (or not) meaningful (or meaningless), among others. Two valuations are of particular importance in the context of sign perception: the feeling that the perceived object, state or situation is meaningful and the feeling of familiarity. The idea of the two distinctions is that familiar objects (in particular utterances) – in contradistinction to unfamiliar ones – do indeed have a different experiential character (Jackendoff, 2007, p. 81). Similarly, we experience differently meaningful entities (in particular utterances) and meaningless ones. According to Jackendoff, if it is true that we experience language in the form of phonological images or, in the case of signed languages, in the form of visual or proprioceptive experiences (Jackendoff, 2007, p. 83), then these images (or other experiences) have an additional "felt" character: we experience them as meaningful.

I would like to push the hypothesis one step further: just as in the case of language, we tend to experience signs (including non-linguistic ones) as meaningful entities as well. Indeed, sometimes one "feels" that their experience (of object, sound or reminiscence) is meaningful despite one's inability to grasp the meaning itself. I suppose that such a feeling of meaningfulness is based on one's past experiences. In such cases the feeling of familiarity has some priority: if we are familiar with some stimuli (e.g. a special pattern of sounds, typical for a given language), my previous experiences (e.g. phonological images) would be responsible for the feeling of meaningfulness and later would trigger mechanisms of interpretation.

The following two properties of valuation features seem to be relevant in the context of semiotic processing. First, these features may be subject to error: it may be the case that one has a feeling of familiarity [when] perceiving completely new objects or situations (*déjà vu*); it may also be the case that one has a feeling of meaningfulness of an utterance while the utterance is meaningless. Second, valution features are, in a sense, independent of perceptual modality. As Jackendoff says (2007, p. 88) they cut across the "vertical" domains of language, vision and so on.

To sum up, when one is experiencing a sign, phenomenal consciousness appears to play the first and main role, i.e. we start with qualia and their

features, and to be precise – with valuation features. Such aspects of phenomenal experience bias (or guide) an agent's behavior or direct further cognitive processing, the process of interpretation included. The phenomenal valuation (e.g. the feeling of being external and meaningful) should form the basis for the initial distinction to be drawn between signs and non-signs. The above suggestion can be supported by the Damasian idea of a somatic marker, elaborated in the next section.

4.3 A hypothetical mechanism: (semiotic) markers

Damasio (1994) suggests the existence of a neural mechanism known as "somatic marker" which, according to him, provides an explanantion for the fast (in fact immediate) decision-making process, a mechanism allowing for quick choices between available alternatives. The mechanism is based on core consciousness and in particular – on emotions. A somatic marker works as an automatic alarm, it warns against possible negative consequences of the choice made. The warning is based on our previous experiences, encoding associations between objects or events and some states of a body. One can think also about somatic markers as directing mechanisms where some alternatives are immediately rejected, leaving a much smaller number of alternatives to be considered.

According to Damasio, "somatic markers [...] assist the deliberation by highlighting some options (either dangerous or favorable), and eliminating them rapidly from subsequent consideration. You may think of it as a system for automated qualification of predictions, which acts, whether you want it or not, to evaluate the extremely diverse scenarios of the anticipated future before you. Think of it as a biasing device" (Damasio, 1994, p. 174).

It is worth noting that a somatic marker is a mechanism shaped by experience; sometimes it is created during processes of socialization.

The idea of a somatic marker was a motivation to stipulate an analogical mechanism responsible for a detection of (at least some) meaningful signals. An initial observation is that human beings, in their everyday functioning, decide quite quickly whether certain objects or events are meaningful entities. Without a mechanism allowing for a quick choice we would be "drowned" in the multiplicity of potential signs (not to mention a number of possible interpretations of each of them). I am convinced that, in some cases an efficient semiotic activity requires some "fast-track" decision mechanism. Such a mechanism, called the "semiotic marker" would be enabled by consciousness. According to this suggestion it is phenomenal consciousness (and its valuation features in particular) that could be responsible for detecting

meaningful entities in an environment. The above proposition should be treated, at least at the moment, as a kind of speculation which calls for detailed empirical examination[19]. To reiterate, this is just an initial step in semiotic activity: experience of meaningfulness is not a sufficient reason for a sign to be recognized.

5. Using a sign as a sign

Although the role of experience and phenomenal consciousness is unquestionable, I am convinced that we cannot explain semiotic activity of a cognitive agent independently of psychological consciousness (awareness). The essence of the claim is particularly clear in the case of metasemiosis (and metacognition in general).

5.1 Metacognition and metasemiosis

Moses and Baird (1999) define metacognition as "any cognitive process that controls or monitors any aspect of cognition". Metaknowledge, in turn, can be defined as "knowledge about knowledge", which embraces, among other things, beliefs about beliefs (metabeliefs). I treat metasemiosis as a metacognitive process that utilizes metaknowledge. Consequently, metasemiosis is understood as a human capacity to reflect on signs and their usage, to analyse and control processes of recognition and interpretation of signs, to detect and correct errors in semiotic activity, etc. There are at least three reasons to discuss metasemiosis in the context of cognitive-semiotic systems.

First, as Petrilli (2014, p. xviii) points out, "human being is [...] an animal capable not only of semiosis, but also of semiotics, that is, of using signs to reflect on signs". In other words, a cognitive semiotic system is able not only to use signs but also to discuss them: define, classify them, reflect on their properties. In general, a semiotic system is able to theorize about signs[20].

Second, metasemiotic activity as characterized above is a semiotic activity *per se*. A theory of signs can be analyzed as an example of (meta-)sign-

[19]The results of the so-called P300 experiment (Chapman & Bragdon 1964) may be interesting in this context. The researchers presented subjects two kinds of (visual) stimuli: numbers and flashes of light. Chapman and Bragdon concluded that ERP responses to visual stimuli differed depending on whether the stimuli had meaning or not for subjects.

[20]It could be a kind of implicit, commonsense (or folk) semiotics.

usage. For instance, a review of this chapter may be analysed as the right (or wrong) interpretation of natural language signs.

The third and main motivation for considering metaknowledge in context is the danger of a behavioristic interpretation of semiotic activity. According to such an interpretation, a cognitive agent is a semiotic system if it reacts in some way to special kinds of stimuli (called signs). I wish to claim that mere reaction is not enough as the systems displaying such ability are just "as-if" semiotic systems, i.e. systems that behave as if they use signs. The danger of such an interpretation emerges from discussions on the possibility of artificial semiotic systems: "For a causal system to be a semiotic system, of course, it has to be a system for which something can stand for something (else) in some respect or other, where such a something (sign) can affect the (actual or potential) behavior of that system" (Fetzer, 1988, p. 139)[21].

As stated above in line with Peircean approach, a cognitive semiotic system must be aware that it uses signs as entities standing for something else (in some respect), i.e. the system needs to have some metaknowledge embracing the usage of signs as well as be able to specify some metaprocesses that control the interpretation and usage of signs. The role of such a metalevel is implied in Fetzer's discussion on the possibility of artificial semiotic systems. Fetzer suggests a test checking whether a cognitive system is a semiotic system as well. The criterion is the capacity to make a mistake. As he indicates, to be a real sign-user, a cognitive agent "has to take something to stand for something other than that for which it stands" (Fetzer, 1988, p. 141, my emphasis)[22]. I would modify the statement: the agent has to be able to make a mistake and to realize the mistake as well. In consequence, a real cognitive and semiotic system should be able to realize (among others) that:

– there are possible alternative interpretations of a sign;

– he or she made a mistake in interpreting a sign;

– the sign used is an inappropriate one (taking into account norms of a community);

[21] As Johan Blomberg noticed (personal communication), some semioticians would reject here the applicability of the notion of sign in the context – they would treat such behavior-evoking phenomena as mere signals.

[22] Taking something to stand for something other than what it stands for (for instance, taking the green light to stand for the obligation to stop at the intersection) implies – according to Fetzer – the capacity to take something to stand for something else (in some respect) in general.

– a sign is unexpected in particular contexts etc.

5.2 Metasemiosis and awareness

Following Chalmers' distinction, metasemiotic processes are examples of conscious processes in psychological sense. The processes seem to be specific in that they require explicit deliberation on semiotic acitivties and they distinguish a special class of semiotic systems. As mentioned above (section 3.1), of particular interest are: the ability to access and report own interpretations of signs; ability to discriminate and categorize kinds of signs, ability to revise interpretations and modify behavior, ability to make deliberate choices, plan usage of signs, etc. To justify the claim I would like to consider the typical mental capacities treated usually as instances of awareness (conscious in psychological sense) in connection with metasemiotic activities. One is aware when one is able to:

– access own mental states – one is not only stopping at a red light, but one knows the reasons for stopping;

– report mental states (the ability assumes introspection and a language faculty) – one justifies crossing the junction despite the red light – "I noticed the red light, but I am in a hurry so I decided. . . ";

– discriminate kinds of signs – when one wonders whether a road sign "dangerous bend to left" is an icon or a symbol, and why;

– integrate information (and is able to solve inconsistencies) – when one sees a red light at an intersection and simultaneously one observes a police officer signaling "go".

All the above examples of metasemiotic activity are clearly dependent on psychological consciousness.

As argued above, Chalmers' awareness seems to be neurally realized in the form of the Damasian extended consciousness. Damasio (1999, p. 195) states: "Extended consciousness goes beyond the here and now of core consciousness, both backward and forward". What happens when a cognitive-semiotic mind perceives a sign, say, a red light? Rather than just access the fact that one experiences redness of a red light at a intersection, one can also survey the facts concerning the situation: where it is located (in front of you), what caused it (an electronic system for managing traffic), when has one experienced it before (ten minutes ago), who has also experienced

it recently (one's mother), who discussed it (one's boss), the fact that one should perceive a green light soon. As can be seen from these examples, the functioning of extended consciousness requires several abilities, including the ability to learn and memorize numerous and various past experiences, the ability to reactivate those memories in connection with "a sense of self knowing", the ability to direct attention to the content of mental states and the ability to predict and plan behaviors. Consequently, it seems that, seen from the context of semiotic activity, the characteristics of extended consciousness and the description of psychological consciousness converge. Both are connected with so-called higher-level cognitive processes; both assume a kind of self-awareness and existence of a self-model.

Simultaneously the two approaches are formulated at different levels[23]. Psychological consciousness is characterized by a role of conscious mental states in the functioning of a cognitive system. To explain a cognitive function like the interpretation of an ambiguous sign, we need only to specify a mechanism that performs the function (Chalmers, 2004, p. 620). It seems that extended consciousness is perfect for this task.

6. Conclusions: towards phenomenal consciousness

Cognitive semiotics is by stipulation closely connected with research on consciousness and priority is given to first-person methods (Zlatev, 2012). In this paper, I have argued that the first step in the course of semiotic activity has such character: it is phenomenal consciousness in general and valuation features in particular that allow us to pre-select meaningful entities. As a result, a cognitive system treats certain entities as meaningful without grasping the meanings of such entities. The putative mechanism explaining the phenomenon (a semiotic marker) is based on past experiences of a system. As a consequence, the mechanism works only in reference to a subset of all possible signs. The "feeling of meaningfulness" becomes now an impulse to recall information e.g. from long-term memory and to further processing. In situations involving phenomenal consciousness, subjective experience could be (and usually is) followed by psychological consciousness (awareness): the feeling that something is meaningful may be followed by an analysis of ground of meaningfulness (similarity, convention) or attempts to elicite a meaning (cf. also Chalmers, 1996, pp. 218–222). The higher-level metasemiotic processes require, it seems to me, psychological consciousness.

[23]In the sense of the Oppenheim-Putnam hierarchy (Oppenheim & Putnam 1958, p. 9).

To appreciate fully the role of consciousness in cognitive and semiotic activities one has to include in the theory one more aspect of a sign relation. Consciousness and awareness have their role in apprehending the relation between a sign and its object (CP 2.247–249). I thus stipulate that different kinds of signs (indexes, icons, symbols) require different types of consciousness. Icons appear to be more closely connected with phenomenal consciousness, whereas the use of symbols seems to primarily depend on psychological consciousness. This initial suggestion, however, can only be confirmed (or not) by further analysis.

Bibliography

Armstrong, D. (1980). *The Nature of Mind and Other Essays*. Ithaca: Cornell University Press.

Beer, R.D. (1995). A Dynamical Systems Perspective on Agent-Environment Interaction. *Artificial Intelligence*, *72*, 173–215.

Block, N. (1995). On a Confusion about the Function of Consciousness. *Behavioral and Brain Sciences*, *18*, 227–247.

Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence*, *47*, 139–159.

Chalmers, D. (1996). *The Conscious Mind.*New York: Oxford University Press.

Chalmers, D. (2004). Facing Up to the Problem of Consciousness. In J. Heil (ed.), *Philosophy of Mind. Guide and Anthology* (pp. 617–640), New York: Oxford University Press.

Churchland, P. (1981). Eliminative Materialism and the Propositional Attitudes. *The Journal of Philosophy*, *78*, 67–90.

Chapman, R.M., Bragdon, H.R. (1964). Evoked Responses to Numerical and Non-Numerical Visual Stimuli While Problem Solving. *Nature*, *203*, 1155–1157.

Damasio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain.* London: Papermac.

Damasio, A. (1999). *The Feeling of What Happens*. Boston, Mass.: Houghton Mifflin Harcourt.

Fetzer, J. (1988). Signs and Minds: An Introduction to the Theory of Semiotic Systems. In J. Fetzer (ed.), *Aspects of Artificial Intelligence* (pp. 133–161). Dordrecht: Kluwer Academic Publishers.

Fetzer, J. (2016). Peirce and the Philosophy of Artificial Intelligence. In M. Bergman, João Queiroz (ed.), *The Commens Encyclopedia – The Digital Encyclopedia of Peirce Studies New Edition*. URL: `http://www.commens.org/encyclopedia/article/fetzer-james-peirce-and-philosophy-artificial-intelligence`, [accessed: 10.01.2016].

Harnish, R.M. (2002). *Minds, Brains, Computers. A Historical Introduction to the Foundations of Cognitive Science*. Malden: Wiley-Blackwell.

Jackendoff, R. (2007). *Language, Consciousness, Culture. Essays on Mental Structure*. Cambridge, Mass.: MIT Press.

Houser, N. (1983). Peirce's General Taxonomy of Consciousness. *Transactions of the Charles S. Peirce Society*, vol. 19, no. 4, 331–359.

Kim, J. (2000). *Mind in a Physical World*. Cambridge, Mass.: The MIT Press.

Kim, J. (2011). *Philosophy of Mind*. 3rd edition, Boulder: Westview Press.

Konderak, P. (2007). *Model kognitywny zdolności językowych*. PhD Thesis.

Konderak, P. (2015). On a Cognitive Model of Semiosis. *Studies in Logic, Grammar and Rhetoric*, *40*(53), 129–144.

Konderak, P. (2017). Between Language and Consciousness: Linguistic Qualia, Awareness, and Cognitive Models. *Studies in Logic, Grammar and Rhetoric*, *48*(1), 285–302.

Konderak, P. (in press). On Evolution of Thinking about Semiosis: Semiotics Meets Cognitive Science. *Avant. Trends in Interdisciplinary Studies*.

Nagel, T. (1974). What Is It Like to Be a Bat?. *The Philosophical Review*, *83*(4), 435–450.

Nęcka, E., Orzechowski J., Szymura B. (2006). *Psychologia poznawcza*. Warszawa: Wydawnictwo Naukowe PWN.

Noë, A. (2004). *Action in Perception*. Cambridge, Mass.: MIT Press.

Oppenheim, P., Putnam, H. (1958). Unity of Science as a Working Hypothesis. In H. Feigl *et al.* (eds.), *Concepts, Theories, and the Mind-Body Problem*, Minnesota Studies in the Philosophy of Science, vol. II (pp. 3–36), Minneapolis: University of Minnesota Press.

Peirce, C.S. (1931–1958). *Collected papers of Charles Sanders Peirce*, vols. 1–6, C. Hartshorne, P. Weiss (eds.), vols. 7–8, A.W. Burks (ed). Cambridge, Mass.: Harvard University Press.

Petrilli, S. (2014). *Sign Studies and Semioethics: Communication, Translation and Values*. Boston: Walter de Gruyter.

Place, U.T. (1956). Is Consciousness a Brain Process?. *British Journal of Psychology*, *47*, 44–50.

Putnam, H. (1975). Philosophy and Our Mental Life. In H. Putnam, *Mind, Language and Reality: Philosophical Papers*, vol. 2, Cambridge: Cambridge University Press.

Putnam, H. (1981). Mind and Body. In H. Putnam *Reason, Truth and History* (pp. 75–102), Cambridge: Cambridge University Press.

Rowlands, M. (2010). *A New Science of the Mind. From Extended Mind to Embodied Phenomenology*. Cambridge, Mass.: MIT Press.

Schraw, G. (1998). Promoting General Metacognitive Awareness. *Instructional Science*, *26*, 113–125.

Smart, J.J.C. (1959). Sensations and Brain Processes. *The Philosophical Review*, *68*, 141–156.

Sonesson, G. (2012). The Foundation of Cognitive Semiotics in the Phenomenology of Signs and Meanings. *Intellectica*, *58*, 207–239.

Taatgen N.A., Anderson, J.R. (2008). Constraints in Cognitive Architectures. In R. Sun (ed.), *The Cambridge Handbook of Computational Psychology* (pp. 170–185), Cambridge, UK: Cambridge University Press.

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology and the Sciences of Mind*. London: Belkarp Press.

Thompson, E., Stapleton, M. (2009). Making Sense of Sense-Making: Reflections on Enactive and Extended Mind Theories. *Topoi*, *28*(1), 23–30.

Varela, F.J (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies*, *3*(4), 330–350.

Zlatev, J. (2009). The Semiotic Hierarchy: Life, Consciousness, Signs and Language. *Cognitive Semiotics*, *4*, 169–200.

Zlatev, J. (2012). Cognitive Semiotics: An Emerging Field for the Transdisciplinary Study of Meaning. *Public Journal of Semiotics*, *IV*, 2–24.

Originally published in *Studia Semiotyczne* 31/1 (2017), 67–89.