

# Balkan Journal of Philosophy

---

ISSN 1313 – 888X

## EDITORIAL BOARD

### Editor in Chief

Vesselin **Petrov** (Institute for Philosophical Research – Bulgarian Academy of Sciences)

### Editorial board

Marin **Aiftincă** (Institute of Philosophy and Psychology, Romanian Academy of Sciences, Bucharest, Romania)

Christina **Ambareva** (Institute for Philosophical Research – Bulgarian Academy of Sciences)

Marina **Bakalova** (Institute for Philosophical Research – Bulgarian Academy of Sciences)

Pavo **Barišić** (University of Split, Croatia)

Betül **Çotuksöken** (T. C. Maltepe University, Istanbul, Turkey)

William **Desmond** (National University of Ireland, Maynooth, Ireland)

Ivajlo **Dimitrov** (Institute for Philosophical Research – Bulgarian Academy of Sciences)

Ina **Dimitrova** (Institute for Philosophical Research – Bulgarian Academy of Sciences)

Boris **Grozdanoff** (Institute for Philosophical Research – Bulgarian Academy of Sciences)

Lilia **Gurova** (New Bulgarian University, Sofia, Bulgaria)

Rosen **Lutskanov** (Institute for Philosophical Research – Bulgarian Academy of Sciences)

Plamen **Makariev** (SU “St. Kl. Ohridski”, Sofia, Bulgaria)

Nenad **Mišćević** (University of Maribor, Slovenia)

John **Norton** (University of Pittsburgh, USA)

Roberto **Poli** (University of Trento, Italy)

Stathis **Psillos** (University of Athens, Greece)

Athanassios **Raftopoulos** (University of Cyprus, Cyprus)

Isidora **Stojanovic** (Institute Jean-Nicod, Paris, France)

Christo **Todorov** (New Bulgarian University, Sofia, Bulgaria)

Bogdana **Todorova** (Institute for Philosophical Research – Bulgarian Academy of Sciences)

Michel **Weber** (Centre for Philosophical Practice, Brussels, Belgium)

English Reader: William Desmond

Available online at: <http://www.philosophybulgaria.org/en/Publikacii/BJP/index.php>

<http://www.cceol.com>

# BALKAN JOURNAL OF PHILOSOPHY

---

*Vol. 1, Issue 2, 2009*

## CONTENTS

### Articles

“Pragmatic Turn” in Contemporary Thinking. the Pragmatic Dimension and Shaping of the Pragmatic / 61

*Alexandru Boboc*

The Transcendental Dimension of Heidegger’s Analytic of Predication / 69

*Dimitri Ginev*

Towards A Phenomenology Of Gratitude—A Response To Jean-Luc Marion / 77

*Peter R. Costello*

*Der Untergang Des Abendlandes* in The Bulgarian Cultural Area / 83

*Nina Dimitrova*

*Non Sequitur* – Some Reflections on Informal Logic / 91

*Danilo Šuster*

Can We Save *A Priori* Knowledge? / 103

*Nenad Mišćević*

Meaning And Truth / 117

*Joško Žanić*

On What Value, My Lord?

How Values Intervene in Hard Legal Cases / 125

*Mladen Domazet*

Should we Consult Kant when Assessing Agent’s Moral Responsibility for Harm? / 131

*Friderik Klampfer*

## **Reviews**

Marin Aiftincă. *The Philosophy of Culture*, Romanian Academy Publishing House, Bucharest, 2008, 164 pages. [Reviewed by Alexandru Boboc, Bucharest University] / 157

Psillos, Stathis. *Philosophy of Science A-Z*. Edinburgh: University of Edinburgh Press, 2007. [Reviewed by Lilia Gurova, New Bulgarian University, Sofia, Bulgaria] / 159

## **Information for contributors**

## “PRAGMATIC TURN” IN CONTEMPORARY THINKING. PRAGMATIC DIMENSION AND SHAPING OF THE PRAGMATIC

*Alexandru Boboc*

Bucharest University – Bucharest, Romania  
steluta@acad.ro

### Abstract:

The study of a pragmatic dimension of semiotics appears only late in the history of this discipline. “Pragmatics” seems the last one called into the dispute of signs. The following study focusses on semiotics and the theory of action, stressing the distinction between “pragmatism” and “practical” matters within this philosophical discipline. The theory of “speech acts” is investigated and related to the new perspectives opened in the semiotics in order to highlight further avenues for research.

**Key words:** sign – semiotic – syntax – semantics – pragmatics; pragmatic dimension – “pragmatic reason” – “pragmatic turn”; pragmatic thinking – pragmatics – pragmatism; language – meaning – reference; “significance – use” – “speech acts”.

1. A study of the pragmatic dimension of the semiotics appears quite late in the history of this discipline. “Pragmatics” seems to be the last arrived at the debate of the signs. It has been noticed that “the division of semiotics into three parts” originated in Ch. S. Peirce’s researches and was updated by Ch. Morris; but not everybody agrees with this fact; H. Hermes and Schroter thought of a semiotic dedicated exclusively to mathematical theories that didn’t imply a pragmatic dimension<sup>1</sup>.

In this respect Carnap’s opinion is significant: “nobody doubts that a pragmatic study of natural languages is of great importance for the understanding of human behaviour, character and culture. I think the construction and the semantic study of the linguistic system are more important nowadays for the particular purpose of the development of logic. For logicians a pragmatic study can be useful<sup>2</sup>.”

Once established, however, the pragmatic tends to restructure the semiotics depending on the new understanding of the communicational process (related to action) and of the language itself as action. It is spoken today even of a “pragmatic reason”: “The idea of a progress of reason isn’t completely repudiated yet. But the Age of Enlightenment, its optimism and its absolute reason have passed. The demand and the project of a promotion strategy of a non-absolute impure rationality, as a “practice of becoming”, as “a practical reason practically extended” – a social demand – “could be considered as one of the great challenges for our present time<sup>3</sup>.”

From the view of the action theory, the same idea has been underlined: “Expression (*Sprechen*) and action are activities that show what the specific nature of the human race is. By expression and through action, human beings actively differentiate one another, instead of being simple different; these are *Modi* through which the human being is revealed”; “only through action and communication do the human beings reveal who they are, show their personality<sup>4</sup>”.

On these bases the word “pragmatism” has been also taken as a name for a philosophical attitude related to modern positivism, that emphasises the actionable side in a utilitarian direction, which is, of course, complementary to relativism. In fact, “pragmatic” means “according to action”, in the service of practice, oriented towards the connections between personal abilities and the consequences of action.

This fact must not lead to a confusion between “pragmatic” and “practice”: *pragma* (Greek “action”) and *prattein* (Greek *praktikós*, “related to action”) are differentiated through the fact that

<sup>1</sup> P. Botezatu, *Semiotică și negație, Orientare critică în logica modernă*, Iasi, Junimea, p. 96.

<sup>2</sup> R. Carnap, „Sinn und Synonymität in natürlichen Sprachen”, in vol. *Zur Philosophie der idealen Sprache*, hrsg. von Joh. Sinnreich, München, Dtv, 1972, p. 146.

<sup>3</sup> H. Lenk, *Pragmatische Vernunft*, Stuttgart, Ph. Reclam jun. 1979, p. 28.

<sup>4</sup> H. Arendt, *Das Handeln*, in vol. *Handlungstheorien Interdisziplinär*, II 1, hrsg. von H. Lenk, München, W. Fink. 1978, p. 14, 18.

“practice” is related to *praxis* (“practice” is different from “theory”), which is related rather to a behavioural, moral action. From the historical perspective of Kantian “practical reason”, one gradually delimits a type of rationality, defined not only in an ethical-behavioural manner, but also in an actionable manner, a fact that led to praxeology (the theory of efficient action).

In an historical approach, “praxis” reproduces the Greek term „*pragmatikós*”, which can be translated as “skilled (experienced) in affairs (practical things) and would mean: 1. ready for action, serving the practice, engaged in practice; 2. serving the public welfare. Pragmatic historiography (developed for the first time by Polybios) had to do with a historical description in the context of which the data are studied according to their internal causal connections, particularly from the view of political actions. In Kant (“Anthropology from a Pragmatic Point of View”, 1798) the term “pragmatic” is employed by opposition with physiological, scientific naturalism, and is somehow identical with an ethic that serves self-knowledge and moral action”<sup>5</sup>.

2. Along with the latest studies focussing on semiotic studies in the pragmatic field, which is, in fact the identification of what was once called “linguistic turn” with the “pragmatic turn” appears a considerable effort to redefine the semiotic and pragmatic terms and to conceive a new historical approach, designed to justify these preoccupations and the status of the disciplines that aim to include and reorganize them from a systematic and conceptual point of view<sup>6</sup>. As we stated before, “pragmatic

thinking” is in the centre of these concerns. It developed “in various philosophical fields, especially in the theory of science”, as “American pragmatism, conventionalism and European operationalism”, but also “the logical empiricism”, the cybernetic movement, and the new “theory of science”<sup>7</sup>. The goal of both historic and systematic dimensions was to reach “a better comprehension of the contemporary practice”, a rethinking of our contemporary position in the world and our future action”<sup>8</sup>.

Peirce's ideas from 1868 to 1890, which became wellknown only in the third decade of the twentieth century after the influence of logical semantics (from Frege to Tarski and Carnap) led to a rethinking of the structure of the semiotic approach (and of its applications for linguistics, logic, and epistemology).

It is worth noticing that the concerns with pragmatics in the field of logic (Tarski, Carnap), related to a better definition of logical semiotics, as well as those of Peirce-Morris tradition gravitate around the problem of “the Interpreter” and of those actions (operations) necessary in order to define the “semiosis.” There is a confrontation between two directions in what concerns the creation of “the theory of significance”: one related to “the semantization of pragmatics”, the second related to “the pragmatization of semantics”<sup>9</sup>, underlined by the later orientation of Wittgenstein toward the analysis of the ordinary language and the connection between the significance and “the use” of the word in a given speech. “The decisive pragmatic orientation in semantics was realized by Wittgenstein. The equation *significance-use*, brought by Wittgenstein, constitutes the most popular concise formula of the new paradigm. The semantics becomes a part of the pragmatics”<sup>10</sup>. From here perhaps derives the need to return to origins, that is to the Greek terms. What is interesting in this approach is not only the *pragma/praxis* synonymy, but also the fact that

<sup>5</sup> Joh. Hoffmeister, *Wörterbuch der philosophischen Begriffe*, 2. Aufl., Hamburg, F. Meiner, 1995, p. 483.

<sup>6</sup> In this respect a vast literature appeared synthesizing new significances, with construction developments, under the title: *PRAGMATIK. Handbuch pragmatischen Denkens*, hrsg. von Herbert Stachowiak. Hamburg, F. Meiner: Bd. I *Pragmatischen Denken von der Ursprünge bis zum 18. Jahrhundert* (1986, 578 p.); Bd. H – *Der Aufstieg pragmatischen Denkens im 19. und 20. Jahrhundert* (1987, 481 p.); Bd. III-*Allgemeine philosophische Pragmatik* (1989, 548 p.); Bd. IV – *Sprachphilosophie, Sprachpragmatik und formative Pragmatik* (1993, 527 p.); Bd. V – *Pragmatische Tendenzen in der Wissenschaftstheorie* (1995, 482 p.). Without attempting a summary, we mention that we have here an elaborated study that benefited from the contribution of several experts who updated the debate of the problem (not only the pragmatic one, but the semiotic, logical and epistemological problems), giving some openings to the current approaching trends. Our references to the mentioned volumes are limited by the paper profile

(the generic language and ontological) and do not present or comment on the study. It is a theoretical and methodological synthesis of reference that we ignore, at the risk of remaining within an outmoded semiotic approach!

<sup>7</sup> H. Stachowiak, *Pragmatische Philosophie*, in *PRAGMATIK*, Bd. V, p. XVII-XVIII.

<sup>8</sup> *Ibid.*, p. XVIII, XIX.

<sup>9</sup> J. Schneider, „Ausprägungen pragmatischen Denkens in der zeitgenössischen Sprachphilosophie”, in: *Pragmatik. Bd IV*, p. 2.

<sup>10</sup> G. Meggle, „Pragmatische Semantik im Ausgang von L. Wittgenstein Sprachspielkonzept”, in: *Pragmatik*, Bd. II, p. 279.

*pragma* leads to "the signified field: action, performance, activity, and fact", and at the same time to a signified field of state of facts and things". The fact was essentially preserved in the adoption by the European languages of the two terms; but of course this process of adoption denotes some changes in what concerns the significance"<sup>11</sup>. Even in its ancient use, the word *πράγμα* denotes "a poor semantic reference, very often used because it determines the most important words in the general communication in the linguistic environment"<sup>12</sup>.

The concept of "pragmatics" becomes itself a "sign" for the different uses of the signs, putting together the logical, empirical, normative, and actional fields.

This true cultural phenomenon, a history of thought *in nuce*, becomes relevant through Peirce and the semiotic shaping which he proposes in order to develop a logic and a better understanding of mathematical essences and functions. Peirce guides his research towards "the sign" in a double direction: sign-action and sign-object, the first being "semiosis", and the second "representamen"<sup>13</sup>.

3. In fact, the pragmatic was introduced by Peirce as both *semiotic* and *philosophy* (the pragmatism). In his well-known study, *How to Make Our Ideas Clear* (1878), Peirce conceived "the pragmatic maxim": "Consider what effects, which might conceivably have practical bearings, we conceive the object of our conception to have, then the whole of our conception of those effects is the whole of our conception of the object"; "our idea of anything is our idea of its sensible effects"<sup>14</sup>. This was called "the Peirce principle". Once extended to the theory of truth, it became "the principle of pragmatism". In a formula named by Peirce "pragmaticism", the principle is: "the content of meaning of every concept is its possible reference to our behaviour."<sup>15</sup>

The interpretation leads to "a community of interpretation", which "have to be a meta-dimension for every theoretico-systematic objectification of the social appearances"; "the schema of such orientation would be a sign-sign; sign-fact; sign-interpreter, that is the three dimensions of the semiotic."<sup>16</sup>

Following Peirce, Ch. W. Morris imposes gradually a three-dimensional semiotics conceived in a behaviourist manner: "the pragmatics of signs" studies the relation between the signs and the person who uses them"; as "a science of human behaviour mediated by signs", semiotics is a pragmatic discipline. This last one can and must understand the operational rules of logical syntax and the rules of signification and truth of logical semantics' taking into consideration the determined rule of behaviour"<sup>17</sup>. It is the pragmatic-behaviourist explanation of semiotics: "semiotics wants to be a general theory of signs in each and every form and shape, for animals or humans, normal or pathologic, persons or social groups. Semiotics became an interdisciplinary procedure"<sup>18</sup>.

4. Pragmatics reorganizes, under many aspects, the linguistic theory, a logic of language proposing a real "pragmatic turn."

Let's take two examples—Wittgenstein's theory of "language games" and the Austin-Searle theory of "language acts"—and consider both of them on the basis of *ordinary language philosophy*, which is dominated by "the so-called theory of employment rule of signification", whose formula is: "the matter isn't the meaning, but the use"<sup>19</sup>.

The term "language games" (*Sprachspiele*) underlines that "the speaking of a language constitutes a part of an activity or a new form of life". That's why a word doesn't have a meaning till it is expressed, or, to put matter even more clearer: "for a

<sup>11</sup> H. Stachowiak, „Einleitung", in: *Pragmatik*, II, p. XIX.

<sup>12</sup> *Ibid*, "Einleitung", in *Pragmatik* I, p. XXI.

<sup>13</sup> G. Deledalle, introduction to Peirce Semiotic", in: *SEMIOSIS* 79/80, No. 3/4 (1995) p. 16.

<sup>14</sup> *Collected Papers of Ch. S. Peirce*, ed. By Ch. Hastshorne and P. Weiss, Cambridge, Harvard Univ. Press, 1931–1958: 5, 402. (The first number indicates the volume, the second the text position).

<sup>15</sup> *Ibid* 5, 460. As I already said, Peirce proposes the formula of "a semiotic transformation of a transcendental philosophy"; "the semiotic logic of the research" instead of "transcendental theory" of experience; "the idea of the triangularity of all spiritual operations is applied to the "cognitive relation" by putting first and foremost "the Interpretant" (K.-O. Apel, "Einführung: Peirces Denkweg

vom Pragmatismus zum Pragmatizismus", in Ch. S. Peirce, *Schriften*, II, hrsg., von K.-O. Apel, Frankfurt a M. Suhrkamp, 1970, pp. 53–54; 79,86).

<sup>16</sup> *Ibid*, pp. 206, 86.

<sup>17</sup> K.-O. Apel, *Transformation der Philosophie*, I, Frankfurt a M., Suhrkamp, 1973, pp. 149,150.

<sup>18</sup> Ch. W. Morris, "Bezeichnung und Bedeutung. Eine Untersuchungen der Relationen von Zeichen und Werten", 1964, in vol. Ch. W. Morris, *Zeichen, Wert, Ästhetik*, hrsg. A. Eschbach, Frankfurt a. M., 1975, p. 199. In this context, the pragmatic stays "that aspect of the semiotics which deals with the origins, the use and the sign actions", with various applications in philosophy. (*Ibid*, pp. 256, 257).

<sup>19</sup> E. von Savigny, *Philosophie der normalen Sprache*, Frankfurt a. M., Suhrkamp, 1974, P–257.

large category of cases of using the word “signification,” this theory can be explained as follows: *the meaning of a word is its use in a language*<sup>20</sup>. In fact, “the practice of language use” consists in this: “one pronounces the words and another one acts accordingly; in learning a language there is a series of things: the one who learns names the objects, i.e. he pronounces the word when the professor points to a rock, for instance. There also could be the case of a simpler exercise: the student repeats after the teacher”<sup>21</sup>.

The text of “Wittgenstein II” (Philosophical Investigations) moves away from the *Tractatus*, which focuses on the “ontological representation” thesis, according to which “the world is arranged in a manner determined by the real facts (*Tatsachen*), which are composed in turn of simple things (objects and attributes)”, the “philosophical criticism” theme being a pretext to offer “a precise picture of this reality”<sup>22</sup>.

But our ordinary language is not built up in this manner, therefore its structure doesn't correspond to any objective relationship between objects and facts, and its meanings “are not precisely and exactly determined”; Wittgenstein had to renounce “the scientific ideal of objectivity” and introduced the following thesis: “the world opens to us only with the linguistic description”; it isn't a given fact in itself, it exists “only through linguistic interpretation”<sup>23</sup>.

Thus Wittgenstein changes entirely the view of language: *language* is an activity among others (which intervenes in different situations and contexts of action), and is “a part of life.” As a result, he proposed the *Sprachspiele* theory, according to which language can be understood and analyzed only by starting from its use in practical determined contexts. “We can think – he wrote – of the entire course of words using as one of those games through which the children learn their mother tongue. I shall call these games, *language games*, and sometimes I will refer to a primitive language in terms of a language game. ... I shall also call the entire language

and the linguistic activities related to it ‘a language game’”<sup>24</sup>.

The phrase emphasizes the fact that “speaking a language is part of an activity or of a life form”<sup>25</sup>. *The meaning* is “the usage in language” and concerns every single word: neither logic nor even less philosophy can overlook this usage”<sup>26</sup>. Maybe that's why Russell considered that Wittgenstein renounced “his best talent”, to invent a theory that “makes futile philosophic activity”<sup>27</sup>. With this sense we must probably take Wittgenstein's formula: “*Denkt nicht, sondern schau*” (don't think, just look)<sup>28</sup>. Thus the idea that “the world” appears as “a situation” (a communication context) and “the rule” doesn't depend on logic but on “the actional situation” “is a pragmatic norm”<sup>29</sup>.

5. Wittgenstein directed his steps towards the pragmatic context, the context of using language. Together with this understanding one goes beyond the narrow abstract understanding of the realist semantics of language, which ceases to be a “simple way of describing”. The semantics itself reveals “a pragmatic background”<sup>30</sup>. Of course, the identification of meaning with its usage in language can raise a series of objections, the first one being that “the usage of language” takes place “in concrete acts of speaking.”

For these reasons, Wittgenstein's theories are likened to those belonging to the theory of *speech acts*. But Wittgenstein's new direction re-inaugurated the pre-scientific, which liberated itself from the prison of “logical atomism.” “Meaning” comes “in the manner of my intention”, in the manner in which I relate to “objects”—a fact that calls for a reintegration of the formal and structural analyses in phenomenology”<sup>31</sup>.

Furthermore, the meaning implies the act of attributing a meaning, the function of signification and “the context of signifiability” (a notion of Dilthey's hermeneutics), which obviously associates the pragmatic and the pre-scientific (together with the

<sup>20</sup> L. Wittgenstein, “Philosophische Untersuchungen”, in *Schriften*, 1., Frankfurt a. M., Suhrkamp, 1980, pp. 300, 310, 311).

<sup>21</sup> *Ibid.*, pp. 292–293. The issue here is the understanding of language in “a pragmatic context” and ascribing the analyses of the signification to this context (Fr. Von Kutschera, *Sprachphilosophie*, 2. Aufl., München, 1975, p. 137), a view that underlines the action of other rules of signs systems: the pragmatic rules.

<sup>22</sup> F. von Kutschera, *Sprachphilosophie*, p. 133.

<sup>23</sup> *Ibid.*

<sup>24</sup> L. Wittgenstein, *Op. cit.*, pp. 292–293.

<sup>25</sup> *Ibid.*, p. 300.

<sup>26</sup> *Ibid.*, pp. 342, 345.

<sup>27</sup> B. Russell, *Histoire de mes idées philosophiques*, Paris, 1961, pp. 269, 271.

<sup>28</sup> L. Wittgenstein, *Op. cit.*, p. 48 (§ 66).

<sup>29</sup> M.K. Munitz, *Contemporary Analytic Philosophy*, New York, McMillan Publishing Co., 1984, pp. 192, 198.

<sup>30</sup> F. von Kutschera, *Op. cit.*, p. 137.

<sup>31</sup> E. Hollenstein, *Linguistik, Semiotik, Hermeneutik, Plädoyers für eine strukturalistische Phänomenologie*, Fr. a. M., Suhrkamp, 1976, pp. 9, 34.

scientific) from a hermeneutic point of view. Zimmermann talked of Wittgenstein's "hermeneutic character of the linguistic thought" which reveals itself "in the delimitation of the natural language as a basis and horizon of any understanding of a 'sense' and 'meaning'"; "the hermeneutics of 'language games'" in *Philosophical Investigations* develops "the most radical critique until now of the traditional representation of 'the language substance' and its functions; in the centre of these opinions there is the theory of picture, developed in *Tractatus*"<sup>32</sup>.

The explanation of *meaning* through "usage" has consequences even for the understanding of "the semiotic triangle." Wittgenstein doesn't criticize "some theory of meaning" but "the form of the argumentation which is common to different standpoints. Since they ask the question – concerning the essence – ('What is the meaning?'), they postulate 'a pure median essence' between the word and the object, sentence and fact, language and reality to be explained; they also ask themselves why a word 'designates' something, a sentence 'presents' something and the language can be 'a picture' of reality"<sup>33</sup>. Therefore, there are three types of relations: I: language-thinking-reality which leads to the "image" (*Abbildung*); II: sentence-idea-fact (*Sachlage*), which leads to "presentation" (*Darstellung*); III: word-concept-object, which leads to "designation" (*Bezeichnung*).

In the scholastic tradition the matter would be expressed as the following: grammar ("*modi significandi*" or "*voces*") – logic ("*modi intelligendi*" or "*conceptions*"); ontology ("*modi essendi*" or "*res*")<sup>34</sup>.

As we can see, "the semiotic triangle" appears three times, its basis lacking "the causal relation", since "the symbol" (in this case: the language, the sentence, the word) stands for referent ('the object of reference'), the relation being an "accepted relation", marked by "the truth"<sup>35</sup>.

6. Above and beyond the difficulties mentioned, the triangular relation through which "the semiotic

situation" appears marks the relations (and the presence) of the semiotic on the other basic levels of "a reference to the world" (*Weltbezug*): linguistic, logic and gnoseologic. Of course, there is no point in connecting the semiotic (as a sign) to the image (copy, reflection, painting etc).

The result is due to the openness through pragmatics to communication and knowledge (more precise: to knowledge in the context of communication). Communication is the most important way of using the language. As Wittgenstein stated: "*Die Sprache muss für sich selbst sprechen*" (the language must express for itself) and that's why the "plural appears": 'language' means "languages", which are 'systems'"<sup>36</sup>.

The reference to communication is, in fact, structural to pragmatics through which result "linguistic alternatives", due to the various relations between the linguistic signs in a particular language. It appears now to the persons and their relations in a community. "The formal aspect of the pragmatic" comprises in its turn the logical syntax and semantic<sup>37</sup>. In fact, on the Peirce-Morris line of thought, the pragmatic tends to become universal, not only erasing the differences but also marking the unity of the three semiotic plans on the formal side. This thing happens not only for the sake of communication needs (as Carnap pointed out), but also for the understanding of truth (especially in formalized languages).

We can conceive syntax, semantics and pragmatics as "dimensions of linguistic truth"; even in Morris's view, in pragmatics as well in semantics and syntax one needs to distinguish not only "a pure formal aspect from a descriptive-empirical one" but also the way in which the formal aspect of pragmatics "circumscribes" logical syntax and "semantics"<sup>38</sup>.

As "a science of human behaviour mediated by signs", semiotics can be considered as pragmatic in its essence. Therefore, we can and must understand semiotically the operational rules of logical syntax and the rules of meaning and truth of the logical semantics as a regulations determined by human behaviour"<sup>39</sup>. This idea originates not only in Peirce

<sup>32</sup> J. Zimmermann, *Wittgensteins sprachphilosophische Hermeneutik*, Fr. a. M., V. Klostermann, 1975, pp. 101, 103.

<sup>33</sup> *Ibid.*, p. 109.

<sup>34</sup> *Ibid.* The *Researches* underline the pragmatic side and the actional function of the language (see Alexandru Boboc, *Limbaj si cunoastere*, Bucuresti, Cartea Universitară, 2005, p. 112).

<sup>35</sup> C. K. Ogden, I.A. Richards, *Die Bedeutung der Bedeutung (The Meaning of Meaning)*, Frankfurt a. M., Suhrkamp, 1974, p. 18.

<sup>36</sup> L. Wittgenstein, *Philosophische Grammatik*, hrsg. von R. Rhees, Frankfurt a. M., Suhrkamp, 1973, pp. 5, 27.

<sup>37</sup> K.-O. Apel, *Transformation der Philosophie*, Bd. I: *Sprachanalytisch, Semiotik, Hermeneutik*, Fr. a. M., Suhrkamp, 1973, p. 150.

<sup>38</sup> *Ibid.*, pp. 141, 149.

<sup>39</sup> *Ibid.*, p. 150.



and Morris, but also in “Wittgenstein II”, in his concept of “meaning-use” which is related to the context of communication and inevitably to human behaviour. This fact can explain the interaction with psychology, with the theory of action and, not least, with phenomenology (intentional analysis).

The unity of speaking-action-thinking demonstrates the so-called “pragmatic turn” in the contemporary thinking (realized in logic, science theory, social theory, philosophy etc). This pre-eminence cannot be generalized: we can talk of universal pragmatic aspects, even of a “pragmatic paradigm”, but there is no use of a reductionism (as a type of pragmatism). The unity of the three dimensions is essential to semiotics, where the sign-function is characterized, first of all, by meaning and interpretation, which are the conditions of accomplishing the speech, of putting it into a shape in order to obtain the impact in action and communication. The theory of “language acts” places the dimensions of signifying and signification, as well as its intention of signifying in a communicational horizon. Using the terminology of “neopositivist logicians” we can say: the Saussurian theory states the need to make a difference between the semantic relation of a statement and its meaning, on one hand, and its pragmatic value given to the statement at the “moment of speaking”<sup>40</sup>, on the other. Searle wrote: ...“a theory of language belongs to action theory; to speak is a form of behaviour governed by rules”<sup>41</sup>.

“The behaviour” is “intentional”, a notion studied by Searle: “the intentionality is a feature that of certain mental states and processes regarding *to* or *about* or *of* certain entities and states of fact or just representing them”<sup>42</sup>.

The true initiator of “language acts” theory was J. L. Austin, representing the so-called “Ordinary Language Philosophy”. He “was the first to develop a systematic theory regarding the question: ‘what can be done with words?’”, which in his specific language had always the form: ‘How to do things with words?’<sup>43</sup>.

In Austin's opinion, “no speech about the usage of words is too precise or systematic”; the problem would be: “what can a person do with a particular

expression?” This leads to the conclusion: “what the expression is in an isolated case depends on circumstances”<sup>44</sup>. “With the help of linguistic expressions we can make all sorts of actions”; such “expressions” are actions or “locutionary acts” and since by “every expression” we always say the “unsaid” too, this kind of action is called “illocutionary act”<sup>45</sup>.

Thus Austin distinguishes meaning (in a semantic sense) from “the illocutionary role of an act of speaking.” The theory of acts of speaking is somehow the opposite of Wittgenstein's late theory, according to which “the use gives a meaning to an expression”. Austin opposes the thesis: we can establish the meaning of a use without being clear how it is used. “I call ‘the act of saying something’ the performing of a locutionary act and, in this context, I name the study of the statements: the study of locutionary acts or of the complete elements of speech”; but “we can say that the performing of a locutionary act is *eo ipso* the performing of an illocutionary act”, for which determination we need to know “how to use the locutionary acts”<sup>46</sup>.

7. Regarding this theory of “speech acts” we need to say that “language games” engaged a “pragmatic turn.”

Briefly, the main thesis is the following: the meaning of a “speech act” is its function; as a function of a linguistic expression, it is based on linguistic conventions which relate to modalities of actions and not to single actions, thus becoming general rules. In a doctrine concerning “those-things-that-are-not-in-their-places”, Austin settles a number of rules; “if they are broken then our speech will be improper”<sup>47</sup>.

The term “accomplishment” derives from the verb “to accomplish”, a verb naturally close to the noun “action”; “it shows that performing a speech means to accomplish an action that isn't conceived as a simple saying of something”<sup>48</sup>. The novelty is: “with the help of linguistic sayings we accomplish different types of actions”; any linguistic saying is at the same time an action (in Austin's terms: a locutionary act) but human language reveals also the

<sup>40</sup> O. Ducrot, “Introduction”, in Searle *Les actes de langage*, Paris, Herman, 1972, p. 10.

<sup>41</sup> J. Searle, *Op.cit.*, p. 53.

<sup>42</sup> J. Searle, *Intentionality and Its Place in Nature*, in *Dialectica*, Vol.38, Fasc. 2–3 (1984), p. 87.

<sup>43</sup> E. von Savigny, *Die Philosophie der normalen Sprache*, Fr. a. M., Suhrkamp, 1974, p. 127.

<sup>44</sup> *Ibid*, pp. 127, 129.

<sup>45</sup> W. Stegmüller, *Hauptströmungen der Gegenwartsphilosophie*, Bd.II, 6, Sufl., Stuttgart, Kröner Verlag, 1979, p. 65.

<sup>46</sup> J.L. Austin, *Zur Theorie der Sprache (How to Do Things with Words)*, Stuttgart, Reclam jun., 1972, p. 114.

<sup>47</sup> *Ibid*, p. 34.

<sup>48</sup> *Ibid*, p. 110.

unsaid (an illocutionary act)<sup>49</sup>. Due to the fact that Austin's text is not quite clear it is absolutely necessary to use some explanations. "The 'locutionary' act is what the speaker said; but in what concerns the 'illocutionary', is not quite clear whether the speech carries out the role of a proposal, of a call, of a disposition, or of a warning"<sup>50</sup>.

8. The theory of "speech acts" raises the issue of dependence of meaning not only to simple language use but on the actional context of speaking; speech is not a simple saying, an enunciation of something. In other words, the limits not only of said things but also of unsaid things—an issue that belongs to the study of human communication in the context of action and behaviour. This fact could justify a study from the viewpoint of logic and the theory of action and also an ontological study in order to understand not only the functioning of the rules but their reason as well. The problem needs a linguistic approach, logical and philosophical at the same time, which means the imminence of the philosophy of language designed to clarify semiotic purposes not only through the pragmatic and actional but also through the study of "the unsaid" side which will come as a "semiotics of silence."

Through philosophy of language is realized "the reference to what we are as generic human beings"<sup>51</sup>. In a semiotic transformation it points to Peirce and to the methodological and theoretical consequences of the "tridimensional" relation of signs. More than that, the philosophy of language is "the science which examines the meaning of language for the human being. Through this characteristic it is different from an unsystematic interrogation unexamined from a critical point of view"<sup>52</sup>. If we take into account the performances of language in their relation to the human being, then we are lead towards anthropology. After all these considerations an interdisciplinary approach seems to necessary. The understanding of the essence and of the functions of language engages, among other things, the metaphysics, poetry and destiny.

<sup>49</sup> W. Stegmüller, *Op.cit.*, p. 69.

<sup>50</sup> E. von Savigny, *J.L. Austin*, in *Grundprobleme der grossen Philosophen*, hrsg. Von J. Spech, *Philosophie der Gegenwart III*, Göttingen, Vandenhoeck und Ruprecht, 1975, p. 209.

<sup>51</sup> J. Simon, *Sprachphilosophie*, Freiburg/München, Alber, 1981, p. 34.

<sup>52</sup> A. Keller, *Sprachphilosophie*, Freiburg/München, Alber, 1979, p. 13.

## References

Apel K.-O. (1973) *Transformation der Philosophie*, Bd. I: Sprachanalitik, Semiotik, Hermeneutik, Fr. a. M., Suhrkamp.

Austin J. L. (1972) *Zur Theorie der Sprache (How to Do Things with Words)*, Stuttgart, Reclam jun.

Botezatu P. *Semiotică și negație, Orientare critică în logica modernă*, Iasi, Junimea.

Carnap, R. (1972) „Sinn und Synonymität in natürlichen Sprachen”, in vol. *Zur Philosophie der idealen Sprache*, hrsg. von Joh. Sinnreich, München, Dtv.

Collected Papers of Ch. S. Peirce, ed. By Ch. Hastshorne and P. Weiss, Cambridge, Harvard Univ. Press, 1931–1958: 5, 402. (The first number indicates the volume, the second the text position).

Deledalle G. (1995) "Introduction to Peirce Semiotic", in: *SEMIOSIS* 79/80, No. 3/4.

Ducrot O. (1972) "Introduction", in Searle *Les actes de lan-gage*, Paris, Herman.

Hoffmeister Joh. (1995) *Wörterbuch der philosophischen Begriffe*, 2. Aufl., Hamburg, F. Meiner.

Hollenstein E. (1976) *Linguistik, Semiotik, Hermeneutik, Plädoyers für eine strukturale Phänomenologie*, Fr. a. M., Suhrkamp.

Keller A. (1979) *Sprachphilosophie*, Freiburg/München, Alber.

von Kutschera F. *Sprachphilosophie*.

Lenk H. (1979) *Pragmatische Vernunft*, Stuttgart, Ph. Reclam jun.

Meggle G. „Pragmatische Semantik im Ausgang von L. Wittgenstein Sprachspielkonzept”, in: *Pragmatik*, Bd. II.

Morris Ch. W. (1975) "Bezeichnung und Bedeutung. Eine Untersuchungen der Relationen von Zeichen und Werten", 1964, in vol. Ch. W. Morris, *Zeichen, Wert, Ästhetik*, hrsg. A. Eschbach, Frankfurt a. M.

Munitz M. K. (1984) *Contemporary Analytic Philosophy*, New York, McMillan Publishing Co.

Odgen C. K., I.A. Richards. (1974) *Die Bedeutung der Bedeutung (The Meaning of Meaning)*, Frankfurt a.M., Suhrkamp.

Rendt H. (1978) *Das Handeln*, in vol. *Handlungstheorien Interdisziplinär*, II 1, hrsg. von H. Lenk, München, W. Fink.

Russell B. (1961) *Histoire de mes idées philosophiques*, Paris.

von Savigny E. (1974) *Die Philosophie der normalen Sprache*, Frankfurt a. M., Suhrkamp.

von Savigny E., J. L. Austin. (1975) in Grundprobleme der grossen Philosophen, hrsg. Von J. Spech, Philosophie der Gegenwart III, Göttingen, Vandenhoeck und Ruprecht.

Schneider J. „Ausprägungen pragmatischen Denkens in der zeitgenössischen Sprachphilosophie“, in: Pragmatik. Bd IV.

Searle J. (1984) Intentionality and Its Place in Nature, in Dialectica, Vol. 38, Fasc. 2–3.

Simon J. (1981) Sprachphilosophie, Freiburg/München, Alber.

Stachowiak H. Pragmatische Philosophie, in PRAGMATIK, Bd. V, p. XVII–XVIII.

Stachowiak H. „Einleitung“, in: Pragmatik, II, p. XIX.

Stegmüller W. (1979) Hauptströmungen der Gegenwartsphilosophie, Bd.II, 6, Sufl., Stuttgart, Kröner Verlag.

Wittgenstein L. (1973) Philosophische Grammatik, hrsg. von R. Rhees, Frankfurt a. M., Suhrkamp.

Wittgenstein L. (1980) „Philosophische Untersuchungen“, in Schriften, 1., Frankfurt a. M., Suhrkamp.

Zimmermann J. (1975) Wittgensteins sprachphilosophische Hermeneutik, Fr. a. M., V. Klostermann.

## THE TRANSCENDENTAL DIMENSION OF HEIDEGGER'S ANALYTIC OF PREDICATION

*Dimitri Ginev*

University of Sofia, and Western Kentucky University  
dimiginev@yahoo.com

### **Abstract:**

The threads of linguistic philosophy in *Being and Time* oppose the prevailing tendency to understand language philosophically within the confines of representational epistemology. In elaborating on the ontological aspects of the view of human beings as inhabiting a linguistically articulated world, the paper stresses the peculiar status of the "fore-structure of understanding" in the discursive articulation as an existential phenomenon of being-in-the-world.

**Key words:** discursive articulation, hermeneutic as-structure, discourse-language distinction, fore-structure of understanding, discursive genres, speech acts.

### **1. Predication and Communication**

In *Being and Time*, Heidegger clearly distinguishes the ontical-logical-epistemological discussion of interpretation as a cognitive procedure from the existential-ontological analytic of the nexus of understanding and interpretation. In so doing, he significantly clarifies his transcendental position. Understanding and interpretation are empirically (ontically) universal. The conditions of possibility of this universality, however, have to be ontologically revealed. Considered in terms of the existential analytic (i.e. as existentialia), the nexus of understanding and interpretation provides a transcendental point of departure in approaching possible kinds of cognizing (including scientific explanation, mathematical deduction, and the "phenomenological intuition of essences"). All of them are "existential derivatives" from that primary nexus "which is one of the constituents of the Being of the 'there' in general."<sup>1</sup> Heidegger's elaborations on this claim put forward a sort of a "transcendental-hermeneutic logic" within the whole program of hermeneutic phenomenology.

At stake in this logic is first and foremost the issue of the status of assertion as a means for "pointing out and determining" a state of affairs within-the-world. In treating the circular relationship between understanding and interpretation (in accordance with the scenario "understanding which interprets"), Heidegger introduces the notion of "hermeneutic as." It characterizes the way of encountering a thing ready-to-hand as an equipment which has been circumspectively taken apart with regard to its "in-order-to." The "circumspective deliberation" delineates the sphere of the pre-predicative experience. *Per definitionem*, the "hermeneutic as" is the phenomenon, which "makes up the structure of the explicitness of something that is understood. It constitutes the interpretation."<sup>2</sup> Encountering a thing ready-to-hand as an equipment amounts to actualizing projected possibilities for using and manipulating that equipment. One encounters such a thing by (re)contextualizing its involvements. In so doing, one is not compelled to take an objectifying distance from the contexts of equipment, predicating thereby the items within this context as "thematic things." The "hermeneutic as" of identifying an equipment in its contextual serviceability and usability operates on a pre-predicative level.

Being embedded in the nexus of understanding and interpretation, the "hermeneutic as" proves to be the principal trait of the paradigm of meaning constitution suggested by hermeneutic phenomenology. Heidegger ascribes a special function to assertion in mediating between the readiness-to-hand and the presence-at-hand. Assertion is a "derivative mode of interpretation" that always already operates within a projected horizon of understanding. As a predicative judgment assertion is grounded on that same "understanding which interprets." Yet in contrast to the interpretation that

---

<sup>1</sup> Martin Heidegger (1962), p. 182.

---

<sup>2</sup> Martin Heidegger (1962), p. 189.

articulates meanings in a pre-predicative manner (through “circumspective deliberation”) within-the-world, assertion classifies (according to the “apophantic as-structure”) what already is meaningfully articulated in the contexts of “concernful dealing.”

Heidegger assigns three basic moments to assertion. It is first and foremost a “pointing out” of what can be spoken out. The “pointing out” is the morphological aspect (in linguistic sense) of the classification achieved by assertion. By means of it a stable semantic content of assertion’s linguistic expression as related to what gets classified comes into play. Through making use of assertions the meaningful world of everyday practices becomes gradually ordered in semantic units. The existentially primordial “articulative interpretation” gets significantly veiled by this semantic-morphological order attained by the classifying power of the “pointing out” as the initial moment of assertion. Yet, being existential condition of the possibility of Dasein’s “average everydayness”, the articulative interpretation (operating through the “hermeneutic as of circumspective deliberation”) is by no means destroyed or replaced by the classificatory use of predicative assertions: While contributing to the formation of language’s morphology, the “pointing out” does not launch an authentication of pure epistemic representations of things that are present-at-hand.<sup>3</sup> (In other words, the “pointing out” does not

transform the temporal unity of the “awaiting-that-makes-present-through-retaining” into a mere presence that is represented by an epistemic subject.)

The next moment of assertion singled out by Heidegger is that of predication. Attributing a predicate to a subject is an “operation” constantly implemented in all talks accompanying everyday practices (or what Heidegger calls “the idle talk of publicness”). Predication is a step of classifying and ordering the world by structuring further (not only morphologically but syntactically as well) its linguistic descriptions. By detaching progressively from “circumspective deliberation” of everyday practices, predication (as it is based upon the initial “pointing out”) indicates a tendency towards an objectifying transformation of what is ready-to-hand within-the-world. However, this tendency proves to be a counterpart of an opposite trend rooted in “circumspective deliberation” – the trend towards keeping “hermeneutic as” intact. Within everyday being-in-the-world constant modifications of utensils that are ready-to-hand in objects that are temporary present-at-hand are taking place. Heidegger’s favored example with the assertion about the heaviness of the hammer is instructive in that regard. Through a predication the hammer is de-contextualized for a while. Its partial objectification, however, serves the function of communicating that hammer’s physical property which makes it dysfunctional in certain contexts of equipment.

The assertion that predicates the hammer as a heavy thing present-at-hand is thus integrated into discursive practices of communication, becoming thereby something ready-to-hand in a concernful dealing. In this case, the as-structure of predication supersedes a “hermeneutic as” of circumspective deliberation in order to induce a new hermeneutic

---

<sup>3</sup> Pointing-out is that moment of asserting which precedes the predication by drawing a preliminary demarcation line between denotation and connotation of what is going to be predicated. Contemporary linguistic theories are facing serious difficulties when trying to equate denotation with extension (the set of all actual entities in the universe of discourse to which the respective word or phrase can be applied), and connotation with intension (all possible entities that the linguistic expression can describe). (See in particular Saeed, 1997. There is also a tendency in linguistics to avoid an identification of denotation with reference. As a result, the way of working out a distinction between denotation and connotation in linguistic semantics seems to be not a matter of substantive difference, but rather of conventionalism: By deciding what semantic characteristics a definite linguistic expression has which allow it to be applied to some things but not to others, one defines the central meaning of that expression (its denotation) as opposed to the secondary meaning (its connotations) acquired through frequent associations. To be sure, this conventionalist approach (which bears also a strong resemblance to later Wittgenstein’s linguistic pragmatism) has its own deficiencies. Yet it is much better than the previous approaches. Hei-

---

degger’s notion of “pointing-out” as the first moment of asserting fits perfectly to the conventionalist way of differentiating between denotation and connotations of a linguistic expression. There is always – so Heidegger’s argument goes – a “hermeneutic as” (a unity of fore-sight, fore-having, and fore-conception) in picking up the central meaning (the basic predicate), when one tries to formulate assertion about something that is “made present” within the interrelatedness of practices of a certain context of equipment. Put differently, there is always a fore-sight, a fore-having, and a fore-conception of defining a denotation of an expression that is involved in a process of predication.

as of communication. To serve as a means of communication's discursive practices is the third moment through which Heidegger specifies assertion. In circulating in a communicative process, the assertion enables the participants to see a definite character of what has been pointed out. Being embedded in an interrelatedness of discursive practices, assertion (as a "derivative interpretation") is characterized by its own contextual fore-structure of understanding. More specifically, it is fore-structured by the contextual possibilities of discursive articulation upon which the communication is projected.

For Heidegger, communication is not an encoded exchange of linguistic expressions but rather an intersubjective being-towards-possibilities of discursive articulation. Communication does not have an intrinsic essence (code, structure). In its projection upon possibilities it is always a potentiality-for-being. On this account, contextual interpretations of assertions within communication are appropriations and actualizations of possibilities of discursive articulation. Regardless of how rigid is the as-structure of predication, each assertion (being fore-structured by such possibilities) is open to further contextual interpretations in the process of communication. On a corollary to this claim, discursive articulation (as a horizon of ongoing appropriation of possibilities) plays the role of a "fore-structure of understanding" for the formation of those linguistic structures (grammatical and morphological paradigms) which are based upon the as-structure of predication.

Generally speaking, there is in Heidegger's analytic of predication an ongoing interplay between both kinds of "as" (i.e. between interpretation that resides in everyday practices and discursive predication). The pointing out and the predication (being embedded in everyday "idle talk" of publicness) are managing to transfer the interpretative way of appropriating possibilities from circumspective deliberation to communication. The "hermeneutic as" operating in a medium of communication (its particular fore-having, fore-sight, and fore-conception) does contribute in its turn to preserving the predicative structure characterizing typically the assertions circulating in the communicative process. The "hermeneutic as" of communication provides a fore-structure (fore-having, fore-sight, and fore-conception) regarding the outcome of actualizing possibilities of being-with through a

dialogue. This outcome – so Heidegger's argument goes – takes on the form of exchangeable assertions.

Now, due to the dominance of "hermeneutic as" over the "apophantic as" there is a strong asymmetry in the aforementioned interplay. The "hermeneutic as" is not solely involved in, but it guides the interplay itself: The exchange of assertions in communications takes place always already in an open horizon of communicative possibilities whereby there is a fore-having, fore-sight, and fore-conception of that exchange.<sup>4</sup> Only the extreme case of formulating a theoretically objective assertion based upon the "mathematical projection of Nature as objective reality" escapes the guidance of "hermeneutic as." (Notoriously, Heidegger addresses this case in his "existential conception of science.") The "hermeneutic as" of communication is embedded in all discursive practices. It is the fore-structure of understanding that makes possible the disclosedness of the world within its discursive articulation. Although not in a systematic fashion, Heidegger makes the case that it is this disclosedness which grounds ontologically the enacting of each and every discursive practice whose linguistic expressions have semantic (ontic) meanings.

The claim that there is a "hermeneutic as" of communication guiding the interplay between the "hermeneutic as" of circumspective deliberation and the "apophantic as" of predication is not to be read in the key of transcendental pragmatics. In mediating between concerned dealings with entities ready-to-hand and thematic predications of entities present-at-hand, communication creates homogeneity within "average everydayness." Yet this does not imply that there is an "idealized rationality" of a dialogue that may provide a normative-structural "framework" (or "codex") determin-

---

<sup>4</sup> Yet to stress this "fore-structure of communication" does not amount to holding that there is an "a priori regulative consensus" embedded in the rational dialogue through the exchange of arguments – a consensus that "in principle" can be made explicit by following the norms of rational dialogue. The "fore-structure of communication" is not a kind of "transcendental inter-subjectivity" since it is not a phenomenon of an inter-subjective (dialogical) mentality that can be normatively reconstructed. It is rather a form of "hermeneutic trans-subjectivity" that does only exist as an ongoing hermeneutic circularity by means of which the articulation of communicative meaning comes into being.

ing the “communicative articulation of being-with one another understandingly.” Heidegger’s hermeneutic phenomenology suggests good arguments against any hypostatization of (emancipatory) communication as a realm of non-instrumental (and non-coercive) rationality of social interaction. However “rational” it could be, the structure of communication through dialogical argumentation cannot provide the “ideal normative principles of rationality” as the champions of transcendental pragmatics are inclined to admit. All processes and procedures of communication presuppose the transcendence of the world through which the meaningful articulation of all modes of being-with-the-others comes into play.<sup>5</sup>

## II. “The Task of Liberating Grammar from Logic”

Interestingly enough, Heidegger touches only incidentally upon one of the most important issues of his analytic of predication. In commenting on the difficulties of having access to the “hermeneutic as”, he makes the case that between the “articulative interpretation” (as an *existentiale*) and the “extreme opposite case” of theoretical assertion about what is an outcome of “objectifying thematization”, there are many intermediate gradations. Cases in point are “assertions about the happenings in the environment, accounts of the ready-to-hand, ‘reports on the Situation’, the recording and fixing the ‘facts of the case’, the description of a state of affairs, the narration of something that has befallen.”<sup>6</sup> These “intermediate gradations” are measured with regard to the degree of detachment of predication’s as-structure from the primordial “as” of circumspective interpretation. In other words, they are indicating the relative weight of the apophantic as-structure within the interplay that is constitutive for (what is now called) a definite “discursive genre.”<sup>7</sup> The examples Heidegger delivers

suggest in addition that one has to scrutinize “discursive genres” in order to establish the “gradation” of a particular kind of assertions. Since the interplay between hermeneutic and apophantic “as” takes place in all discursive practices constituting a particular space of communication, the production of discourse gets specified as a genre whose elements as “units of meaning” do not exist before the meaningfulness of the whole. The genre specification of a discourse consists in that tendency of articulating an open horizon of possibilities of communication, which is characterized by the moments of fore-having, fore-sight, and fore-conception regarding the outcomes of performing speech acts involved in interrelated discursive practices.

Following that line of reasoning, Heidegger goes on to undertake a *prima facie* strangely sounding “task of liberating.” He states: “The task of *liberating* grammar from logic requires *beforehand* a *positive* understanding of the basic *a priori* structure of discourse in general as an *existentiale*.”<sup>8</sup> If we take this claim seriously as a part of the agenda of working out the existential-ontological foundation of discourse, then we have to look for the steps Heidegger envisages in accomplishing the “task of liberating.” He is convinced that the traditional way of making grammar explicit (via linguistic theories) is by searching for grammar’s foundations in (first-order) predicate logic. Yet it is this logic precisely that is based upon the “ontology of the present-at-hand” (*Vorhandenheitsontologie*). One can strive for “deriving” linguistic structures from predicate logic only by assuming that discourse is a set of assertions. This assumption, however, involves a double misinterpretation. First, one wrongly admits that predicate logic provides the foundations of language’s readiness-to-hand within-the-world. In other words, one admits that the “regional formal ontology” of first-order predicate logic lays bare

defined by the way of appropriating possibilities through speech acts, whereby the appropriation constitutes an “implied receiver” who is expected to have a competence of how to make use of the expressions that reach her/him. Thus, the concept of “discursive genre” refers to a pretty complex hermeneutic circularity involving the open horizon of discursive articulation, the performer of speech acts who actualizes possibilities by constituting an “implied receiver” committed to the same horizon of articulative interpretation, and the ongoing grammatical structuring of language.

<sup>8</sup> M. Heidegger (1962), p. 209. Italics in the original.

<sup>5</sup> See on this point Dimitri Ginev (1999b).

<sup>6</sup> M. Heidegger (1962), p. 201.

<sup>7</sup> The concept of “discursive genre” proves to be shared by post-structuralist linguistics, interpretative anthropology, literary criticism, history of ideas, and cultural studies for a very conspicuous reason. By implementing that concept in all studies where the focus is on a “language in status nascendi”, one is no longer trying to codify the linguistic structures in accordance with a set of formal rules. Being specified by genres, the process of structuring is projected upon possibilities. A discursive genre is

the grounds of linguistic (communicative) behavior. Second, the assumption implies (in an objectivist manner) that grammar (as a set of morphological and syntactical paradigms) determines discourse as a medium of communication.

In underlining the significance of the task of liberating grammar from logic, Heidegger looks for a non-mentalist theory of grammar. His elaborations on "discourse and language" in *Being and Time* anticipate the contemporary research programs (not only in linguistics but in cognitive science as well) denouncing the doctrine that the language faculty as a whole should itself be regarded as a distinctive part of human mental equipment. Roughly speaking, Heidegger's outline of a non-mentalist grammar leads to a theory that gets rid of any search for an "ultimate layer of linguistic thinking" distinguished by a totally decontextualized "phrase-structure rules" for generating sentences. For him, the generation of grammatical structural patterns is guided by the interpretative relationships of the "in-order-to", the "towards-which", the "towards-this", and the "for-the-sake-of" of the discursive-communicative articulation of meaning. Notoriously, these are the relationships characterizing both the temporal-ecstatic unity of Dasein with what is ready-to-hand within-the-world, and the "transcendence of the world."

Heidegger is not interested in devising transcendental arguments regarding the formation of linguistic structures (in which predication plays a significant role) through reflecting upon the hermeneutic circularity of discursive articulation. Yet addressing this issue would be a central task within the scope of a "Heideggerian linguistics" – a linguistics that is presuming neither the Cartesian mentalism of "universal grammar", nor the normative mentalism of "phrase-structure grammar", nor the "transcendental inter-subjectivism" of Karl-Otto Apel's pragmatics. The hermeneutic circularity of discursive articulation is not to be located in a kind of "linguistic consciousness." It is rather (to use a Heidegger's expression) "equiprimordial" with the situated transcendence of existence, exhibiting thereby an ontological aspect of the "meaning of Being." A reconstruction of Heidegger's prominent "language-discourse" distinction suggests that discursive articulation consists in genre specified appropriation of possibilities by means of speech

acts.<sup>9</sup> A discursive genre opens up a horizon of possibilities that get actualized by performing speech acts. Neither the horizon nor the particular acts could claim priority in this process. Their involvement in a hermeneutic circularity rules out any kind of causal determinism governing the structuring of language. Consequently, all linguistic structures are open since they are fore-structured by that circularity. To be sure, Heidegger's hermeneutic view of discourse-language is built upon that model of (phenomenological) constitutional analysis which he associates with the nexus of understanding and interpretation. By implication, discursive articulation (and its "ontic counterpart": linguistic structuring) are predicated on an "ecstatic temporality" that becomes temporalized through speech acts. (Briefly, discursive articulation is also a "temporalizing of temporality.")

Being embedded in discursive practices of communication, speech acts (like promising, raising a question, praying, making a threat, ordering, requesting, or describing) take place within the aforementioned relationships of the "in-order-to", the "towards-which", the "towards-this", and the

---

<sup>9</sup> A few words about Heidegger's interpretation of this distinction are in place here. Tentatively speaking, the priority of discourse (articulative interpretation) over language (structure) is a corollary to the basic claim in hermeneutic phenomenology that each existiale (interpretative circularity providing an ontologically necessary condition of meaning constitution) grounds that structure through which it gets an ontic (in particular linguistic) manifestation. Discourse is an existiale whose manifestation in ontic (linguistic) categories gets achieved through syntactic and morphological structures. Worked out as a specification of the "ontological difference", Heidegger's distinction is irreducible to Saussure's celebrated opposition between *parole* and *langue*. It is an opposition that is, first, carried out entirely on ontic level, and, second, based on strong causal determinism. For Saussure, discourse is the linguistic behavior expressed by speech acts produced by real people in real situations. By contrast, language is a pure essence beyond "real people in real situations". Language is an abstract symbolic system that can only be reconstructed as syntactic and morphological paradigms. Accordingly, *langue* (the grammatical structure) determines *parole* (discourse that is similar to Chomsky's "performance"). Heidegger's view of discursive articulation puts forward also a radical alternative to the Sapir-Whorf linguistic relativity hypothesis, according to which the structure of language determines the way of perceiving the world.



“for-the-sake-of.”<sup>10</sup> Since the latter refer to the “transcendence of the world”, Heidegger tacitly assumes that the circularity of discursive articulation is itself fore-structured by the “temporality of Being.”<sup>11</sup> There are clear indications that the issue of how this fore-structuring should look like belongs to the unwritten Part Two of *Being and Time*.

Heidegger differentiates between two aspects in the “task of liberating grammar from logic”: one that has to do with the “existential meaning” of the “is” as the principal means of discursive articulation, and another that takes into account the “temporalizing of temporality” corresponding to that articulation. (Being irreducible to the grammatical copula, the “is” for Heidegger is the basic “operator” of discursive articulation, whose “existential meaning” refers to the whole interplay of hermeneutic and apophantic “as.”) In repudiating the view about grammar as a set of inflected forms grounded upon predicate logic, he makes the case that the hermeneutic circularity of discursive articulation grounds the structuring of syntactic and morphological patterns (paradigms). This claim brings into focus the kind of grammar Heidegger addresses in the liberating task. It is a grammar that unfolds structures of language *in the process of discursive articulation*, or in slightly modified formulation, linguistic structures that are not disentangled from its fore-structure of understanding.

Heidegger undertakes in this regard a criticism of the way in which the “copula is” (i.e. the reduction of the “existential meaning” of the “is” to a “linguistic presence-at-hand”) acquires a privileged status in the mainstream grammatical doctrines. (Of course, Heidegger has in mind the doctrines from the 1920s.) It is a criticism that is in line with the tenets of speech act theory which set the scene of analytical philosophy thirty years after *Being and Time*. This theory unveils a diversity of modes of

using the “is” that corresponds to a diversity of existential modalities as configurations of (illocutionary) expectations (attributed to speech acts) and (perlocutionary) actual effects produced by these acts. In stressing the “existential meaning” of the “is”, Heidegger makes it clear that the “is” “contains” the whole hermeneutic circularity of discursive articulation. The “is” can be reduced neither to a verb which serves only to express identity, nor to a class membership, nor to a bond or binding within a sentence. Actually, the whole existential analytic addresses this meaning since the “is” (the linkage between understanding and interpretation rendering possible Dasein’s being-in-the-world) is grounded in the “ecstatico-horizonal unity of temporality” in a manner that provides an access (through a transcendental reflection) to the unity itself.

The plasticity of the “is” makes possible in the first place the whole spectrum of “intermediate gradations” stemming out from the interplay of hermeneutic and apophantic “as.” Furthermore, the “is” involved in predication is “temporalizing” that horizon in which the predication is fore-structured. In avoiding any substantiation of predicate as a grammatical category in terms of predicate logic that deforms the existential meaning of the “is” (by reducing it to a mere “copula”), one opens the avenue not only to spelling out the phenomenal diversity of predicate in terms of the interplay between hermeneutic and apophantic “as”, but also to studying this diversity against the background of the ways in which “temporality temporalizes itself” as a discursive articulation of meaning. Thus, to overcome the “superficial grammatical theory of propositions and judgments” (Heidegger’s expression) amounts to gaining access to the temporal-ontological foundation of the “is” of predication.

Let me in conclusion return to the way in which Heidegger advocates in *Being and Time* the priority of discourse over language by treating them in terms of the distinction between existentialia and categories. Language is a system of categories (aspect, case, tense, voice, gender, modality, etc.), each of them predicated on the effect of modifying the forms of some class of words. Discourse like understanding and state-of-mind is an existiale (ontological characteristics of Dasein’s being-in-the-world). Following the postulates of existential analytic, Heidegger ascribes a specific “temporalizing of temporality” to discourse. Yet in

<sup>10</sup> More specifically, these relationships are contextualizing the ways of differentiating the three dimensions (locutionary, illocutionary, and perlocutionary) of the speech acts embedded in an interrelatedness of practices of communication.

<sup>11</sup> Let me remind that for Heidegger the “temporality of Being” (*die Temporalität des Seins*) is the ultimate aim of existential ontology, whereas the “temporality of Dasein” (*die Zeitlichkeit des Daseins*) is only a preliminary task within the scope of existential analytic. Notoriously, the former task remained unaccomplished in Heidegger’s work.

contrast to the existentialia of understanding (which is “basically futural”), state-of-mind (which temporalizes itself “primarily in having been”), and falling (which has its existential meaning in the present), discourse does not temporalize itself in any definite “ecstasis.”<sup>12</sup> Heidegger rather adheres to the view that each particular discursive genre has its own “ecstatic regime” of temporalizing of temporality.

The grammatical categories are founded upon discourse as an *existentiale* with a diversity of regimes of “temporalizing of temporality.” The categories do have an “existential-temporal-ontological foundation.” I should like to make sense of this claim by addressing Heidegger’s preoccupation with *aspect* as a grammatical category. In *Being and Time* this is the only explicit discussion of a linguistic issue. Why does Heidegger pay special attention to this category precisely? In my view, because it provides the best illustration of how the liberation of grammar from logic does work. As a grammatical category, aspect refers to the temporal structure of an event. In contrast to tense, however, which is representing processes encountered “in time” (as something present-at-hand), aspects are directly related to what Heidegger calls the “ecstatic unity of temporality.” He makes the following observation: “Aspects have their roots in the primordial temporality of concern, whether or not this concern relates itself to that which is within time. The problem of their existential-temporal structure cannot even be formulated with the help of the ordinary traditional conception of time, to which the science of language needs must have recourse.”<sup>13</sup>

Among the linguistic categories, aspect is that one which is closest to the genre articulation of discourse. Aspect is expressing distinctions in the temporal structure of an event by stressing implic-

itly a mood “embedded” in that structure.<sup>14</sup> Thus, a sentence formulated in perfective aspect expresses an event as an unanalysable whole where the state-of-mind (and mood) of indifference prevails. The description of an event through an assertion formulated by means of progressive aspect expresses usually moods of concern and disturbance about what is going on. A typical use of habitual aspect is related to the linguistic expressivity of boredom. The mood of resoluteness finds in the first place its adequate linguistic form of expression in assertions formulated in conclusive aspect. Sentences in inchoative aspect are tailored to suit the linguistic expressivity of enthusiasm and excitement. The linguistic category of aspect provides an access to the intimate relationship of three things: existential temporality, ways of appropriating possibilities in a given event, and ways of expressing moods of being thrown in that event. This is why the category of aspect brings into being that grammatical structure of language which is closer (as compared with the other grammatical categories) to the existential-ontological totality of discourse. Aspect illustrates in the best way the linguistic disclosedness of being-in-the-world as constituted by understanding, state-of-mind (mood), and falling.

The task of liberating grammar from logic does coincide with the task of grounding the studies into the formation of grammatical structures upon rhetoric.<sup>15</sup> Since discursive articulation is a process of choosing possibilities, there is always the need of justifying the choice. By implication, a process of discursive articulation involves a special art of persuasion regarding the justifiability of the chosen possibilities. For this persuasion belongs to discourse as an *existentiale*, it is also fore-structuring the formation of grammatical categories and structures. This issue, however, requires a separate inquiry.

---

<sup>12</sup> To each of these three basic existentialia corresponds an ecstasy of temporality. Heidegger raises this claim against the background of his doctrine that temporality (as the existential intertwining of the “outside-of-itself in and for itself”) temporalizes with regard to each basic existentialia possible ways of itself. Thus, temporality is not “an entity which first emerges from *itself*; its essence is a process of temporalizing in the unity of the ecstases.” (M. Heidegger, 1962, p. 377. Italics in original.)

<sup>13</sup> M. Heidegger (1962), p. 400.

---

<sup>14</sup> I am speaking here of “mood” as a particular *existentiale* (the disclosing of being-in-the-world as thrownness), and not as a linguistic category expressing the kind of reality attached to an utterance. See in this regard F. Palmer (2001).

<sup>15</sup> See on this task, Dimitri Ginev (1999a).

## References

- Ginev D. (1999a) "From a Strong Hermeneutics of Science to a Strong Rhetoric of Science", *Philosophy and Rhetoric*, 32, pp. 247–281.
- Ginev D. (1999b) "The Hermeneutical Circle of Linguistic Transcendentalism: Intersubjective Validity of Argumentation or the Dialogue that We are", *Thesis Eleven*, 58, pp. 1–18.
- Heidegger M. (1962) *Being and Time*, Translated by J. Macquarrie and E. E. Robinson, Harper, San Francisco.
- Palmer F. (2001) *Mood and Modality*, Cambridge University Press, Cambridge.
- Saeed J.I. (1997) *Semantics*, Blackwell, Oxford.

## TOWARDS A PHENOMENOLOGY OF GRATITUDE— A RESPONSE TO JEAN-LUC MARION

*Peter R. Costello*

Providence College  
pcostell@providence.edu

### Abstract:

Jean Luc-Marion's assertion that Heidegger has not sufficiently addressed the notion of gratitude and the Call is incorrect. Based on Heidegger's discussion in *What is Called Thinking?* of thankfulness and its relation to thinking, I argue that Heidegger indeed articulates a place for gratitude as the proper situation, the proper attitude of phenomenology. While I make an apology for Heidegger, I also note, however, that Husserl's own discussions require more authentic reappraisal within the context of Heidegger's work, thereby reinforcing the notion that gratitude has something to say in terms of the way phenomenology gets built up over time, both in form and content.

**Key words:** phenomenology, Marion, Heidegger, Husserl, gratitude.

"No one knows who is thanking whom for what"—

Jacques Derrida

In his *Reduction and Givenness*,<sup>1</sup> Jean-Luc Marion labels Edmund Husserl "intoxicated" and "bedazzled"<sup>2</sup> by "objectivity," by the richness of mean-

ing as it is brought to light within the phenomenological reduction. For Heidegger, Marion has what seems to be a much harsher judgment. Heidegger was ungrateful: "All that matters is that he himself did not always do full justice to gratitude, whose unconditional anteriority alone was supposed to respond fittingly to the claim of Being" (188). I would like to address both these claims that Marion makes, to challenge them *as* claims, by examining Heidegger's *What Is Called Thinking?*<sup>3</sup> For it is in this work that I find evidence that Heidegger was consistently grateful, that Heidegger recognized the claim that Marion seems to take as Marion's addition to Husserl and Heidegger, and that Heidegger recognized the claim, the call to be beyond Being, to be a call not only towards that which initiates and sustains thinking but also back towards Husserl, to the blossoming tree of phenomenology with whose fruits the rest of us, especially Marion, are in a significant way still nourished.<sup>4</sup>

To begin, I would like first to establish the experiential structure of gratitude, the way it is experi-

<sup>1</sup> I acknowledge that this work of Marion's is part of a trilogy and that Marion himself claims that his intention is not to remain simply faithful to phenomenology but to go beyond *both* Husserl and Heidegger toward a kind of mystical (negative) theology. However, it is the *manner* in which Marion attempts to go beyond them that is problematic for me. This problem is also something Derrida recognizes in his discussion with Marion as cited in *God, the Gift, and Postmodernism*. There, Derrida asks whether or not Marion is "authorized to go directly from the phenomenological concept of *Gegebenheit*, givenness, to the problem of the gift" (57). At issue for Derrida is whether or not Marion is trying to invoke God in the proper way. Marion counter-claims that he is going in the opposite direction, back from gift to givenness, and that thereby "I am not trying to reduce every phenomenon to a gift and then to say that, after that, since this is a gift, and given to a finite mind, then there is perhaps a giver behind it all" (70). Marion claims he is "trying to be fair to some phenomena which cannot be described either as object or as being" (70).

<sup>2</sup> In *God, the Gift, and Postmodernism*, Marion makes an interesting claim by which Marion might absolve Husserl

of his own intoxication (though not Heidegger of ingratitude): "To have an experience of the impossible [i.e., of the gift] means to have an experience of impossibility *prima facie*, which I call the 'counter-experience' of bedazzlement, of astonishment, or *Bewunderung*" (75). It is quite informative to note John Caputo's rejoinder to Marion on this point of bedazzlement: "I do not know how to distinguish the confusion of bedazzlement or excess from the confusion of defect" (78).

<sup>3</sup> Marion claims that Heidegger has not been consistent with regard to privileging gratitude as the source of thought. Heidegger has revised certain manuscripts and exchanged or replaced *Danken* with *Denken*. However, the consistent invocation of gratitude in *What Is Called Thinking?* leads me to believe that Heidegger, as late as the 1950s, and later than his revisions of the texts in question, still privileges gratitude.

<sup>4</sup> In his "The Other First Philosophy," which follows *Reduction and Givenness*, Marion seems to be more sensitive to Heidegger as the one who asks after the meaning of givenness, *Gegebenheit*, and thus the way phenomenology and theology might begin to establish a relationship of gratitude with one another.

enced, in order to devise more clearly what seems to be at issue in Heidegger's *What is Called Thinking?*<sup>5</sup> Authentic gratitude in straightforward lived experience involves both an individual and a communal (or mutual) act of recognition.<sup>6</sup> On the individual side, gratitude is a preservation of what one is thankful for based on the fact that one recognizes that that object, situation, or person has enlarged or coalesced one's experience. In gratitude, one keeps something in mind because it has opened one onto a particular (and particularly relevant) stream of experiences, has provided an organizing route through the multiplicity that before had perhaps receded from view as undifferentiated. In being grateful, one recognizes one's vulnerabilities, one's limits, and uses them in order to allow more sophisticated appearances to appear within one's home.<sup>7</sup>

Still within one's individual experience of gratitude, one maintains a certain posture, that of at-homeness, of reflection, of relaxation and of freedom from anxiety. When grateful, one is at home with the world and with oneself and attributes that at-homeness to a well-maintained tension between self and other, to the others whom one still bears in mind. When one is grateful, one is open towards acting—

one *could*<sup>8</sup> act if it were necessary—and one is incapable of acting out of compulsion—since one is engaged precisely in a kind of reflective awareness upon experience that allows for time to show itself as a meaning that one participates in rather than a measure that constrains.

The communal or mutual act of recognition that authentic gratitude involves is implicit in the individual act, is intertwined with it. In gratitude, one's posture is one's response to the other, one's preservation of that other. But what this means is that one recognizes the other as another self, as a co-participant who has the right to her own paths. Authentic gratitude thus involves a sense of mutual recognition that preserves a duality—I see you or your words or your actions as *recognizing me* as valuable; I see you as valuing me in a way that corresponds to the way I value you; however, I still see the separation between me and you that allows your view to work with and on me. If this communal act is missing, if I cannot handle the duality that gratitude preserves, then what is left in my act of preserving, of holding you in mind is not gratitude but an individual act of self-abnegation or self-subordination that might very well happen to produce something unpalatable, e.g., vendetta or masochism. If this communal act is present as a preservation of duality, however, all parties concerned have the possibility of offering something meaningful and good to one another, of taking what is most valuable in each other to heart and letting it continue to blossom forth.

Gratitude is a structure, an attitude, a posture of human experience that also allows for particular emotions of thankfulness to be felt to various degrees of authenticity.<sup>9</sup> In short, gratitude is the way in which

<sup>5</sup> Lawrence Hatab in *Ethics and Infinitude* notes that it is precisely in Heidegger's pairing of thinking and thanking that his "ontological sense of thankfulness for being can serve as a founding disposition for ethics, a primal mood of grateful reception and openness to the world, to others, and to the fragile marvel of human possibilities" (199). Hatab, however, locates this thankfulness in Da-sein's indebtedness, which I believe Heidegger to be working out of.

<sup>6</sup> Fred Berger in his excellent article "Gratitude" notes that with gratitude "a relationship of moral community is established, maintained, or recognized, consisting of mutual respect and regard. Reciprocation makes the relationship two-way" (302). I would argue that the gratitude that takes one beyond being is still communal, though in a way that articulates community as beyond the particular subjects drawn together.

<sup>7</sup> I also draw a good deal of material on gratitude from Melanie Klein's wonderful essay "Envy and Gratitude." In that essay, Klein notes that gratitude is a structure of human behavior that is absolutely necessary to develop: "Gratitude is essential in building up the relation to the good object and underlies also the appreciation of goodness in others and in oneself. Gratitude is thus rooted in the emotions and attitudes that arise in the earliest stage of infancy, when for the baby the mother is the one and only object" (187). To label Heidegger as ungrateful, then, would be to accuse him of envy and thus of resentment, of not being able to locate the claim or the call. Klein notes: "A particular cause of envy is the relative absence of it in others. The envied person is felt to possess what is at bottom most prized and desired—and this is a good object, which also implies a good character and sanity" (203).

<sup>8</sup> Much of the ethical literature on gratitude is grounded in the *subjunctive* character of its obligations. It could or should be what we enact. A. D. M. Walker in "Obligations of Gratitude and Political Obligation" continues an ongoing argument with George Klosko as to whether "all obligations of gratitude have the characteristic of being generally weak" (362). Walker notes that, even if this is so, and even if we are not experiencing gratitude towards our nation as Socrates did to Athens, nevertheless that is no argument against the power and necessity of gratitude. In his earlier article "Political Obligation and the Argument from Gratitude," Walker argued two important claims. First, he said that "gratitude to institutions must always be derivative in this way from gratitude to individuals" (198). And second, he noted that "the grateful person seeks to communicate to his benefactor that he has certain attitudes" (200). The invocation of the reductions here, phenomenological, existential, and theological, seems pertinent. It is our *attitude* that gratitude establishes, whether toward consciousness, being, or the call.

<sup>9</sup> Patrick Fitzgerald in "Gratitude and Justice" takes a line that is parallel to that of Melanie Klein when he argues vociferously for the inclusion of emotions in ethics and insists that gratitude shows how important cultivating

one always already negotiates givenness, embraces it, and has the possibility of continuing to direct its notice of givenness in its bringing more and more of what is implicit to the fore in light of some greater, more original pull or claim.<sup>10</sup>

To receive a gift well is to receive the other's presentation of something that goes beyond the present of a clock or a quilt;<sup>11</sup> to receive well, to be grateful, is to recognize explicitly that the very possibility of giving comes from the other's participation in or instantiation of a movement of givenness from a source beyond it.<sup>12</sup> This is why gratitude relaxes our anxiety, makes us feel the edges of the world soften at the approach of the gift, why we feel comforted. Yes, we experience our gratitude as opening us onto the other, whom we see as valuing us and our relationship, but the experience of gratitude changes our en-

tire perception of the world, it opens us onto some other vector, changes our mood.

Our ingratitude, conversely, is not simply a refusal of this person and of this relationship. It is a refusal of an entire way of viewing, of a possibility to grasp the impossible. We have not just acted unvirtuously with respect to this person; we have acted to shut down this person's (and our own) claim to participate in goodness itself, to bear references to that which cannot be otherwise named and presented.<sup>13</sup>

Marion, to return to the argument from this description, has accused Heidegger of being inconsistent with regard to gratitude, in short, of being ungrateful. Why has Marion done this? One reason that Marion *could* find Heidegger to be ungrateful, though Marion himself passes over it, is the fact that Heidegger notes in section 7 of *Sein und Zeit* that existential phenomenology begins by taking seriously the notion of concealment or *Verdeckung*. Marion means to emphasize Heidegger's passion for concealment as a compliment, i.e., as a way Heidegger has pushed past Husserl: "Heidegger does not conclude the necessity of a drastic reduction to presence that is given up absolutely for consciousness; he concludes, against Husserl, that 'Covered-up-ness is the counter-concept to 'phenomenon' [*Verdecktheit ist der Gegenbegriff zu 'Phänomen'*]" (57).<sup>14</sup> However, Marion has not noticed something about this issue of concealment. What he had not noticed is that Husserl had seen and used it.

For Husserl, the experience of concealing [*Verdeckung*] was the counter-concept (or the complement at the very least) to the experience of overlaying or *Deckung*. In fact, Husserl linked both concealing and overlaying to unconcealment in many of his works from the 1920's onwards, and especially in the *Lectures on Passive Synthesis*. There, Husserl speaks of the different syntheses of overlaying and their connection to concealment and discovery as follows:

emotions become for the sake of a morality of duty: "Gratitude, unlike duty, has an emotional component....It offers a way to avoid psychological patterns that can harm us physically and psychologically, destroy our most important relationships, and lead to actions that harm others" (153). To stop doing and to feel appropriately, that is important for Fitzgerald. It is also important for phenomenology and for postmodernism—the real reduction is the reduction to gratitude.

<sup>10</sup> Walker notes that "obligations of gratitude *are* generated when benefits are the result of mixed motives and conferred from a mixture of self-interest and goodwill" (1988, 200). I would claim that it is gratitude that allows us to sort out the mixture, to see through it towards what is better. It is gratitude in the face of the going and coming of experience that allows us to move beyond *ressentiment*, as Heidegger argues.

<sup>11</sup> Christopher Wellman, contrary to Berger, uses these objects, the clock and quilt, to talk about gratitude as a virtue and not a duty in his article "Gratitude as a Virtue." For Wellman, gratitude as a virtue that must be developed and nurtured is necessary to preserve reciprocity, which is always threatened by our freedom that is not fully inscribed by duties. Wellman only notes gratitude within the relationship of exchange, and shame should be placed on those who do not recognize what the other has given us. I feel compelled to ask, following both Heidegger and Derrida—particularly in Derrida's analysis of the impossibility of the gift in *Given Time I: Counterfeit Money*, the following question: Is not the experience of gratitude as an emotion and a structure of experience, that is, as an attitude, what precisely takes us out of the cycle of exchange? Is not the 'moral phenomenology' that Wellman claims to be operative one in which gratitude does its work best precisely by no longer holding within explicit awareness the prior gift that motivated our grateful posture? Wellman maintains it is always clear to whom one owes gratitude. Heidegger is right in acknowledging that thanking in terms of thinking has a particularly difficult kind of clarity.

<sup>12</sup> Berger notes that gratitude is "a response to benevolence" (299) that makes gratitude different from other kinds of reciprocity.

<sup>13</sup> Berger notes that gratitude can have pathological enactments that destroy its power. Gratitude that seeks to place the other in debt, and gratitude that "involves the tendency to overemphasize it and to ritualize it" lead to oppression (304). This last concern appears to be something that Heidegger and Derrida both recognize.

<sup>14</sup> See note 3 above. Marion in *God, the Gift, and Postmodernism* notes that the power of the 'counter-experience' of wonder, of the impossible "is not nothing, but a new kind of modality. As Jacques Derrida explains it very well, if the possible, in this moment of philosophy, that is, after the end of metaphysics, is precisely the experience of the impossible, then the only rationality able to match the impossible as such will be the experience of the counter-experience" (75). It seems to me that Marion is in debt to Heidegger here for the notion of counter-experience; it also seems to me that Marion is acknowledging the kind of duality that Heidegger will start to claim at the end of *What Is Called Thinking*?

“In comparing matters with similar characteristics [*Ähnlichen*] we find two things that stand out, (i) the synthetic coinciding in a commonality [*synthetische Deckung in einem Gemeinsamen*], that is in sameness, and yet (ii) the synthetic conflict [*synthetischen Widerstreit*] of particular matters of this commonality that repress one another reciprocally in the process of overlapping [*Überdeckung*]. Repressing means that one conceals [*verdeckt*] the other, that the concealed element [*Verdeckte*] tends toward unconcealment [*Aufdeckung*], then breaking through conceals [*verdeckt*] the previously unconcealed element [*Aufgedeckte*], etc.” (LPS, 176).

This kind of description is also found in *Experience and Judgment*<sup>15</sup> and in the volumes on intersubjectivity, where, curiously enough, Husserl disposes of the notion that *Deckung* means simple “coincidence” by noting that overlaying, in the kind of overlaying and pairing proper to the Fifth Cartesian Meditation, the pairing between oneself and the alien, happens at a distance, with absence already part of the equation:

“For example, next to one tree appears another tree, and in similar manners of appearance. Thus I cannot be conscious of them in any other way than as ‘both’ together, as a pair. From the one to the other as thus appearing goes an awakening and a type of *Deckung par distance* to the other and vice versa” (H XV, 26, my translation).

All of this is to suggest that Heidegger *might* have been ungrateful had he tried to suggest that it was his insight alone that enabled him to come up with the significance of concealment at the heart of meaning, at the heart of Being, at the heart of Da-sein. However, neither Marion nor I condemn Heidegger for this.

What then does Heidegger do to warrant Marion’s disapproval? Why can Marion claim that there is a third reduction necessary, beyond Heidegger’s new reductions of boredom and anxiety? Heidegger’s analysis of boredom does, Marion acknowledges, make “the call as such appear” (198). What Heidegger does not do is to acknowledge the “pure form of the call” as preceding both objectivity and Being, both constitution and interpretation: “That which gives itself gives itself only to the one who *gives himself over* to the call and only in the pure form of a

confirmation of the call, which is repeated because received. The claim of Being itself can call only in putting on this pure form—which Heidegger, however, persists in silencing” (197-198). Marion would “think givenness as such” but Heidegger would stop short with givenness as Being (205).

I find that Marion basically accuses Heidegger of not being attentive enough to Husserl, to the Husserl of *Formal and Transcendental Logic* and of *Ideas I*. The Husserl who spoke of the region as itself a pure form, the region of regions, is necessary; one has to, like Pascal, carry him in one’s pocket, if one is to find the pure “form of the call.” Heidegger has turned away too fast, Marion implies. Heidegger is a re-working of Husserl that needs to be re-worked again, re-conducted toward what Heidegger covered over, concealed, silenced, perhaps without recognizing it, through Heidegger’s own ingratitude.

I would like now to demonstrate Heidegger’s gratitude to Husserl, contrary to Marion’s arguments and silences. In *What is Called Thinking?* Heidegger moves through and sanctions Nietzsche’s rejection of *ressentiment* for the one who can think the eternal recurrence in order to deliver the call of thinking. To move beyond resentment, to free the will “from its revulsion against time, against time’s mere past, when it steadily wills the going and coming, this going and coming back of everything” (WITC, 104)—this is the move towards gratitude. And indeed being freed of resentment allows one to be thankful: “In giving thanks, the heart gives thought to what it has and what it is. The heart, thus giving thought and thus being memory, *gives itself* in thought to that to which it is held” (141, my emphasis). Heidegger then does recognize that thought is a kind of call and that one must give oneself over to this call—not in a mode of exchange<sup>16</sup> and not as a duty. The holding that holds the heart is at once higher and more gentle and more terrible than either of these relations.

Heidegger is most grateful, his gratitude is most apparent, however, in his own invocation of Husserl throughout this book. He acknowledges his debt to Husserl not only in his examination of the “conjunction” of the logical and the noetic. But he also acknowledges Husserl’s example with an example, the

<sup>15</sup> Husserl describes the relation of likeness in *EAI* in the following manner: “the feature of B coincides [*überdeckt*] with the corresponding feature of A, and coincides without any ‘gap’ [*ohne jede ‘Abstand’ überdeckt*], is completely one with it, so that what is covered [*Verdeckte*] is seen entirely through the covering [*durch das Verdeckende*]” (EAI, 190). It should be noted that there is also the possibility of seeing *with* a gap.

<sup>16</sup> “But thanking enacted by itself, as payment and repayment, remains too easily bogged down in the sphere of mere conventional recompense, even mere business” (WITC, 141). Heidegger here presages Derrida’s *Given Time*. But I would want to argue that *What Is Called Thinking?* links Marion and Derrida. The final discussion of unconcealment and its relation to gratitude is something that moves between the gift as non-causal and the gift as non-indebtedness. See John Caputo’s remark at the end of the discussion between Derrida and Marion in *God, the Gift, and Postmodernism*: “Economy for Marion means causality. Economy for Derrida means credit and debit. I worry whether we do not end up in debt in Marion” (77).

concreteness of phenomenology in the most concrete terms. In one of the last sections of the book Heidegger returns to section 88 of Husserl's *Ideas*, the section in which Husserl uses a blossoming apple tree to articulate the relationship between the noetic and the noematic, between the logic of thought and its motivational structure, between immanence and transcendence. In Lecture X of Part II in *What is Called Thinking*, Heidegger notes that thinking calls us to honor the duality of beings and Being. It is here, at the end of this lecture, that Heidegger suggests that the response one offers to the call is and can be thinking—thinking as enacting the duality of beings and Being in a particular way, in the way of gratitude: "Useful is the letting lie-before-us, so (the) taking-to-heart too: beings in being" (*WICT*, 228).

Like Husserl, then, Heidegger found that there was something momentous in our call to hold together two counter-concepts—letting lie, and taking to heart. Both demand that we appreciate not only that which lies and that which we hold but *also* that which calls us to do both, that which allows us to do both, without our being torn apart. For Husserl, that transcendence appears within immanence but that transcendence nevertheless maintains its references that point outwards and backwards, its references to a place that is not strictly locatable within immanence—this is what causes wonder. And gratitude is called for in this recognition, just as a blossoming tree calls for pleasant reverie and an open, relaxed hand.

It is the "blossoming" that has allowed Heidegger to get to this point, the "blossoming" that designates both the tree and the enactment of the opening, the participle that holds the duality of noun and verb together without enacting a sheer coincidence: "these words are participles because what they state is always applied to what is in itself twofold" (*WITC*, 221). It is Heidegger who pays homage to givenness and to the relaxation of gratitude in the blossoming of Husserlian insights. That Heidegger moves out of the Husserlian framework of the noetic-noematic correlation is correct but is a much more problematic remark, then, if it is true that in this experience of the duality of "blossoming" one finds some kind of conjunction of *legein* and *noein*.

In the very last lecture of this book, Heidegger notes what has been previously discussed: "We learned that it [thinking] signifies memory, thanks, thinking that recalls. Since then, we have heard no more of such matters along our way" (*WITC*, 229). Heidegger thus recognizes the lack of the explicit appearance of gratitude as important. We have not heard it spoken since he announced it explicitly. Does that mean we have not been involved with it all along? Does Heidegger's silence mean he has concealed something, forgotten it, or has it become part of the way in which the revelation of what has be-

come unconcealed, namely the call that calls for thinking, arises? Has Heidegger's inability to get beyond a duality, even in embracing it with gratitude, not prevented him from hearing "another call—no doubt the call of the Other" (*RAG*, 197)?

It seems that Heidegger's concern to note the "conjunction of *legein* and *noein*" (*WITC*, 231) and to ask through them "what is That which directs us into thinking" (230) have everything in common with Marion's "pure call." Marion's claim is that the very structure of Da-sein as it recognizes itself in its existential analytic, the "there" that is hyphenated both *away from* and *towards* its being, suggests a "there" outside of being. Following Levinas, then, Marion wants to maintain that my most fundamental reduction is beyond consciousness, beyond Da-sein, toward the "me" that is always already called. But Heidegger's conclusion at the end of *What is Called Thinking?* seems to be precisely this.

At the end, instead of looking at Da-sein or at Being through Da-sein, Heidegger is looking at "unconcealment" and gratitude, at the very experience of the "there" outside of being: "Even, and in particular, that unconcealment in which this rise and entry takes place, remains concealed, in contrast to the unconcealed present things" (237). This unconcealment, which one recognizes only by turning to it in gratitude, allowing it to show itself as well as taking it to heart, this unconcealment, this call from beyond, is what Heidegger himself recognizes as beyond phenomenology, both transcendental and existential, in a certain sense: "Let us note well...the presence of what is present, and not what is present as such and not Being as such, nor both added together in a synthesis, but: their duality, emerging from their unity kept hidden, keeps the call" (242).

At the end, Heidegger *asks*. He gives himself over to that which calls him. He asks, with Husserl and not against him, with the technical skill of a philosopher and with the humility of one whose past commitments remain within the difficult willing of the eternal recurrence. Heidegger ends by asking what he felt compelled to find in the past of language and philosophy: "Can thinking take this gift into its hands, that is, take it to heart, in order to entrust it in *legein*, in the telling statement, to the original speech of language?" (*WITC*, 244).

Heidegger asks not whether we can let *be* the unconcealment, let *be* the duality of beings and Being. We presumably have begun that task in the very asking of the question. Now, we must let *be* in order to *take in*, to "re-conduct" in Marion's sense, this gift towards language. It is after all in language that the traces of this gift are sedimented, just as in English the root of thinking lies in *thanc*. And it is we who, through examining the sediments, have come to stand



on the Earth<sup>17</sup> in a new way, with open hands, reduced to an informed openness and a will to move beyond the eternal recurrence towards that which allows the circle to circle. It is we who are learning how to approach the gift with renewed heart, blossoming into a new appreciation, into a new attitude.

## References

Berger, Fred. R.( 1975) "Gratitude." *Ethics* 85 (4), July: 298–309.

Caputo, John, ed.(1999) *God, the Gift, and Postmodernism*. Bloomington: Indiana University Press.

Fitzgerald, Patrick.(1998) "Gratitude and Justice." *Ethics* 109 (1), October: 119–153.

Hatab, Lawrence J.(2000) *Ethics and Finitude: Heideggerian Contributions to Moral Philosophy*. New York: Rowan and Littlefield.

Heidegger, Martin.(1968) *What is Called Thinking?* Trans. Fred D. Wieck and J. Glenn Gray.New York: Harper and Row.

Husserl, Edmund.(1966) *Analysen zur Passiven Synthesis. Aus Vorlesungs- und Forschungsmanuskripten 1918–1926*. In *Gesammelte Werke*, Band IX. De Haag: Martinus Nijhoff.

*Cartesianische Meditationen. Die Krisis der Europäischen Wissenschaften und die transzendente Phänomenologie. Eine Einleitung in die phänomenologische Philosophie. Gesammelte*

*Schriften*. Band 8. Ed. Elizabeth Ströker. Hamburg: Felix Meiner Verlag, 1992.

*Cartesian Meditations*. Dordrecht: Kluwer Academic Publishers, 1977.

*Erfahrung und Urteil. Untersuchungen zur Genealogie der Logik*. Ed. Ludwig Landgrebe. Claassen Verlag Hamburg, 1954.

*Experience and Judgement. Investigations in a Genealogy of Logic*. Eds. John Wild and James M. Edie. Evanston: Northwestern University Press, 1973.

*Formal and Transcendental Logic*. Trans. Dorion Cairns. The Hague: Martinus Nijhoff, 1978.

*Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy*. Trans. Richard Rojcewicz and Andre Schuwer. Dordrecht: Kluwer Academic Publishers, 1989.

*Analyses Concerning Passive and Active Synthesis: Lectures on Transcendental Logic*. Dordrecht: Kluwer, 2001.

*Zur Phänomenologie der Intersubjektivität. Texte aus dem Nachlass. Erster Teil: 1905–1920*. Ed. Iso Kern. Den Haag: Martinus Nijhoff, 1973.

*Zur Phänomenologie der Intersubjektivität. Texte aus dem Nachlass. Zweiter Teil: 1921–1928*. Ed. Iso Kern. Den Haag: Martinus Nijhoff, 1973.

*Zur Phänomenologie der Intersubjektivität. Texte aus dem Nachlass. Dritter Teil: 1929–1935*. Ed. Iso Kern. Den Haag: Martinus Nijhoff, 1973.

Klein, Melanie. *Envy and Gratitude and Other Works*. London: Hogarth, 1984.

*The Psycho-Analysis of Children*. Delacorte, 1975.

Klosko, George.(1989) "Political Obligation and Gratitude." *Philosophy and Public Affairs* 18(4), Autumn: 352–358.

Llewelyn, John. "A Footnote to Phusis," *Research in Phenomenology* 31: 39–61.

Marion, Jean-Luc. (1998) *Reduction and Givenness. Investigations of Husserl, Heidegger, and Phenomenology*. Trans. Thomas A. Carlson. Evanston: Northwestern University.

"The Other First Philosophy and the Question of Givenness," *Critical Inquiry* 25 (Summer 1999): 784–800.

Walker, A. D. M. (1989) "Obligations of Gratitude and Political Obligation." *Philosophy and Public Affairs* 18(4), Autumn: 359–364.

"Political Obligation and the Argument from Gratitude." *Philosophy and Public Affairs* 17 (3), Summer 1988: 191–211.

Wellman, Christopher. (1999) "Gratitude as a Virtue," *Pacific Philosophical Quarterly* 80 (3), September: 284–300.

<sup>17</sup> John Llewelyn in "A Footnote to Phusis," explains how phenomenologically, there is, in addition to a readiness-to-hand and presence-to-hand, Heideggerian distinctions that Husserl's reduction discovers in the manner of something's appearance, also a readiness-to-foot or *Zufussenheit*: "Momentarily without movement, not going anywhere yet, yet the [resting] climber is all feet and all earth....The climber is entirely captive to *Zufussenheit*" (51). A reduction within an attitude of rest, sprawled out on the rock, enjoying the earth, one's relation to it, the tiredness that comes from enacting it, Llewelyn notes that "so the presence that is enjoyed is not that of presence-at-hand, which is a present of such straight-ahead intentionality. If there is any intentionality left, it is the intentionality qualified by Levinas as inverse" (52-53). It is then, for Llewelyn, gratitude for phusis that is 'behind' or founding the gratitude for the other: "it may be a debt of gratitude for the other human being, and even of gratitude for that gratitude" (59). It is the relation with the earth as more than the ground of our thoughts that allows us to glimpse the source of givenness itself in an experience that makes sense. Like the climber resting, as we too give heed to our standing on the earth, we find ourselves "ultimately and initially relying on something as inhuman as a rock beneath its foot, and therefore, neither simply a being nor simply being, but owing, owing thanks" (60).

## ***DER UNTERGANG DES ABENDLANDES IN THE BULGARIAN CULTURAL AREA***

***Nina Dimitrova***

Institute for Philosophical Research – BAS, Sofia, Bulgaria  
ninaivdimitrova@abv.bg

### **Abstract:**

The main purpose of the research over the presence of Spengler's famous book "Der Untergang des Abendlandes" in the Bulgarian intellectual area between the two world wars is to find explanation about the power and significance of the prophecy that the book contained – especially in the "Bulgarian case." An accent is put on the interaction between Spengler's ideas about the decline of the West, and the Eurasian movement whose manifesto carried the emblematic title "Back to the East." The conclusion is that Spengler's book played its most important role with regard to this vital question for Bulgarian national self-consciousness – the definition of Bulgarian cultural identity.

**Key words:** Spengler, national and cultural identity, autochthonous nationalism, eurasianism.

"The Decline of the West." Moderate prices.  
Something never seen before."

Hermann Hesse. Steppenwol

"No one can deny the enormous influence that Oswald Spengler exercised on political, social, philosophical and historical ideas, not only in Germany but throughout the cultured world. His book *The Decline of Europe*, translated into all languages, came as a forewarning of the decline of the Western world at a time when this decline was indeed expected, when all humanity, after the war, had premonitions of the approaching end of a world, and that what had been created in the course of thousands of years, would disappear into the ocean of time, even as the cultures of known and unknown ancient peoples had disappeared." (Janev 1934a: 97) The Bulgarian philosopher Janko Janev wrote this at a time when, in his words, "Spengler is no longer in vogue", but "still stands in the center of the cultural problem field of our time." (Ibid: 98) When the first volume of Spengler's *The Decline of the West* appeared in 1918, immediately after the staggering experiences of the First World War, after the complete loss of all footholds, it proved to accord exactly with the general apocalyptic tone of society.

It is quite understandable that the idea of the decadence of West European culture provoked a significant response in Bulgaria. Of course, the wave of moods resulting from the thought of the actual or approaching decline of West European culture was not solely connected with the name of Spengler. Ortega-y-Gasset pointed out in *The Revolt of the Masses* "in recent years there is much talk about the decline of Europe. I ardently plead that people not be so naive as to think of Spengler whenever there is talk about the decline of Europe or the West. Before his book appeared, the whole world was talking about this, and the success of the book was due to the fact that such a doubt or concern, engendered by a variety of feelings, had already taken hold of everyone." (Ortega-y-Gasset 1993: 131) But the theme of Spengler's book thrilled the Bulgarian intelligentsia not so much in connection with the actual or alleged decadence of the West, as by the prophecy it contained about the expected flourishing of the East, of Slavic culture. This aspect of the widely discussed work caused considerable excitement in Bulgarian cultural life, and provoked reasoning on the historical destiny of Bulgaria.

The high interest in the prophecies of the German philosopher subsided in Bulgaria as everywhere else, but its presence endured in the Bulgarian intellectual area (according to data presented by Vladimir Svintila, lectures on Spengler, in a critical aspect, were held by the professors Dimiter Mihalchev and Assen Zlatarov in about 40-50 library centres throughout Bulgaria. (Svintila 2006: 234-240)) We must stress another important circumstance in the interwar period in Bulgaria, which served as a catalyst for the reception of Spengler's work. The Eurasian movement had made its much talked about start, and its manifesto carried the emblematic title "Back to the East." Of course, we should recall that the collective work of the first four Eurasians was published in Sofia in 1921, so

the interaction between the two theses had a truly explosive impact. Both books affirmed the idea of the uniqueness of national development, and that a nation's path was separate, so that "catching up with" others became therefore unnecessary; such ideas were soothing to the sensibilities and national complexes of the Bulgarian intelligentsia at that time. In addition to the autochthonous nationalism, which was later to take intense nationalist forms, also emblematic for the early 1920s were philosophical discussions based on the opposition between culture and civilization, where the decaying Western civilization was denounced and national culture was extolled. Before enumerating the currents of influence and the forms in which Spengler's work was present in Bulgaria, I would like to give some brief facts: until the first complete translation into Bulgarian that came out in 1994, only the introduction to the celebrated work had been published (in 1931), but Bulgarian philosophers and scholars of culture, etc., were certainly acquainted with the full two volumes. Starting from the Liberation (1878), philosophical thought in Bulgaria was oriented to European models; understandably Germany, a center of European philosophical activity, was particularly attractive for Bulgarian intellectuals. Most young Bulgarian philosophers had received their education in Germany and had the opportunity to establish direct contacts with many eminent figures of the German cultural scene at that time. This tendency continued till the end of the interwar period.

As for the large number of publications devoted to *The Decline*, this can be explained, as Atanas Stamatov pointed out, by the fact that "Spengler was not simply 'the latest fashion' in Europe, but a challenge to the whole gamut of European philosophical and historical literature. His words sounded prophetic; his tone, apocalyptic; the message was understandable for practically any more or less intelligent person. In such a situation, it was a matter of proving one's professionalism to try to respond to the challenge. Almost all Bulgarian authors permanently engaged in the problems of philosophy and history did make this effort. This is why I would tentatively define the attitude to Spengler in Bulgarian philosophical and historical thought as a 'ritual' one." (Stamatov 2000: 94)

In enumerating the Bulgarian intellectual responses<sup>1</sup> to Spengler, first in order is definitely the

book by the noted Bulgarian specialist in German studies, Konstantin Galabov *The End of Western Culture according to Spengler*, published in 1925. In periodicals devoted to the humanities, such as "Zlatorog", "Scientific Thought", "Scientific Tribune", etc. there were articles and materials that aimed at popularizing the work. A noteworthy article by Rayna Stancheva-Andreeva, devoted to Spengler and entitled "Spengler and Nietzsche (Dependence and Opposites)" appeared in the journal "School Review" in 1944. Since only the first part of her study was published – the rest was lost during the bombardment of Sofia that year, we can only judge that the text pointed out what the author saw as the limitations and schematic nature of Spengler's conception of history. (Stancheva-Andreeva 1944: 63) In the only *specialized philosophical journal* of that time, "Philosophical Review", two theoretical articles were published on topics connected with *The Decline* and the fate of this work, as well as several briefer reviews. "Philosophical Review" in general was a journal whose authors who did not special sympathize with the rising spirit of nationalism in the mid 1930s, and this tone was, no doubt, set by the editor in chief, **Dimitar Mihalchev** (In connection with this he was occasionally qualified and identified in various respects with Marxism and Marxist internationalism). Whereas the problem field around nationalism, national identity, national destiny, etc., was closely connected with the reception of Spengler's work in Bulgaria. The head of propaganda in the Third Reich, Otto Dietrich, in his "Philosophical Foundations of National Socialism" (a work that was translated into Bulgarian), wrote that "Individualistic thought ends in pessimism and the prophecy of the 'decline of Western culture'. Community-minded thinking and national socialist action open the door to a new epoch of social advance, a new flourishing and development of West European culture." (Dietrich 1941: 86) The basis of refutation of Spengler's prophecy about the breakdown of Western culture, lies in the new interpretation of the *subject* of that culture, in the shift of emphasis from the individual to the united "we", to the cohesion of the monolithic national body, a sure guarantee for the advance of a new Europe. (Perhaps we should recall that quite a few Bulgarian public figures and intellectuals who were adherents of the "new order",

interest in Spengler has not passed in our country either." (Davidov 1994: 37) However, a number of publications at that time disprove this estimate.

<sup>1</sup> In the forward to the new, 1994 edition of Spengler's work, Assen Davidov writes: "...though sporadic, the

were also passionate anti-Europeans, who perceived Europe as synonymous with a decadent, decayed civilization.)

Focused as it was on the value of one's own nation, Bulgarian nationalism, which was quite a strong current during this period, understandably depreciated the European; this attitude was provoked initially by the unjust peace treaties that *Europe* had ratified, but was due also to ardent rejection by the "new order" of the values of liberalism and democracy. *Europe* became a symbol of civilization in the Spenglerian sense; *European* acquired a pejorative connotation; it embodied everything that an autochthonous culture in course of formation should preserve itself from. (An emblematic example in this respect is **Janko Janev's** essay "The Philosophy of the Fatherland", 1934, in which the dichotomy "culture – civilization" and the problem of "building" European culture into a civilization were expressed entirely in tune with Spengler.) Anti-Western feelings were not simply fashionable, they had an impressively wide scope, encompassing the radical attitudes of the proponents of the Arian Proto-Bulgarian principle, but also all sorts of defenders of national uniqueness.

Similar to the solution proposed by national socialism as a way out of Spenglerian pessimism, **Dimitar Mihalchev** also insisted on a different understanding of the subject of future European culture, but his course was in a contrary direction to Spengler's and also to the "latest" German ideology. In his journal Mihalchev printed a letter from a reader, who wrote: "Mr. D. Mihalchev! *Philosophical Review* has published on several occasions articles on Spengler and his ideas about the end of European culture, and also on the corresponding views of Paul Valéry... I feel that you *personally* owe your readers a clarification as to the question of the destiny of present-day culture. Are there no signs that the new European world is now old and the time is coming when its culture will be exposed to the same dire lot that befell the cultures of ancient Egypt, of Babylon, of ancient Greece, Arabia, etc.? In any case, the German *kultur* philosopher Spengler, the one who most systematically and dismally portrayed the end of Western culture, seems to be right in his diagnosis of our times and in his prognosis about the future. But would you express your opinion on this question?" (Mihalchev 1981: 584)

In his answer Mihalchev pointed out that West European culture was undoubtedly undergoing some

kind of crisis, something was indeed decaying (obviously he was thinking of "greedy imperialism") and disappearing, something else was coming in its place. But Western European culture in itself is not, as Spengler claims, some isolated phenomenon; in this respect, according to the Bulgarian philosopher, what we have is simply a change of the social subject of this culture, which Mihalchev connected with the success of the colonial movement.

We see that the assessments of Spengler and his book in Bulgaria were situated in quite a wide range – from complete support to strong rejection. Occasionally there were views of the following type (curiously, these were voiced by theologians): the German author was claimed to be a "great thinker", and his work "epoch-making", for "he was the first who dared speak about the future with the confidence of a clairvoyant..." (Panchovsky and Panchovsky 1943: 209). But in general, a more sober and critical tone prevailed in Bulgarian assessments.

Overall there were three main dimensions of the presence of Spengler in the attention of Bulgarian intellectuals:

1. Some Bulgarian authors discussed Spengler's concept in its quality of a philosophy of history and a philosophy of culture. (As the philosopher Janko Janev pointed out, the past years since the publication of *The Decline* had dispelled the initial enthusiasm about the book, had revealed the falsity of a number of assertions contained in it, sensational views had ceded to more objective considerations, but Spengler remained in the focus of attention – "In any case, nowadays no one can examine the problem of culture or history without taking a stand regarding Spengler." (Janev 1934a: 98-99));

2. Commentaries on the decline of the West in connection with the destiny of Slavic peoples. Bulgarian Slavophilism and Bulgarian Europeanism were both rather marginal currents in the interwar period;

3. The search for a national cultural identity as far as it was motivated by Spengler's work.

\* \* \*

### 1. Spengler's historiosophic conception

Spengler's ideas about the hermetic and isolated nature of separate cultures met with hardly any approval by Bulgaria authors. Most of them were

critical on various points in this conception. For instance the literary critic Georgi Konstantinov was categorical that the Spenglerian type of generalizations “can in no case win a place either as a method or a variant of penetrating into the impulses of our history, of our spirit.” (Konstantinov 1933: 151) But even **Konstantin Galabov**, Spengler’s most enthusiastic Bulgarian admirer, viewed the German thinker’s views about the decline of the present-day epoch of Western civilization, analogous with that of Greek civilization, as problematic and unconvincing. He also rejected the categorical pessimism of Spengler, which Galabov considered to be an inevitable consequence of the view that cultures were unable to interact. Like most Bulgarian authors, Galabov did not agree with the view that cultures were isolated from one another. He definitely believed – and many Bulgarian intellectuals agreed on this point, that *history* did not signify the history of separate peoples, and that unity was achieved by continuity of cultural forms, which enables us to view history as a unity: “A cultural past in which are interwoven people of completely different mentalities, some sort of Martians, for instance, would be absolutely inaccessible to our experience, hence to our understanding. History as an experience and understanding are possible only because people are identical.” (Galabov 1925: 59-60) Thus, very generally speaking, the most outspoken Bulgarian admirer of Spengler’s work, for all his respect, claimed that this work “will not play the role of a gospel.” (Ibid: 54) But he left his colleagues with the impression that this was precisely how he perceived Spengler’s book: as a revelation. In fact Galabov’s attitude was ambivalent: in addition to conceptual critique, shared by most Bulgarian authors in their assessment of Spengler, there was also something very essential and emblematic in Galabov’s views: in his own works he attempted to apply precisely those aspects of the theory that he rejected. Not overlooking the element of bias contained in **Vladimir Vassilev**’s judgment (stemming from personal disagreements with Galabov at the time), we cannot deny they are justified: “There is not a single article, Vassilev wrote, in which [Galabov] does not play with the name of Spengler, making references to “his own” article about him. Spengler’s view that cultures live out their lives and others come in their place, a view he deduced from the history of humanity as a whole, Mr. Galabov is trying to “apply” to quite specific phenomena in our history and culture. The result of

applying a global and millennial scale to them can, naturally, be nothing but comic...” (Vassilev 1992: 337) In this commentary Vassilev’s reference is to Galabov’s efforts to present the three phases that the *Bulgarian soul* goes through in similarity with the stages of all cultures: birth, maturity, and decline. The analogies of these stages are: feeling, reason, and will. Thus, inspired by Spengler, in his study “The Physiognomic” [of Bulgarians], Galabov shows how the man of feeling will be surpassed by the man of reason, and the latter in turn by the man of action, of will. This application is probably the only *true* example of the presence in Bulgaria of Spengler’s culturological and philosophical conception, adapted to Bulgarian conditions.

We mentioned the position of **Dimitar Mihalchev**, stemming from his principles: he does not agree with the main view of the German philosopher, who denies *regularity* in historical development. Mihalchev’s position in the field of philosophy of history was often and justifiably identified with the positions of historical materialism. A true Marxist would have criticized Spengler in much the same way. Moreover, Mihalchev believed that Spengler was giving his imagination free rein (“It takes a lot of *fantasy* to *unite*, under the common denominator of the Faustian yearning for the eternal and boundless, the spires of the Cathedral of Amiens with the oratorios of Bach with Leibniz’s integral calculus!” (Mihalchev 1981: 588), and what he presented was not exactly a philosophy of history or culture but a kind of *historiosophy*; considering the Bulgarian philosopher’s aversion to any kind of non-scientific approach masked as philosophy, his categorically negative attitude to the idea of *The Decline* is understandable.

**Ivan Shishmanov** also leveled a conceptual critique at Spengler’s conception. Invited to Germany to lecture on Slavic culture, the eminent Bulgarian scholar commented in his article “The German and the Slavic Worlds”, 1929, on the vagueness of Spengler’s notion of Proto-soul (Urseele): “But what is the proto-soul? We hardly know what the soul is, let alone the proto-soul. Is this expression not simply an empty word within a brilliant but superficial construction? Yet Spengler goes on with his argument, unperturbed. Although the Russian proto-soul is inscrutable, he is obviously quite familiar with the inner life of its feelings, its most secret passions and emotions. He is even informed about the underground instincts of Russia and has much to say about the terrible proto-Russian

hatred of the West, a hatred equal to that of the Apocalypse towards ancient culture, a hatred possessing the gloomy bitterness of the Maccabees.” (Shishmanov 1924: 85)

## 2. The Decline of the West and the Destiny of Slavic Nations

In his well-known essay entitled “The Soul of Europe”, published in the journal “Zlatorog” in 1932, the philosopher **Janko Janev** wrote: “The history of Europe is full of gospel premonitions that the end of civilization is near and in its place a new world will appear. Slavic revolutionary idealism is especially full of such visions, and nearly all social and political Utopians in Europe have not rarely foreseen the approaching end of Western man, weary and without faith in his own forces...” (Janev 1932: 41) Before National Socialism came to power, **Dimitar Mihalchev** noted, Spengler had pointed to Eastern Slavdom as a *dawn* of the European future. (Mihalchev 1981: 589)

Thus the presence of Spengler is especially prominent in the Slavophile trends among Bulgarian intellectuals (trends that were comparatively rare in the interwar period), due to many assertions that tended to exalt the national feelings of “Easterners.”

Curiously enough, the chief proponent of Spengler’s Slavophile messianism in Bulgaria was, once again, **Konstantin Galabov**, a specialist in German studies (and, incidentally, the only authentic “pro-Westerner” among Bulgarian intellectuals at that time), while **Boris Jotsov**, a Slavophile of long standing, in the late 1930s and early 1940s began depicting the Bulgarian as “disappointed in his Slavic self-consciousness”, and stressing the need for closer cultural cooperation with the German people, appealing for the creation of a German Cultural Center in Bulgaria, etc. (Jotsov, the deputy chairman of the Slavic Charity Association in our country, explained that the Slavic principle in Bulgarians was not sufficiently developed to be perceived as a worthy racial trait, which was the cause of the decline of “Slavic illusions” among them. The skepticism of Bulgarians in this respect could not be overcome even by such German thinkers as Herder and Spengler, who “give cultural-philosophical expression to their faith in the genius and mission of Slavs.” (Jotsov 1937-1940: 122) The future destiny of Slavic nations, at least as Spengler outlined it in the first volume of his work, naturally appealed to the Slavophile feelings in Bulgaria, which were rather scarce in the interwar period.

**Konstantin Galabov** pointed out with regard to Spengler’s book: “You might not agree with what you read, you might consider books like these harmful, even immoral, but when you turn the last page, you have the feeling you have come out of the gloomy labyrinth of a pyramid... onto the glorious top, upon which the beams of a new sun are shining, the dawning culture of Slavs.” (Galabov 1925: 62) And further on, regardless of whether Spengler was right about the decline of the West, “for us Slavs, may his second prophecy, about the dawn of Slavic culture, sink deep into our souls – may it become a mystic revelation, something in which one believes without asking for reasons, as one believes in the words of prophets.” (Ibid: 22)

Quite similar are the thoughts of **Aleksander Nikolov** about the “sated and weary world”, which is now waiting for Slavs to say their word. Although in his essay on the Slavic spirit there is something of a political curtsy (it was published in 1947), his characterizations of the Slavic conception of the world, of Slavic culture as a system of spiritual values, are, overall, in line with Russian Slavophile traditions (but they omit the element of mysticism, which Nikolov considered unessential for Slavic culture). As a “member of the great family of Slavic nations”, Bulgarians share such fundamental Slavic characteristics as the yearning for justice, the search for perfection, the willingness to struggle in the name of ideals, the attraction to a global man, a global humanity. Incorporating Russian mysticism, Polish and Serbian romanticism, as well as rugged Bulgarian practical-mindedness, “the Slav is a person who, out of the deepest inner necessity, refuses to set limits to his strivings. Opposed to the hedonism of the West, there is in him an elemental self-sacrificing will burning to perform an apostolic mission, and the ethics of sacrifice and suffering in the name of humanity.” (Nikolov 1947: 124)

The essay discusses Hegel’s well-known thesis about the “non-historic peoples”, with whom he classified the Slavs, and shows it to be untenable; Nikolov affirmed the Slavic spirit as a fact of a cultural-historical order. This point is constitutive of a “Slavophile” philosophy of history in Bulgaria.

## 3. Spengler as Instrumental for Our Own National Destiny

But while for a small part of Bulgarian thinkers the belief in the Slavic future was inspiring with regard to the cultural advance of our country, as part of the advance of all Slavic peoples, other authors

questioned precisely the affiliation of Bulgaria to the Slavic cultural type. The role of Slavophilism as a social and cultural factor in our times was also questioned. (Cf. Mihalchev 1936) The search for the Bulgarian national identity that various representatives of the Bulgarian intelligentsia were then conducting persistently (as well as some manifestations of nationalism) is the third and most important mark of the intellectual presence of Spengler in our country. (One example out of many: when the philosopher and theorist of art **Atanas Iliev** grounded the national characteristics of culture, he referred to Spengler (Iliev 1944: 32-33), as did a number of Bulgarian culturologists). The alternative to *Bulgarian Slavophilism* was the search for a special, unique identity of our own. This approach was very characteristic of **Janko Janev** for instance, but also of others who were overcome by the strong nationalist mood of the interwar period. I believe it would be fitting to place this Bulgarian philosopher among the host of famous thinkers who made the destiny of their own people the object of their cultural-philosophical work, such as Lucian Blaga, Mircea Eliade, Mircea Vulcanescu, and Emil Cioran in Rumania; Lajos Prohászka and Gabor Liuko in Hungary; and Vladimir Dvornikovic and Nikolai Velimirovic in Yugoslavia. (Balázs 2001) The words of Vladimir Dvornikovic, a “theoretician of the Yugoslav race” as he came to be known in Bulgaria as well, sound similar to texts by Yanko Yanev. *The Crisis of West European Culture* (Sarajevo, 1923) is the characteristic title of Dvornikovic’s book from which the following quotation is taken. “This decayed West European cultural type, which so many European thinkers have sentenced to death, will it not eat into our healthy national body as well?” (Quoted from Malinovska 1931: 460)

But for **Janko Janev**, brilliant stylist and connoisseur of Western culture, Bulgarians must not look to the West. For the West “today has no foothold, no dogma, no religion. What can we expect from a world that is on its way to disappearing?” (Janev 1994: 340) This assertion is based on a reading of Spengler as well, and it is categorical. On the other hand, the Bulgarian author definitively rejects Spengler’s ideas about the future blossoming of Russian-Siberian culture and the prophecies of “Spengler & Co.” about the future of the Slavic world. Janev’s attitude to the Slavic cultural-historical factor is, of course, contradictory. In some places he defines Slavophilism as a category of the historiosophic type of thinking, and

as entirely outdated; he writes that the Slavic soul has completely changed, has lost the power of “the apocalypse and the metaphysical hymnic, of religious depth, of national romanticism, and the living God-bearing Word.” (Ibid: 339) This discouraging condition of the “Slavic soul” implies that Spengler’s prophecies are illusory, invalid. Janev expresses these thoughts in an essay published in the journal “Zlatorog” in 1933. In his book *The Heroic Man*, published the following year, Janev seems to be rehabilitating Slavs with regard to their historical importance; he points out the possible role of Slavic nations in world history, who have remained unaffected by the “herd period”, the period when Europeans grow petty and uniform, and which is the last stage in Western culture. According to Janko Janev only Slavs have remained uncorrupted by Western civilization (this term is used precisely in the Spenglerian sense, although in other respects Spengler is used by the Bulgarian author as a support for National Socialist ideas); that is why it can be said they have a future. Slavdom can be conceived of as a separate historical category, and, despite the fact that in our times the Slavic spirit has practically rejected itself with the advent of Bolshevism, the Bulgarian author claims: “But the Slavic spirit of Dostoevsky will be fateful. The soul of Russia will repent and return to itself. It is impossible for the Russian people, divided between the great tenderness of the child and the ferociousness of the villain, to die without achieving its mission...” (Janev 1934b: 32-33)

However this destiny and this mission, which the Russian national consciousness has always ascribed to Russia, Janko Janev assigns to the Bulgarian people. Precisely Bulgaria must be the link between East and West, the great East-West that is traditionally identified with Russia. The Bulgarian people, which has synthesized in itself the Western and the Slavic principle, both law and knowledge, and anarchy and intuition, this people harbors in itself the force needed for a great renaissance: “We are the most primal and virginally elemental people, which has not yet said its word.” (Janev 1994: 341)

In his passionate efforts to separate Bulgarians as a special, unique cultural type, and pointing out their pure pagan and primal character, Janev presents frankly nationalist ideas. Moreover, in drawing closer to the spirit of National Socialism, he moves away from Spenglerian pessimism: “...Germany today, inspired by ideals and seeking new values, wanting to restore the greatness of the



German race, denies the weary and fatalistic spirit of Spengler's philosophy." (Janev 1934a: 109)

The attempts to define the Bulgarian cultural type in the interwar period were the main result, and one of the factors for it was the influence of *The Decline of the West*. The conceptual work achieved in the field of philosophy of history, and provoked by this widely discussed book, was not without importance, as likewise the cultural diagnostic of modern times; but nevertheless Spengler's book played its most important role with regard to this vital question for Bulgarian national self-consciousness – the definition of Bulgarian cultural identity.

## References

(Note: all texts are in Bulgarian except for that by T. Bálazs, which is in English)

Bálazs, Tencsényi (2001) The Nationalization of Philosophy: Bulgarian "National Ontologies" in the 1930-40s." – [www.cas.bg/obl\\_downloads.3\\_3/Balazs](http://www.cas.bg/obl_downloads.3_3/Balazs)

Davidov, Assen. (1994) "The Fateful Event Oswald Spengler." – In: O. Spengler. *The Decline of the West*. Vol. 1, Sofia.

Dietrich, Otto. (1941) *Philosophical Foundations of National Socialism*. Personality and Society. Sofia.

Galabov, Konstantin. (1925) *The End of Western Culture According to Spengler*. Sofia.

Iliev, Atanas. (1944) *National Education*. Sofia.

Janev, Janko. (1932) "The Soul of Europe." – *Zlatorog*, issue 1.

Janev, Janko. (1934 A) "The Rout of European Civilization. Spengler's New Book." – *Social Development*, issue 3.

Janev, Janko. (1934b) *The Heroic Man*. Sofia.

Janev, Janko. (1994) "East or West?" – In: *Why Are We Such? In Search of the Bulgarian Cultural Identity*. Sofia.

Jotsov, Boris. (1937-1940) "Bulgaria and the Slavic World." – *Slavic Voice*, issue 1.

Konstantinov, Georgi. (1933) *Bulgarian Literature after the War (Theories on the Bulgarian Spirit)*. Sofia.

Malinovska, Sofija. (1931) "Slavs and their Cultural Orientation." – *Philosophical Review*, issue 5.

Mihalchev, Dimitar. (1936) "Slavophilism and the New Patriotism in Soviet Russia." – *Philosophical Review*, issue 9.

Mihalchev, Dimitar. (1981) "The Problem of the End of West European Culture." – In: Mihalchev, Dimitar. *Collected Works*. Sofia.

Nikolov, Alexander. (1947) *Philosophical Studies*. Sofia.

Ortega-y-Gasset, José. (1933) *The Revolt of the Masses*. Sofia.

Panchovsky, Ivan; Panchovsky, Petyr. (1943) *The Road of Modern Man Towards God*. Sofia.

Shishmanov, Ivan. (1924) "The German and the Slavic Worlds." – *Zlatorog*, issue 4.

Stamatov, Atanas. (2000) "Ideas about History." In: *Bulgarian Cultural-philosophic Literature [1878-1948]*, Sofia.

Stancheva-Andreeva, Raina. (1944) "Spengler and Nietzsche (Dependence and Opposites)." – *School Review*, issue 1.

Svintila, Vladimir. (2006) *Studies in the Ethnic Psychology of Bulgarians*. Sofia.

Vassilev, Vladimir. (1992) *Studies, Articles, Polemics*. Sofia.

## Selected Bibliography (in Bulgarian)

Spengler, Oswald. Zalezyt na Zapada. Vyvedenie, prev. Kr. Kyrdzhiev (Transl. by Kr. Kyrdzhiev). Sofia, 1931.

Spengler, Oswald. Zalezyt na Zapada, Vol. 1–2. Sofia, 1994, 1995.

Spengler Oswald. Dushata na kulturata. – *ΦΙΛΟΣΟΦΙΑ*, 1991, № 1.

Davidov, Assen. Sydbovnoto sybitie Oswald Shpengler. – O. Spengler. Zalezyt na Zapada. Vol. 1. Sofia, 1994.

Florov, Ljuben. Zalezyt na Zapada. – *Filosofski pregled*, 1936, № 3; *Filosofski alternativi*, 1994, № 5.

Galabov, Konstantin. Kraiat na Zapadnata kultura spored Shpengler. Sofia, 1925.

Gylybov, Konstantin. Upadyk na zapadnata civilizaciia. – *Polet*, 1924, № 6, 7, 8.

Galabov, Konstantin. Spengler za upadyka na zapadnata kultura. – *Zlatorog*, 1923, № 1.

Janev, Janko. Iztok ili Zapad? – Zashto sme takiva? V tyrsene na bylgarskata kulturna identichnost. Sofia, 1994.

Janev, Janko. Razgrom na evropejskata civilizaciia. Novata kniga na Shpengler. – *Obshtestveno razvitie*, 1934, № 3.

Kamenov, Vladimir. Oswald Spengler. – *Nauchna misyl*, 1936, № 29.

Manov, Andrey. O. Spengler. – *Panagiurski glas*, 1927, № 5.

Mihalchev, Dimitar. Problemata za kraia na zapadnoevropejskata kultura. – *Filosofski pregled*, 1938, № 2.



- Nikov, Konstantin. Edin germanski filosof v usluga na reakciata i mrakobesieto (Shpengler kato filosof na vojната). – *Literaturen glas*, 1932/33, № 211.
- Nikolov, Alexander. *Filosofski etjudi*. Sofia, 1947.
- Shishmanov, Ivan. Germanskiiat i slavianskiiat mir. – *Zlatorog*, 1924, № 4.
- Shopov, Kiril. Oswald Spengler i Arnold Toynbee. – *Filosofski alternativi*, 1996, № 6.
- Stamatov, Atanas. *Idei za istoriata v bylgarskata kulturfilosofska knizhnina [1878–1948]*. Sofia, 2000.
- Stancheva-Andreeva, Raina. Spengler i Nietzsche (Zavisimost i protivopolozhnost). – *Uchilishten pregled*, 1944, № 1.
- Vladimirov, Sergey. Oswald Spengler. – *Misyl*, 1936, № 10.

## NON SEQUITUR – SOME REFLECTIONS ON INFORMAL LOGIC

Danilo Šuster

Department of Philosophy, University of Maribor, Slovenia  
danilo.suster@uni-mb.si

### Abstract:

Some general, programmatic points about informal logic are addressed. The informal approach to argument analysis faces serious foundational problems which have been recognized by its practitioners – but informal logic has yet to come together as a clearly defined discipline. Another problem is the dilemma of the dialectician (Sextus Empiricus): informal logic is either trivial or powerless on its own (field expertise is needed). According to Johnson and Blair the central notion in theory of argument is cogency which replaces soundness. An argument is cogent if and only if (i) its premises are rationally acceptable, (ii) its premises are relevant to its conclusion and (iii) its premises provide sufficient reason to accept the conclusion. I propose to understand cogency as a broader notion that includes deductively valid inferences. The criteria of cogency are simply the basic ideals of scientific methodology which requires a respect for available evidence and “reasonable” inference, an awareness of alternatives and a willingness to modify or reject those beliefs that fail to conform to the evidence. Informal logic in the sense of elementary scientific methodology is concerned with proper reasoning and not with proper dialogue. Informal logic involves non-trivial argumentative skills and abilities applied to the subject area and accessible to every normally intelligent and educated person.

**Key words:** informal logic, argument, cogency, scientific methodology, everyday reasoning.

1. I would like to address some general, programmatic points relating to so-called informal logic. My reflections on this issue were provoked, initially, by the practice of teaching logic to undergraduate students of philosophy. Contemporary formal logic offers a well worked out theory; the problem is simply how to make it relevant to philosophical or even everyday reasoning. (According to B. Russell’s ironic remark, strict proofs are to be found in mathematics and theology only, so at least some philosophical relevance of formal logic is guaranteed.) Nowadays, everyday reasoning is more appropriately studied in courses of informal logic and critical thinking. The core notion is usually the no-

tion of the argument and questions such as: What distinguishes good from bad arguments? What makes a good argument succeed? What makes a bad argument fail? It seems sensible to teach elementary “informal” reasoning skills and perhaps combine classes of standard formal logic with classes which outline various non-formal techniques of argument analysis.

Now, by the end of a typical “combined” course almost all of the students will have mastered the required symbolic manipulations (natural deduction, axiomatic proofs). However, when required to informally analyse a piece of “ordinary” reasoning many of them fail badly. Elementary logical “calculi” are apparently much easier to master than the techniques of applied logic. What then should we do with informal logic? What should we teach and how should we test the results? (Mis)achievements are difficult to evaluate – is failure to be explained by lack of knowledge in some specific field of everyday argumentation (not something to be taught in the *logic* class) or by lack of some special logical knowledge (something one *should* master in the logic class)? Are there any *special* skills of informal logic at all? Is it all just a matter of general intelligence and the depth of one’s knowledge?

Peter Medawar, a Nobel laureate in medicine, gave the following comment on scientific method and philosophy of science in general: “If the purpose of scientific methodology is to prescribe or expound a system of enquiry or even a code of practice for scientific behaviour, then scientists seem to be able to get on very well without it. Most scientists receive no tuition in scientific method, but those who have been instructed perform no better as scientists than those who have not. Of what other branch of learning can it be said that it gives its proficient no advantage; that it need not be taught or, if taught, need not be learned?” (Medawar, 1969: 8). Well, informal logic sprang naturally to my mind. Not surprisingly – informal logic, or logic in a broad sense, is basi-

cally just elementary scientific methodology, or so I will argue. Medawar also added: "Of course, the fact that scientists do not consciously practice a formal methodology is very poor evidence that no such methodology exists" (Medawar 1969:9). So, all hope for a more theoretical approach is not lost.

Informal logic, variously called the logic of "real" arguments, the study of the practice of argumentation, critical thinking, argumentation theory, theory of practical reasoning, applied logic etc., is a blooming field of theoretical and practical activity. In their survey article (2002) Johnson and Blair describe informal logic as (a) the theory of fallacies and (b) the theory of (actual, natural) arguments. There is a strong tendency to associate *informal* with *anti-formal*. The growing disenchantment with the capacity of formal logic to set the standards of good reasoning has led to informal logic being described as any approach in logic which either avoids or minimizes the use of formal logic as the theory of analysis (the notion of logical form) and as the theory of evaluation (the notions of validity and soundness) of arguments. The quotes below illustrate this tendency:

By 'informal logic' we mean to designate a branch of logic whose task is to develop non-formal standards, criteria, and procedures for the analysis, interpretation, evaluation, critique and construction of argumentation in everyday discourse (Johnson and Blair, 1977: 148).

... the subject matter of informal logic is different from that of formal logic. Informal logic has the task of developing norms for the evaluation of arguments; formal logic has the task of developing the norms for formally valid implication relations (Johnson and Blair, 2002: 360).

In the emphasis on investigations of natural arguments in their actual settings as opposed to formal techniques of contemporary logic one may detect the remnants of the old divide of the sixties: ideal language vs. ordinary language theorists. Formal plays the role of the "Ugly, Dirty and Bad." Another opposition often found in characterizing the distinction between the two approaches is the one between semantics and pragmatics:

Formal logic has to do with the forms of argument (syntax) and truth values (semantics)... Informal logic (or more broadly, argumentation, as a field) has to do with the uses of argumentation in a context of dialogue, an essentially

pragmatic undertaking (Walton 1998: 418-19, quoted by Johnson and Blair, 2002: 362).

From the pragmatic point of view, any particular argument should be seen as being advanced in the context of a particular dialogue setting. Sensitivity to the special features of different contexts of dialogue is a requirement for the reasoned analysis of an argument (Walton, 2008: 2).

To be fair, not all theoreticians subscribe to the view that informal is simply anti-formal. In his survey article on informal logic Hitchcock remarks:

In any case, the research programme of informal logic does not preclude the use of formal methods or appeal to formal logics. Its distinctiveness consists in its consideration of a set of questions that are not addressed in the specialist journals of formal logic, such as the *Journal of Symbolic Logic* and the *Notre Dame Journal of Formal Logic*, or in such histories of formal logic as that by William and Martha Kneale (1962). It might in fact better be called "theory of argument." Its questions have however traditionally been regarded as part of logic, broadly conceived. The name can thus be taken to refer to that part of logic as traditionally conceived that is not covered by contemporary formal logic (Hitchcock, 2006: 101).

I agree with the spirit of this paragraph, although it does not offer much in the way of a definition of informal logic ("that part of logic that is not covered by formal logic"). Problems with the very definition of the discipline should give us cause for concern. According to Walton and Brinton (1997: 9, quoted in Johnson and Blair 2002: 352): "Informal logic has yet to come together as a clearly defined discipline, one organized around some well-defined and agreed upon systematic techniques that have a definite structure and can be decisively-applied by users."

In his analysis of the structure of science Kuhn introduced a notion of "pre-science" which, so it seems to me, fits the theoretical situation of informal logic well. Before a scientific discipline develops into "normal" science, there is normally a long period of somewhat inchoate, directionless research into a given subject matter. There are various competing schools, each of which has a fundamentally different conception of what the basic problems of the discipline are and what criteria should be used to

evaluate theories about that subject matter. Some disciplines never reach the status of a “normal” science. If we take, now, the *Fundamentals of Argumentation Theory* (1996) as our guide in the theory of argument, the following approaches are listed: theory of fallacies, new rhetoric, critical thinking, dialogue logic and formal dialectics, pragmatodialectics, theory of dialogue, game theory. A variety of approaches signals foundational problems. What exactly is informal logic? Is it really a branch of logic? Is it, or ought it to be, a distinct discipline? How, if at all, is informal logic to be marked off from neighboring subjects such as critical thinking, dialogue logic, and argumentation theory as a branch of rhetoric?

Some of the problems with the theory of argument were noticed long ago, at the very beginning of the discipline, which was first canonized in Aristotle’s *Organon* (the theory of fallacies in *De Sophisticis Elenchis*). Sextus Empiricus in his *Outlines of Pyrrhonism* presented what I would call the dilemma of the dialectician (mentioned briefly by Walton, 1987: 303):

Perhaps it will not be out of place briefly to consider the topic of sophisms, since those who glorify logic say that it is requisite for explaining them away. Thus, they say, if logic has the ability to distinguish true from false arguments, and if sophisms are false arguments, logic will be capable of discerning these as they abuse the truth with their apparent plausibilities. Hence, the logicians, pretending to be giving assistance to tottering common sense, try earnestly to give us instruction about the concept, the types, and the solutions of sophisms.

In the case of those sophisms that logic seems particularly capable of refuting, the explanation is useless; while as regards those for which explanation is useful, it is not the logician who would explain them away but rather those in each art (*techné*) who have got an understanding of the facts (*Outlines II*: 236; Mates, 1996: 167).

This sounds like Medawer on philosophers of science teaching scientific method! Sextus, a physician himself, gave the following example:

In the abatement stage of disease, a varied diet and wine are to be approved. In every type of disease, abatement occurs before the first three days are up. Therefore, it is necessary for the most part to take a varied diet and wine before

the first three days are up. The logician would have nothing to say toward explaining the argument away, useful though such an explanation would be; but the doctor will do so, knowing that “abatement” is ambiguous and refers either to that of the entire disease, or to the tendency to betterment, after crisis, of each particular attack; and knowing also that the abatement of the particular attack occurs for the most part before the first three days are up but that it is not in this abatement but in the abatement of the whole disease that we recommend the varied diet.

Whence he will say that the premises of the argument are incoherent, with one kind of abatement—of the whole disease—taken in the first premise, and another kind—of the particular attack — taken in the second. And thus, in the case of the sophisms that can be usefully explained away, the logician will have nothing to say, but instead he will propound to us such arguments as these: Snow is frozen water. But water is dark in color. Therefore, snow is dark in color. And, when he has gathered together a collection of such nonsense, he knits his brow and takes logic to hand, trying very solemnly to establish for us by means of syllogistic proofs that ... snow is white ...

If, then, logic fails to explain away such sophisms as might usefully be solved, while in the case of those that somebody might suppose that it does explain, the explanation is useless, then logic is simply of no use in the solution of sophisms (*Outlines II*: 236; Mates, 1996: 167-68).

We may take “logic” in this text as logic in the broad traditional sense mentioned by Hitchcock, so here is our dilemma: (informal) logic is either trivial (“snow”) or powerless on its own (field expertise is needed). And to follow the lead of Sextus: the analysis of reasoning is either “contentful”, but this is done by experts, or topic-neutral. Concentrating on topic-neutral reasoning skills will only bring us back to the old and “impractical” formal logic. But learning to reason substantively involves learning about actual subject areas and thus requires expert knowledge. The dilemma is reflected nicely in contemporary textbooks on informal logic (critical thinking): on one side there are general introductions to scientific methodology, enlightened “common sense,” with long passages and an analysis of scientific articles. Working laboriously through the material one

slowly acquires the knowledge of an expert. On the other side there is an open market in literature introducing readers to the wonderful power of how to tell a good argument from a bad one and influence people by being able to put a Latin name on short snippets from the press.

Johnson and Blair (2002: 352) remarked that formal logic began as a revolution at the level of theory that later filtered down into logic textbooks. Textbooks are instruments for passing the latest theory on to the next generation and hence the theoretical developments found in textbooks normally represent the precipitate from the theoretical controversies that occurred earlier in journal and monograph literature. In informal logic developments at the theoretical level were largely motivated by the attempt to teach students how to assess arguments in use. “Be clear about what tools are needed or need to be designed to do the job properly, then develop the theory needed to support the toolkit” (Johnson, Blair 2002: 352). But this “do-it-yourself” theory does not really help with the problems of foundations and our dilemma of the dialectician!

2. According to Hitchcock the research programme of informal logic does not preclude the appeal to formal logic. It should also be mentioned that the research programme of formal logic does not preclude the analysis of real-life arguments. The latest developments in formal logic try to capture the subtleties of natural reasoning (non-monotonic, dynamic, preferential logics etc.). Some formal approaches confront “definitory” rules of standard formal logic with strategic-procedure rules which tell you how to win an argument (Hintikka, 2001). Jacquette proposes a different line of division – natural language versus symbolic language:

I propose that a logical theory or procedure is formal if and only if it adopts a specialized symbolism for representing logical forms that does not occur in ordinary nonspecialized nonsymbolic thought and language. Although I acknowledge that all of logic has to do with logical form, I do not agree that all expressions of logical form must themselves be formal. This distinction captures much of the received concept, since it includes all of symbolic logic and excludes nonsymbolic evaluations of validity or invalidity. As we might expect, formal logic by the proposed distinction will roughly include

everything belonging to what has become the de facto criterion for formal logic in relations expressed by means of standard and nonstandard notational variations and extensions of the propositional and predicate-quantificational calculus. The definition additionally includes schematic and graphic treatments of syllogistic logic that have traditionally been regarded as more properly within the aegis of informal logic and critical reasoning. Informal logic by contrast on the present proposal is limited to the consideration of a proposition’s or argument’s logical form by discursive reconstruction within natural language, the use of counterexamples to discredit inferences, identification of arguments as committing any of the so-called rhetorical fallacies, and the like (Jacquette, 2007: 134).

I do not see a sharp line of division, though. There is the practical value of “formal” patterns recognized by informal theoreticians devising symbolic representations – diagrams, schemes, letters, etc. True, this is not a “pure” symbolic language – but these representations are on a par with, say, graphic treatments of syllogisms. Frege compared the use of symbolic language instead of natural language to the use of a microscope instead of the naked eye in biology. But even in biology one does not always need a microscope and sometimes too much precision is “over-kill.” Jacquette also adds:

The relegation of syllogistic logic, square of opposition, and argument diagramming methods to the genus of informal logic can now be seen as a kind of historical accident. Were it not for the emergence of more powerful algebraic methods of formal logic . . . , there is little doubt that the logic of syllogisms, Venn and other styles of diagramming, etc., would constitute the whole of formal logic as opposed to purely informal nonspecialized nonsymbolic logical methods (Jacquette, 2007: 134).

Yet it is difficult to see where the line should be drawn between an approach that uses argument diagramming methods and one that uses reconstruction within natural language – both will be used in a typical course on informal logic, though neither of them will find a place in a course on mathematical logic. Venn (or Euler) diagrams are sometimes useful in “informal” analysis, but they are, typically, part of more formal approaches.

But I agree with the starting point – all of logic is connected with form in one way or another. Skills of informal logic require a “formal” ability to move in the modal space: what could be true? Non-sequitur is a violation of form insofar as it represents a possibility for your premises to be true and your conclusion false. To take a simple example: “One who does not play cards can still be a bad judge, so one who does play cards can still be a good judge.” (Here is a counterexample: One who is not bribable can still be a bad judge, so one who is bribable can still be a good judge?) The procedure of resolving questions of form is always the same: keep something fixed (logical “constants” in classical logic) and vary the other parts. But what is fixed is not always given in advance and the possibility is not always logical – only relevant options are to be considered and it is notoriously difficult to explicate the notion of relevance. And of course, imagination is restricted by knowledge: if you know nothing about a certain field, then patterns alone will rarely be of any help (the moral from Sextus Empiricus). This explains one of the discrete charms of formal logic: even if ignorant of the topic, look for the logical form of an argument and calculate the result (whether the conclusion follows). But the most interesting arguments are not plain deductions, and for those that are (or are made such in a reconstruction) the interesting question is why one should accept the premises and what the alternatives are.

Johnson and Blair replace the standard of soundness (premise truth and deductive validity) in argument evaluation with a broader notion of argument cogency. An argument is cogent if and only if (i) its premises are rationally acceptable, (ii) its premises are relevant to its conclusion and (iii) its premises provide sufficient reason to accept the conclusion (the R.S.A. test for argument cogency). These three criteria are sometimes augmented with a fourth: (iv) that there are no known better reasons for an opposite conclusion. Fallacies are explained as being violations of one or more of these criteria.

I find the theoretical austerity of this proposal very attractive. But are these not just the ideals of enlightened common sense or elementary scientific methodology which require a respect for available evidence and “reasonable” inference, awareness of alternatives and a willingness to modify or reject those beliefs that fail to conform to the evidence (cf. a popular presentation in Sokal: 2008)? True, Johnson and Blair understand cogency as a replacement of soundness – in order to be good, arguments do not

require true premises linked to their conclusions by valid inferences. But why should cogency not include deductions?

We can agree that in an argument the arguer proposes to the addressee that a set of propositions supports another and thereby invites the addressee to infer the latter from the former. Also, the relation of support may, but need not be, one in which the supporting propositions purportedly deductively imply the supported proposition. As Johnson and Blair say:

Part, though not all, of the normative theory of argumentation has as its subject-matter the answer to the question, “What sorts of support relationship are logically cogent?” We take it that one subset of logically cogent support relationships are ones in which the supporting propositions logically imply the supported proposition, and we take the subject matter of formal deductive logic to consist (inter alia) of the study of forms of entailment, that is, valid implication relationships among propositions or sentences. It will be relevant to whether the addressee ought to infer the supported proposition from those adduced as support for it to know that the latter logically imply the former, but that is neither a necessary nor a sufficient condition of a justified inference....

In addition, cogency includes standards that the grounds or premises of arguments must meet. Depending on the context, they must satisfy a requirement of truth, reasonableness or dialectical acceptability. We take informal logic to be the study of all the conditions under which an addressee ought to infer the target proposition from those adduced by the arguer as support for it. Thus the subject matter of informal logic is different from that of formal logic. Informal logic has the task of developing norms for the evaluation of arguments; formal logic has the task of developing the norms for formally valid implication relations (Johnson, Blair 2002: 360).

I agree with most of what is said, but the last sentence is puzzling. According to Johnson and Blair informal logic is about arguments and formal logic is not? They deliberately go against a use of “argument” according to which logical proofs or demonstrations are counted as arguments. I agree that the sets of unrelated sentences or propositions they quote as an example of valid arguments should

not be counted as arguments (of the type "The moon is not made of green cheese. So,  $2 + 2$  is 4.") But why could a (strictly deductive) proof that the square root of two is irrational not be a good argument? And what is wrong with more mundane reasoning that uses deductions, say: "The player who won both Wimbledon and the US Open in 1982 was the greater player in that year. Connors won both. Therefore, Connors was the greater player" (suppose I am wondering whether McEnroe or Connors was the greater player in 1982 and a friend of mine offers me this argument, Jackson, 1987: 102). The argument offered is a perfect logical deduction, but it is a good and helpful argument!

As Jackson notes: "The act of propounding an argument may have brought a half-buried piece of information to the surface, may have alerted me to the relevance of certain facts to my final concern, or may simply have enabled me to see how to get it altogether, so as to make transparent what I want to know" (Jackson, 1987: 102). Jackson calls this the "teasing-out" function of propounding arguments. There is another function of valid arguments – when an argument is propounded with the purpose of convincing someone of the truth of its conclusion then the person propounding the argument usually offers the hearer considerations that support the premises. The person to whom the argument is propounded is implicitly invited to "borrow" these considerations and to be persuaded by them of the truth of the premises, and hence of the truth of the conclusion. In short: "The utility of valid argumentation (over and above the teasing-out function) in convincing audiences of conclusions lies in the evidence implicitly offered for borrowing by the presentation and selection of premises" (Jackson, 1987: 107).

The evidence offered for borrowing grounds the premises of an argument, but we need not go into the details of this account to see that exclusion of the deductive from the realm of the "real" argument is unjustified. For Johnson and Blair an argument is a set of reasons, that is, has been, or might be, offered to another person or other persons with the intention of persuading them, or some audience, to modify their beliefs, attitudes or behaviour (Johnson, Blair 2002: 360). This seems to be too narrow – the "teasing-out" function of propounding arguments is ignored. Well, they might point out that reasoning about the greater player in 1982 and a demonstration that the square root of two is irrational is a good piece of reasoning but not an argument (complex social, speech activity). The "teasing-out" function

pertains to proper reasoning and not to proper argument.

We may grant that argumentation is a social and public speech activity involving more than one party, with practical goals (persuasion). One cannot argue without at least an imaginary audience or interlocutor. Reasoning, on the other hand, is a mental activity that can be performed privately and much reasoning is done before and outside the context of argument. Argumentation requires that its participants reason, so reasoning is necessary for argumentation. At the heart of the activity of argumentation is the offering of and response to arguments in the more narrow sense of reasons offered in support of or against claims (Blair 2001: 372). According to Blair 2001 (and perhaps not entirely consistent with other characterizations offered by Blair and Johnson) informal logic is restricted to the normative study of the cogency of argument, which is only one part of critical thinking and argumentation theory. I agree with this restriction, the relationship between reasoning proper and arguments within informal logic is complex, but, to be brief – since reasoning is necessary for argumentation, standards of good reasoning should always be respected in argumentation and good reasoning includes deductions. The norms of informal logic in so far as they embody norms of basic scientific methodology are norms of reasoning, not those of argumentation in the sense of proper dialogue.

Some theoreticians, most notably Walton (2008), understand informal logic as essentially pragmatic. For Johnson and Blair argumentation is a social activity paradigmatically carried out in the medium of a natural language. But even their approach to informal logic is described by Walton as a narrow, text-based, product-oriented approach to argument. In pragma-dialectics, as one of the influential theoretical models of informal logic, argumentative discourse is conceived as a certain activity (a critical discussion) aimed at resolving a difference of opinion by putting the acceptability of the "standpoints" at issue to the test by applying criteria that are both problem-valid as well as intersubjectively valid.

The relationship between reasoning and argumentation within informal logic is complex. Pragmatic factors should be taken into account. Pragmatics is sometimes characterized as dealing with the use of language and effects of context. But this has various meanings – features of the context can be epistemic features indexed to the audience or the

arguer (their belief systems, reasons, doubts, justifications), or conversational features of dialogue. In argument analysis, should we consider epistemic norms or conversational norms? Recall that an argument is cogent only if its premises are rationally acceptable. According to the first interpretation we should look at the beliefs of the audience (and / or the arguer). According to the second we should look at the conversation itself – what the audience has said, not what they think. Fallacies are explained as breaking the rules of proper dialogue. Circular arguments, for instance, are fallacious because they violate normative rules of dialogue which demand consensual starting points (cf. van Eemeren & Grootendorst, 2004).

I think that the pull toward the pragmatic in the conversational sense should be resisted. In a critical discussion the initial conflict of opinions could be resolved, but both parties might still be wrong. But we should be interested in the truth of the matter and not just in justifying our thesis and questioning or refuting the other party's thesis, by reasoned means, using accepted standards of evidence. This looks more like a theory of conflict resolution through argumentative discussion (compare Wreen, 1994: 302). Wreen wonders whether this is a theory of argumentation at all. I am willing to concede that the pragmatic approach is analysing argumentation in the sense of critical discussion as a complex social, speech activity. But informal logic understood as a set of norms of elementary scientific methodology is concerned with proper reasoning.

3. Since the theory of arguments in everyday discourse is at issue, let me proceed by way of two examples. The first one is from a typical textbook – an “enlightened common sense” type (Thomson, 1996: 45):

If people became healthier as the affluence of the country increased, we would expect the population to be healthier now than it was thirty years ago. But over the last thirty years new illnesses, such as chronic fatigue syndrome, have appeared, and we have become more vulnerable to old diseases such as heart disease, strokes and cancer. So the increased wealth of the country has not produced improvements in the health of the population.

The passage illustrates two common flaws of everyday reasoning. Insufficient or suppressed evi-

dence is the first. There is no information given in the passage about the general health of the population – perhaps some old diseases like smallpox and bronchitis are much less common. Perhaps people are relatively healthy for longer periods of their lives, before becoming ill in old age. And, secondly, the key term in this passage, “the health of the population,” is ambiguous. Its meaning is either the percentage of people's lives during which they are free from illness or the number of diseases in the population. If the first, then the argument fails – we do not have enough information upon which to base the conclusion. It seems even more likely that nowadays people live longer without having any disease. The number of diseases has perhaps increased precisely because we live longer (people usually become ill in old age).

I have more or less followed the analysis given by Thompson. The example should remind us of Empiricus and his treatment of “abatement.” This example (from a newspaper report) is somehow dry and colourless. Now add some emotions, engagement in hot issues and the pattern will reappear in more interesting clothing. Consider Paul Feyerabend's drastic views that science is just one of the many ideologies that propel society and that it should be treated as such. Feyerabend claims that negative opinions about astrology and other alternative practices are not justified by scientific research. I understand informal logic as embodying elementary scientific methodology which requires a respect for available evidence and “reasonable” inference, awareness of alternatives and willingness to modify or reject those beliefs that fail to conform to the evidence. The alternative practices so vigorously defended by Feyerabend (herbalism, alternative medicine, judgments based on intuition, rain dances etc.) very often fail to meet precisely those standards. So what are Feyerabend's arguments for his astonishing claims?

In his plea for democratic relativism Feyerabend dismissed the predominantly negative attitudes of scientists towards such phenomena as elitist or racist. In “The Unusual Story of Astrology” (1978) he critically analyses “Objections to Astrology,” a statement by 192 leading scientists (first appeared in *The Humanist* of September/October 1975). The scientists declare that concepts of modern astronomy and physics undermine the principles of astrology. Feyerabend first interprets the principles of astrology as a claim that celestial events (positions of planets, the Moon and the Sun ...) exert influence on



human life. This is a very broad interpretation of astrology – a different, narrower and much more common definition is implied in the text of the 1975 statement: “It is simply a mistake to imagine that the forces exerted by stars and planets at the moment of birth can in any way shape our futures.”<sup>1</sup> This statement is now interpreted by Feyerabend as: “It is simply a mistake to imagine that celestial events (positions of planets, the Moon and the Sun ...) have any influence on human life.” And there is yet a third meaning of astrology in Feyerabend’s text, namely the thesis that celestial events (positions of planets, the Moon and the Sun ...) exert influence on life (in general) on Earth.

Nelson Goodman once wrote (1956) that the practical scientist does the business but the philosopher keeps the books. So let us do some book keeping. We have astrology-narrow (and interesting): the forces exerted by stars and planets at the moment of birth can shape the futures of human beings (their characters, personalities). Next we have astrology-broad: celestial events exert influence on human life. And finally we have astrology-empty (and largely uninteresting): celestial events exert influence on life (in general) on Earth. Well, Aristotle believed that the supra-moon region that covers the Moon and everything beyond is completely different from the terrestrial one and exerts no influence on the sub-moon world. So at least one historical theory is a target of Feyerabend’s ferocious attack. And here is the trick: replace astrology-narrow with astrology-broad, compile a mass of evidence against the denial of astrology-empty and present this evidence as a counter-argument to the denial of astrology-narrow.

Feyerabend meticulously surveys various research reports (the influence of the activities of the Sun on life on Earth, correlations between organic and inorganic processes on Earth and various parameters of the Sun and the Moon; he quotes articles such as “Possible Effects of Extra Terrestrial Stimuli on Colloidal Systems and Living Organisms”, etc.). A lot of respect for scientific results and scientific methodology from someone so strongly opposed to “objective” scientific standards! After being bombarded with this counter-evidence, we read (as a minor objection, according to Feyerabend) that the scientists in the statement also complain that psychologists have not found the slightest evidence that astrology is of any, or even minimal value as an

indicator of the past, present or future features of one’s life. Feyerabend’s reply would exhilarate every PR agency advising politicians: “This argument (against astrology, D.Š.) has no value, when we remember that astronomers and biologists were not even able to find the evidence already published and contributed by their own professions” (presumably the articles quoted in length by Feyerabend).

For starters: first replace astrology-narrow (“the claim that knowledge of the apparent relative positions of celestial bodies is useful in understanding, interpreting, and organizing information about personality and human affairs and other terrestrial matters” is suggested by Wikipedia; “type of divination that involves the forecasting of earthly and human events through the observation and interpretation of the fixed stars, the Sun, the Moon, and the planets” according to Encyclopædia Britannica) with astrology-broad. Next refute astrology-empty (celestial events exert influence on life on Earth) and ignore powerful objections to astrology-narrow on the basis that those who objected to astrology-narrow were not able to find any evidence for astrology-broad. They were not even looking for it, but this will not prevent Feyerabend from switching over to the offensive. PR agencies advising political parties on how to win the elections prescribe the following rule: defense is for losers, when under attack (and justifiably so) do not, ever, retreat but make an assault. So Feyerabend, when faced with a decisive argument against his position (no evidence for astrology-narrow), drops in an offensive ad hominem (more will follow when we approach the end of “The Unusual Story of Astrology”): scientists are incompetent and they are unable to find any evidence.

Basically we have the same blunders as in the “health case” spiced with ad hominem: obliterating different senses of a single term (astrology-broad-narrow-empty) and suppressed evidence. True, life is short and our cognitive resources are limited, but here we encounter a deliberate selection of information. There is a ready explanation for this selection. Some of our beliefs are very precious to us, perhaps we have “invested” a lot in them and would like to protect them at all costs. Denunciations of Western imperialism, his critique of science itself, as well as his concern for environmental issues ensured that Feyerabend has become a hero of the anti-technological counterculture. One might sympathize with Feyerabend’s idea of a free society in which “all traditions have equal rights and equal access to

<sup>1</sup> Compare:  
<http://www.americanhumanist.org/about/astrology.html>

the centres of power.” One attached to those ideals will easily ignore or suppress counter-evidence to the conclusion that “objectively” there may be nothing to choose between the claims of science and those of astrology, voodoo, and alternative medicine. This is the familiar “confirmation bias”, a useful definition of which can be found in Wikipedia ([http://en.wikipedia.org/wiki/Confirmation\\_bias](http://en.wikipedia.org/wiki/Confirmation_bias)):

Confirmation bias is a tendency to search for or interpret new information in a way that confirms one's preconceptions and avoids information and interpretations which contradict prior beliefs. It is a type of cognitive bias and represents an error of inductive inference, or as a form of selection bias toward confirmation of the hypothesis under study or disconfirmation of an alternative hypothesis. Confirmation bias is of interest in the teaching of critical thinking, as the skill is misused if rigorous critical scrutiny is applied only to evidence challenging a preconceived idea but not to evidence supporting it.

And what to make of Feyerabend's claim that theoretical anarchism is more humanitarian and more likely to encourage progress than its law-and-order scientific alternative? Hugh Trevor-Rope once remarked that one of the early effects of the discovery that nature is strictly governed by impersonal laws was to reduce the enthusiasm for burning witches.

4. Let me now try to draw some morals about informal logic. I do not think that there was a great deal of theory involved in spotting “non sequitur” in the arguments above. The main features traditionally ascribed to “the scientific method” are a clear statement of a problem, careful confrontation of theory with fact, open-mindedness, and (potential) public availability or replicability of evidence (entry on “Scientific method” in Routledge Encyclopedia of Philosophy, Version 1.0, London: Routledge). In the spirit of elementary scientific methodology, the starting points of argument analysis are very simple, almost trivial: first, identify the phenomenon you are dealing with. So, what does that (astrology, health ...) mean? What exactly is claimed (conclusion) and what evidence is offered for those claims? I assume that ambiguity could sometimes be interpreted as the problem of relevance and sometimes as the problem of acceptability of a premise (once the various

meanings are clearly distinguished one realizes that the premise is irrelevant or unacceptable) so all of the R.S.A. criteria of argument cogency were involved in evaluation of the two pieces of reasoning above. I could speculate that the pervasiveness of the fallacy of the ambiguity (amphiboly) can perhaps be explained with the way knowledge is encoded. Cognitive economy requires only one “slot,” different senses are then triggered by different contexts. The diagnosis of a dialectician remains the old scholastic: *distingue* (be aware of different senses).

Let us now turn to the dilemma of the dialectician. Here is what Sextus has to say about amphiboly:

In view of these points, then, the logical treatment of sophisms that is so much boasted about by the logicians is useless. We say similar things about the distinguishing of amphibolies. For if an amphiboly is a linguistic expression having two or more meanings, and if linguistic expressions have meaning by convention, then those amphibolies that are worth resolving — such as occur in some practical situation — will be resolved, not by the logician but by the people practiced in each particular art, who themselves have the experience of how they have created the conventional usage of the terms to denote the things signified, as, for example, in the case of the amphiboly “In periods of abatement one should prescribe a varied diet and wine” (*Outlines* II: 256; Mates, 1996: 168).

What kind of knowledge is required to notice the ambiguity in “the health of the population” or in Feyerabend's definition of astrology? Do we have to be experts in the field (physicians, astronomers, astrologists)? I do not think so. General alertness, knowing something about different subjects and general education is required, but one need not be a specialist. And knowing something about different subjects has never been easier than today. I take the gist of Medawer's pessimistic remark on scientific method to be that a typical scientist will absorb the elements and skills of scientific methodology through her education and practice. Something similar is true of informal logic. It is the attitude toward claims, positions, theses ... in general, including those defended or put forward in the public sphere of knowledge and opinion (usually media, matters of “polis” and the like). Informal logic at its core embodies the attitudes of basic scientific methodology

applied to the subject area accessible to every intelligent and educated person. To repeat – cultivating respect for available evidence and “reasonable” inference, awareness of alternatives and a willingness to modify or reject those beliefs that fail to conform to the evidence.

The view of informal logic as basic scientific methodology accords well with Toulmin's (1958, 2003) suggestion that argument analysis is like applied epistemology – in a successful argument sufficient evidence is given to support the conclusion, but the criteria of sufficiency depend on the nature of problems at issue. Take *ad hominem* – it is difficult to believe a broker who advises you to sell your stock while he himself is buying the very same stock. But ignoring the evidence for the proposition that celestial events exert influence on terrestrial life in general does not make scientists skeptical about astrology incompetent and unworthy of belief! According to the textbook *ad hominem* facts about the proponent of the argument are irrelevant for the assessment of her argument. But when practical reasoning is at issue (what one should or should not do), some facts are clearly relevant. It is just not given in advance which facts to take into consideration.

Toulmin considers that logic is “concerned with the soundness of the claims we make” (1958:7), rather than with the task of telling valid from invalid inferences, and he also says that this requires more than mere calculations, it requires “experience, insight and judgment” (1958:188). According to my view the ability to tell valid from invalid inferences apart is a part of logic, but I agree with the stress on practice. In certain important aspects “applied” or informal logic is more like skill (playing piano or chess) and less like a firm corpus of theoretical knowledge. Memorizing dozens of fallacies (“theoretical” knowledge), all of them context dependent (not always fallacious), is useless without a “plain” ability to recognize a non sequitur: the conclusion does not follow – never mind the name – here is a counterexample. Elements of skill, pattern recognition, practice and the general knowledge required to assess arguments on various topics explain the difficulties in teaching this subject. The ability to recognize the same pattern in different contexts and the ability to find an example or the counterexample to the pattern (inference scheme, typical fallacy etc.) seem to be one of the main features of a “critical thinker.”

In the spirit of Medawer's initial remark, the fact that we do not consciously practice a theory is very poor evidence that no such theory exists. We are sometimes confronted with the objection that there is no theoretical coherence to informal logic, which should instead be viewed simply as applied epistemology. I have argued that recognizing a non sequitur is largely a matter of “knowledge-how” (to be developed by practice) and not “knowledge-that.” But viewed from the theoretical perspective, informal logic comprises a variety of “tools.” In a sense, informal logic is a theoretical retro movement. Fallacies, inductive logic and scientific method have been routinely discussed in the much despised elementary logic textbooks since the nineteenth century. Standards were set by Aristotle – recall that Aristotle regarded logic as a tool for every science. This conception is reflected in the collective title *Organon* that we still use to refer to Aristotle's logical writings. *Organon* contains elementary deductive logic and principles of argumentation, the study of fallacies, philosophy of language, the basic principles of definition and explanation, the basic principles of scientific methodology and epistemology etc. Informal logic, or logic in a broad sense, from the theoretical perspective, is such a toolbox, which perhaps explains the variety of theoretical approaches and foundational proposals.

Let me conclude these reflections on informal logic with an observation that ideals of logic in the broad sense are in certain aspects like those of democracy. That is, an extreme respect for form and procedure – even valuable beliefs are cancelled by bad reasons and fallacious reasoning. And the lesson can be more generally applied. We know from legal procedure that if an avowal is extorted, it is not valid. Gettier cases in epistemology have taught us that justified true beliefs do not count as knowledge if not supported in the right way. Some do not like this obsession with procedure. Yet, according to Sokal (2008), ideals of clear thinking, combined with a respect for evidence – especially inconvenient and unwanted evidence, evidence that challenges our preconceptions – are of the utmost importance to the survival of the human race in the twenty-first century. Strong words, but do they justify special courses of informal logic in which “theoretical” terms such as confirmation bias, amphiboly, suppressed evidence, *ad hominem*, etc. are introduced and explained?

As we have learned more about the world, we laugh, indulgently, at the “expertise” of Sextus.

Wine would be very rarely prescribed nowadays in the abatement stage of disease. Expertise is changing, but even today there is an ambiguity to be spotted by the careful reasoner. And it is useful to master the ability to spot a non sequitur. I think that despite Sextus' criticism and Medawer's sarcasm, it is still useful to teach this discipline, if only to reaffirm the value of logic as part of the old *artes liberales*. The education and training deemed suitable for free persons was supposed to equip one with general knowledge, and interest in the pursuit of knowledge in the strict sense (and not, for instance, because of economic interests, characteristic of *artes illiberales*, supposedly less free vocational education). The distinction is nowadays perhaps obsolete, but only because we recognize the value of universal tools in any kind of "artes." Logic in the broad sense, as elementary scientific methodology applied to subject areas accessible to every informed intellectual, is one such universal tool. Matters of politics typically are, but molecular biology and nuclear physics are typically not part of general knowledge – though things are changing rapidly. We live in the era of the explosion of knowledge. It has never been easier to access information; we are only ever a few clicks away from the manual which explains how to build an atomic bomb or the recipe for salmon with almonds. True, specialization is a law of every profession. Yet experts entering the public sphere of knowledge still have to make their results accessible to the rational audience. In order to pass a law on a nuclear power station, members of parliament need not know much (if anything at all) about gaseous diffusion separators. But they should be able to recognize a non-sequitur (require clarification or ask for more information).

## References

- Blair, J. A. (2001) "Walton's Argumentation Schemes for Presumptive Reasoning: A Critique and Development." *Argumentation* 15: 365–379.
- Craig, E. (ed). (1998) *Routledge Encyclopedia of Philosophy*, Version 1.0, London: Routledge.
- Feyerabend, P. (1978) *Science in a free society*. London: New Left Books.
- Goodman, N. (1956) *The Problem of Universals*. Notre Dame: University of Notre Dame Press.
- Hintikka, J. (2001) "Is Logic the Key to All Good Reasoning?" *Argumentation* 15: 35–57.
- Hitchcock, D. (2007) "Informal Logic and the Concept of Argument." In Jacquette, D. ed. *Philosophy of Logic. Handbook of the Philosophy of Science*. Elsevier B.V.
- Jackson, F. (1987) *Conditionals*. Oxford: Basil Blackwell.
- Jacquette, D. (1994) "Formalization in philosophical logic." *Monist* 77: 3, 358–376.
- Jacquette, D. (2007) "On the Relation of Informal to Symbolic Logic." In Jacquette, D. ed. *Philosophy of Logic. Handbook of the Philosophy of Science*. Elsevier B.V., 131–154.
- Johnson, R. H. and Blair, J. A. (1977) *Logical Self-defense*. Toronto and New York: McGraw-Hill Ryerson.
- Johnson, R. H. and Blair, J. A. (2002) "Informal Logic and the Reconfiguration of Logic," in Gabbay, D. M., Johnson, R. H., Ohlbach, H. J. and Woods, J. eds. *Handbook of the logic of argument and inference*. North Holland. Elsevier, 339–396.
- Mates, B. (1996) *The skeptic way: Sextus Empiricus's Outlines of Pyrrhonism*. New York, Oxford: Oxford University Press.
- Medawar, P. (1969) *Induction and Intuition in Scientific Thought*. American Philosophical Society (collected in *Pluto's Republic*, Oxford: Oxford University Press, 1982).
- Sokal, A. (2008) "What is science and why should we care?", [www.physics.nyu.edu/faculty/sokal/sense\\_about\\_science\\_PUBL.pdf](http://www.physics.nyu.edu/faculty/sokal/sense_about_science_PUBL.pdf)
- Thomson, A. (1996) *Critical Reasoning*. London: Routledge.
- Toulmin, S. (2003) *The Uses of Argument*. (1958, 1st edition). Cambridge: Cambridge University Press.
- van Eemeren H. Frans et al., eds. (1996) *Fundamentals of Argumentation Theory, A Handbook of Historical Backgrounds and Contemporary Developments*, Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey.
- Van Eemeren, F.H. van, and Grootendorst, R. (2004) *A Systematic Theory of Argumentation: The Pragma-Dialectical Approach*. Cambridge: Cambridge University Press.
- Walton D. N. (1987) *Informal Fallacies*.

Amsterdam, Philadelphia: John Benjamins Publishing Company.

Walton, D. N. (1998) *The New Dialectic: Conversational Contexts of Argument*. Toronto: University of Toronto Press.

Walton, D. N. (2008) *Informal Logic: A Pragmatic Approach*. 2nd edition. Cambridge: Cambridge University Press.

Walton, D. N., Brinton, A. eds. (1997) *Historical Foundations of Informal Logic*. Ashgate Publishing Company.

Whitehead, A. N. (1911) *An Introduction to Mathematics*. Oxford: The Clarendon Press.

Wreen, M. J. (1994) "Look, Ma! No Frans!" *Pragmatics & Cognition*, 2:2, 285–306.

## CAN WE SAVE A PRIORI KNOWLEDGE?

Nenad Mišćević

University of Maribor  
and  
Central European University  
vismiscevic@ceu.hu

### Abstract:

The paper joins Horwich's criticism of stipulationist accounts of a priori knowledge, and raises some problems for his own account of the a priori. It first questions the assumed separability of scientific investigation and non-scientific assertoric practices in regard to norms of adequacy. It also questions Horwich's Restriction Assumption according to which only the former are answerable to the standards of empirical adequacy and overall simplicity (which threaten apriority in the case of science). Finally, it criticises his argument that inability to think otherwise might guarantee apriority, pointing to science-driven reflective revisability of possibly innate beliefs.

**Key words:** a priori, stipulation, meaning-constitutivity.

### 1. Introduction.

Can stipulation yield factual, substantial *a priori* knowledge?<sup>1</sup> And is our knowledge of logic a central candidate for being accounted for by such an appeal to stipulation? P. Horwich has been defending the negative answer in his ongoing polemic with P. Boghossian, and has proposed a very tentative solution of his own in his paper "Stipulation, Meaning, and Apriority" from the Boghossian and Peacocke volume of "*New Essays on the A Priori*" (2000); I shall be mostly discussing this paper in the sequel. My bad luck is that Professor Horwich has already signaled some change of mind, by taking the crucial pages containing a sketch of his positive account of the *a priori* out of the revised version published in the more recent book (2005). His disagreement is with Boghossian's recent interesting form of stipulationism which attempts to synthesize stipulationist

view of analytic sentences with a factualist account of analytic truths. I want to build upon his negative and his positive results. I should say that I have learned a lot from Horwich, since his truth book, up to his criticism of stipulationists. Now, back to the opening question about the powers of stipulation.

Boghossian has recently argued for an affirmative answer to it, distinguishing stipulationism from conventionalism, and leaning on Kripkean intuitions about contingent *a priori* (see his papers collected in his<sup>2</sup>). He has thus joined, in an original and creative fashion a dominant line in the apriorist or rationalist approach to logical, mathematical and philosophical armchair beliefs and hopefully knowledge. This line relies upon cognizers mastery of relevant definitions-concepts: the mastery gives the cognizer access to relevant facts and thus ultimately helps solving the dilemmas that have plagued the epistemology of the *a priori*. I shall call it "conceptualism", taking as its main proponents Boghossian together with C. Peacocke, and to some extent B. Hale and C. Wright. Here we are interested in specifically epistemological issues, so by "conceptualism" I will in the sequel mean above all an epistemological position. It rejects any reliance on *a posteriori* sources of justification or warrant for logical knowledge, mostly rejects or downplays the need of any kind of broadly causal (or causal-like) explanation of our logical capacities, both of our having them and of their reliability. Boghossian is arguing that stipulations and implicit definitions generated by them offer the royal epistemic road to *a priori* truths, preserving at the same time their factual character. The solution is the compatibilism between the factual and the stipulative.

Horwich is critical of this proposal, and I very much agree with him.

On the positive side Horwich offers an interesting conjecture about why we hold some truths to be

---

<sup>1</sup> The paper was presented at the conference dedicated to the work of P. Horwich in Pecs in May 2006. I am keeping the original format, and therefore talk about "questions for professor Horwich". I would like to thank professors Horwich and Boros and other participants in Pecs conference, and my students Adrian Briciu and Eva Ferlež for stimulating discussions.

---

<sup>2</sup> Boghossian, P. (2008) *Content and Justification*. Oxford Univ. Press.

*a priori*. His examples encompass truths of logic and sentences like “Spinsters are not married”, “If London is larger than Paris, then Paris is not larger than London” and “If John persuaded Mary to go, then Mary intended to go.” One possible source of their apriority are purely meaning-constituting decisions, another one pragmatic considerations, and the third and most challenging one “is the innate structure of our minds.” But the innate assumptions defining our familiar logical concepts might turn out to be false, defeated by our best science. “Thus we might end up with two logic: an *a priori* one involving our familiar logical concepts, and an *a posteriori* one deploying technical versions of ‘and’, ‘not’, ‘every’, etc.” This is “the tale of two logics” alluded to in his title.

I agree with Horwich’s negative answer, and somewhat disagree with his positive alternative, so I would like to do two things, or two and a half things, to be exact. First, join Horwich in the main struggle with conceptualist-stipulationist of Boghossian’s stripe, and offer an additional argument for the negative answer, addressing Boghossian’s most recent attempts at salvaging his proposal. My criticism concentrates on two points. Point one: Boghossian’s early analogy with reference fixing, his only proof of possibility of a factual truth generated by stipulation; I argue that his analogy is wrong-headed. Point two: his rejection of “substantive” conceptual-analytic truths together with his acceptance of “factual” ones; I am doubtful about there being a relevant difference. The first part of the present paper briefly argues that *the powers of stipulation have been overestimated and that it cannot yield rich substantial a priori knowledge*. Together with Horwich and against Boghossian, it argues that stipulation yields no substantial, factual and *a priori* knowledge and that conceptualism-stipulationism entails conventionalism. If this negative result holds, conceptualists-stipulationists face a dilemma: either retain the view that rich conceptual contents are due to stipulation, giving up their substantial character, or retain the substantial character of conceptual contents, and jeopardize their *a priori* status, and thereby also the *a priori* status of conceptual analysis. The first horn leads to conventionalism and anti-realism about conceptual knowledge: it is just meta-conceptual knowledge of our stipulation, the way logical positivists had proposed. The second horn leads to the view that our concepts are our best attempts at encapsulating our hoped for knowledge about their referents, and beliefs in conceptual truths retain the original character of the knowledge encapsulated.

Next—and this is the second thing I want to do in the paper—I turn to the disagreement within the

family, so to speak, all this in order to raise questions for Horwich concerning his positive account

Finally, the “half” in the two and the half projected things I want to do. I shall briefly conclude that the second horn of the above dilemma points to the correct account of concepts and conceptual knowledge. Many concepts and their attendant concept-analyzing propositions are empirical, among them some that are crucial for our epistemic lives. Being knowable or believable *a priori* is not necessarily a property of conceptual truths, i.e., concept-analyzing propositions. This then is the task. However, in order to accomplish all this I need to summarize the debate, at risk of repeating what is well known. This I do in the second section, first summarizing Boghossian’s strategy for the conceptualist-compatible, and then very briefly Horwich’s criticisms. In the same second section I develop my own argument to the effect that stipulation and similar means are not sufficient for *factual a priori* knowledge. The third section begins by the summary of Horwich on the *a priori* and continues with problems and questions for Horwich.

## 2. Stipulation. Factuality and *A Priori* Knowledge

### 2.1. Boghossian’s compatibilist strategy

The center of the Horwich-Boghossian debate is the issue of compatibility between the stipulated and the factual. As we mentioned, Boghossian has in his (1996) argued that stipulation can yield factual *a priori* knowledge, and has done this by distinguishing stipulationism from conventionalism, and by leaning on Kripkean intuitions about the contingent *a priori*. The motivation of the account and of the debate is epistemological, i. e. answering the question of how a cognizer can acquire *a priori* knowledge. More precisely, how one can reach substantive and factual, mind-independent truths from one’s armchair, without indulging in an contact with relevant facts (logical, mathematical, philosophical)—which seems difficult to explain if not downright inexplicable? Traditionally, the combination of stipulationism and the belief that all *a priori* knowledge is analytic was associated with various non-factualist accounts of candidate conceptual-analytic truths. It went together with a general anti-realist attitude to objects of *a priori* knowledge, that eschews the issue of contact with mind-independent facts, by depicting *a priori* truths as being truths of our making, linguistic or conceptual that we impose upon the reality, where the order of determination going from mind/language to reality. It was opposed to factualist accounts, which insist that it is the mind-independent reality that determines truth value

of logical and mathematical truths (and other conceptual truths as well). Factualist accounts themselves came in two varieties: either the more Platonist ones in which abstract reality forces itself upon the cognizer's mind, or tough empiricist variety in which general structures of empirical reality force themselves upon the cognizer's mind, doing it either in a piecemeal, Millian, or in a holistic, Quinean fashion. Factualism thus seemed firmly opposed to stipulationism.

Still, might the two seemingly opposite poles, the traditional conventionalist-stipulationist one, and the factualist-realistic one, be joined together? Boghossian thinks so: yes, if we give up conventionalism, we can retain stipulationism, since there is an intermediate view combining advantages of stipulation and factual truth.

This compatibilist view enjoining us to combine the conceptualist-stipulationist view of analytic sentences with a factualist account of analytic truths centrally claims the following: We stipulate the truth value of relevant sentences and thereby hit upon important factually and necessarily true analytic propositions.

What I dispute is that it follows from the fact that a given sentence *Q* is being used to implicitly define one of its ingredient terms, that *Q* is a non-factual sentence, not a sentence that “tells us anything that is the case” (reference to Coffa, *The Semantic tradition from Kant to Carnap*, 1991, CUP p.265 summarizing the views of early Wittgenstein and of Carnap). These two claims seem to me to be entirely independent of each other. (Boghossian, 1996:159).

Horwich's “Stipulation, Meaning, and Apriority” opens with a useful summary of the stipulationist factualist view, ascribed to Boghossian and Peacocke, that he wants to criticize. Let ‘#f’ stand for a conjunction of sentences containing the word ‘f’ and perhaps sufficient to determine its meaning. Then the stipulationist argues as follows:

- (1) I am entitled to mean what I want by ‘f’.
- (2) I stipulate (perhaps implicitly) that ‘f’ means what will make ‘#f’ true.
- ∴(3) I am entitled to hold *a priori* that ‘f’ means what will make ‘#f’ true.
- ∴(4) I infer that ‘#f’ is true.
- ∴(5) I infer that #f
- ∴(6) I am entitled to hold *a priori* that #f
- ∴(7) My belief that #f is justified and *a priori*. (Boghossian and Peacocke 2000:151; compare the revised version in 2005: 137).

The parade examples of such semantogenic stipulation is first, fixing the meaning of logical constants by holding true the requisite axioms, and second, the use of abstraction principles in fixing the meaning of number-terms by holding true some version of Hume's Principle (stating that two sets have the same number of elements, iff they can be put into 1-1 correspondence). Here are two quotes from Boghossian on logic:

If ‘f’ is to mean what it does, then *A* has to be valid, for ‘f’ means whatever logical object makes *A* valid ‘f’ means what it does, therefore

*A* is valid. (Boghossian, 1996:159).

and:

“logical constants mean what they do by virtue of their conceptual role: ‘if, then’ means what it does by virtue of participating in some inferences and not in others” (Boghossian and Peacocke 2000 “Knowledge of Logic” 2000:248)

The set of inferences encompasses those “that are meaning-constituting for a thinker”, i.e. the subset responsible for fixing its meaning. Boghossian then proceeds to argue for the compatibility between the factual and the stipulative from the idea of contingent *a priori*.

It is agreed on all sides that contingently true propositions are factual. But Kripkean stipulation yields contingently, and therefore factually true propositions. If Kripkean stipulation yields factually true propositions, then meaning stipulation can also yield factually true propositions, Boghossian invites us to conclude, thus extending the model of “contingent *a priori*” to stipulation of meaning. Remember Kripke's example of the introduction of the term “meter” where the stipulator stipulates

“1m = the length of this stick”

By stipulating the proposition, he comes to know it, indeed in an apparently *a priori* way. But if Kripkean stipulation can yield *a priori* knowledge of factual truth, why not other kinds of stipulation?<sup>3</sup> If the

<sup>3</sup> One more quote: “Knowing all this may not be much of an epistemic achievement, but that isn't the point. The point is that there appears to be no inconsistency whatsoever between claiming that a given sentence serves to implicitly define an ingredient term and claiming that that very sentence expresses something factual” (Boghossian, P. (1996) “Analyticity” in Hale, B and Wright, C. (Eds.), *A Companion to the Philosophy of Language*, Blackwell, Cambridge, p. 350).



suggestion is correct it unifies semantics and epistemology of logic and offers a simple and intuitively appealing starting point for theorizing guarantees factual status to a priori knowable propositions.

In his “Knowledge of Logic” Boghossian reiterates his view about logical constants that they “mean what they do by virtue of their conceptual role: ‘if, then’ means what it does by virtue of participating in some inferences and not in others” (2000:248). The set of inferences encompasses those “that are meaning-constituting for a thinker”, i.e. the sub-set responsible for fixing its meaning. The new point is that thinker needs no explicit justification for making inferences from this set. She does have a warrant, from the meaning constituting character of each of the inferences. That such a warrant will not move a determined skeptic does not deprive it of its warranting powers. That such a warrant is rule-circular is besides the point: first, the rule-circular argument involved is not trivially guaranteed to succeed (I suppose Boghossian thinks of pseudo-constants like “tonk,” born as failures). Second, “by relying on a small number of applications of a particular rule, a successful rule-circular argument delivers the conclusion that that rule is necessarily truth-preserving in all applications” (2000:254). Finally, the thinker is entitled to these applications on the ground of meaning-constitutiveness of the rule.<sup>4</sup>

Before passing to discussion, let me quote a recent paper by Boghossian in which he returns to the topic. This time it is the issue of substantial nature of meaning-constituting propositions. Such propositions notoriously have some very problematic cousins, for instance propositions defining pejorative expressions like Boche. Now, if being German is sufficient for being a Boche, and being a Boche for being cruel, a very unpleasant “conceptual truth” follows. In order to avoid such a bad company, Boghossian enjoins us to avoid commitments to substantive beliefs or theories when stipulating meaning.

<sup>4</sup> In the polemical part of the paper Boghossian disposes of other alternatives: non-factualism about justification is of no help since it does not answer the following: “with what right do I accept epistemic norms that are unconditional? (2000:244). Nor does it give me a basis for insisting that others give up their unconditional norms in favor of mine ones. Equally, acquiescing in accepting default reasonableness, brings relativization to particular thinkers. Second, take someone (I would suggest Hegel), who thought that negating the law of non-contradiction is a basic rule. He is then default reasonable in doing this (2000:239).

Here is the quotation from the concluding section of his “Epistemic Analyticity: A Defense”

The most fundamental relation between grasp of meaning and entitlement occurs when a thinker is entitled to reason in accord with a certain rule R simply by virtue of the fact that R is constitutive of a non-defective concept of his.

This Constitutive model could be extended to account for the thinker’s entitlement to have certain beliefs, provided that those beliefs were similarly constitutive of the possession of a non-defective concept. I think that something like this story will explain how we get to be entitled to believe in the principle of non-contradiction and in the link between truth and meaning that featured in the templates described above.

The substantial nature of the commitments is illustrated by the fact that someone could choose to rationally dispute them. But the concept is set up in such a way as to preclude that sort of rational disagreement.

‘Boche,’ then, is problematic in that possessing it requires a thinker to believe a substantive theory about the world. And it does seem intuitively right to insist that, in so far as is possible, a concept should not be set up in such a way that only those who are prepared to affirm a controversial creed are allowed to possess it. (web, p. 13)

Let me conclude by noting that Boghossian has been developing the first line, on meaning-constitutiveness as a priori justifier, combined in his “Knowledge of Logic”, with occasional remarks on compellingness, i.e. on the alleged fact that “it is not open to us to regard our fundamental logical beliefs as unjustifiable (in Boghossian and Peacocke, 2000:253). For instance, in the same paper Boghossian argues for the warrantedness of logical rules mainly from negative compulsion, i.e. from deeply felt unacceptability and inconceivability. He does not offer any causal or psychological explanation of compulsion, which is, after all, a felt item. Here is the relevant quote:

we cannot accept the claim that we have no warrant whatsoever for the core logical principles. We cannot conceive what such a warrant could consist in...if not in some sort of inference using those very core logical principles... It is not open to us to regard our fundamental logical beliefs as unjustifiable (2000:253).

An opponent might claim that the evidence of these passages is not a knock down one, since the author talks about inconceivabilities concerning matters at the level of epistemic justification, the meta-level, so to speak, and not the first level concerning logical principles themselves. But they suggest a style of justification which very strongly appeals to compellingness, cogency and the like.

## 2.2. *Comrades against compatibilism: the issues of factuality*

I will now very briefly just remind the reader of Horwich's line against the compatibility between the factual and the stipulative, and then propose another line of argument for the same line, by pointing to two additional problems for the compatibilist, both concerning factuality, one old and one new.

First, the all-too-brief summary. Against the appeal to the truth of stipulations, and the alleged need to believe in them, Horwich stresses that it is enough that the stipulator accept '#f' in order to effect the stipulation; there is no need for it to be true. (Williamson has been developing an analogous line on pejoratives in his recent debate with Boghossian (2003); we shall briefly appeal to it in towards the end of this section.) He also argues that there is no guarantee that stipulation will be successful. But Boghossian has already proposed a conditional re-rendering of the original stipulation: "If 'f' has a use, then that use makes '#f' true." Horwich presents him with a dilemma. Either, the conceptualist-stipulationist has not given 'f' a use, and then the conditional is true but vacuous and unsuccessful. Or, if I do give 'f' a use "then I can hold true the conditional only by accepting its consequent." But the consequent just might be false, and I do not know a priori it is not. Horwich's next move is to point out that usually only a part of accepted theory ('#f') is strictly necessary for constitution of meaning. Moreover, the apriorist puts the explanatory cart before the explanatory horse. It is not the case that our holding the object of stipulation to be a priori true results from its being meaning-constitution. The proper order is the reverse one: we hold some propositions to be true a priori. "And this fact is responsible for 'f' meaning what it does." (161). The application of analytic factualist view to logical constants and numbers is then questioned on the following ground. We can understand a constant and disagree on the embedding logical theory, as witnessed by the classicist-intuitionist debate. We can understand what  $2+3$  means and disagree on the ontology of numbers. So, logical theory or ontology of numbers are not constituting for the respective meanings. And

when we turn to the purely meaning-constituting, semantogenic core, the explanation offered by Boghossian gets the order wrong. It is our holding some truths a priori that explains constitution of meaning, and not the other way around. His conclusions are the following. First, stipulation and similar means are not sufficient for factual a priori knowledge. Second, there is a serious question of objectivity of logic, not touched by stipulation-based-accounts. Third, for all we know, logic might be revisable in the light of scientific discoveries. End of the summary.<sup>5</sup>

Let me now try to complement Horwich's criticism concentrating upon the issue of factuality. The initial doubts about factuality being compatible with stipulation derive from general consideration of the order of determination. Distinguish truth-bearers, sentences and propositions from truth-makers, facts,

<sup>5</sup> Bob Hale and Crispin Wright in their paper "Implicit Definition and the *A priori*" offer an original view of meaning-constituting, semantogenic practice. They call "the thesis that at least some important kinds of non-inferential *a priori* knowledge are founded in implicit definition" *the traditional connection* (2000:286). They start by criticizing Horwich's earlier claims against the traditional connection and analytic factualism, namely that mere stipulation that '#f' is true is vulnerable to several risks, which deprive stipulator's belief that #f of its alleged *a priori* justification. The risks are, first, existence problem, i.e. that there might not be a candidate meaning for 'f' around, second, uniqueness problem, that there might be several such candidates, third, possession problem, of how, even granted existence and uniqueness, we can be sure that 'f' actually does possess the right meaning, and fourth, the explanation problem, namely to explain how 'f' comes to possess it. Hale and Wright deny that there is an existence and uniqueness problem. The worry that there might be one "seems to be largely inspired by a combination of platonic imagery and a misconceived analogy between implicit definition and reference fixing" (p.292). The wording seems to me unjust to Horwich. It is not him, but his postulationist opponent that has originally introduced the "misconceived analogy" between meaning constitution and reference fixing (see, Boghossian, P. (1996) "Analyticity" in Hale, B and Wright, C. (Eds.), *A Companion to the Philosophy of Language*, Blackwell, Cambridge, p. 350). The point is, however, independent from the attribution of blame. While fixing the reference ("Let the murderer be called Jack the Ripper") points to further, independent ways of identifying the referent, implicit definition does not presuppose any *independent way* of arriving at the same meaning. The cure Hale and Wright propose is quite radical: their account of explanation via implicit definition leaves room "for the capacity of such explanations to *invent* meanings" (295).

states of affairs, or whatever is involved in factual truth. Then, the most general line about non-factuality of truth allegedly obtaining merely in virtue of relations of concepts is the following:

A: Truth bearer is supposed to fit truth-maker, not the other way around, (Direction of fit consideration).

B) Concepts (meanings) are parts of truth bearers, so conceptual necessity should be determined by truth bearer alone.

C) conceptual (semantic) necessity dictates necessity in the truth-maker.

D): Truth-bearer does not determine truth maker, also not its modal properties. Order of alethic and modal determination therefore

NOC) There is no conceptual (semantic) necessity that is at the same time factual.

Now, stipulation, as described by Boghossian, concerns relation between two kinds of truth-bearers, a purely linguistic one (which is only a potential truth-bearer) and an abstract-propositional conceptual one. The stipulator fixes the truth value of her sentence. This then determines a proposition (that is independently true) and the meaning of 'f'. But the purely linguistic truth bearer 'f' does not determine anything about (non-linguistic) facts in the world. The conceptual one does not either. Therefore, stipulation cannot yield factual truth, and, a fortiori, factual knowledge.

I am not claiming that the argument is conclusive in any way. It just indicates that there is a good ground for incompatibilism. Given this, and the fact that stipulation has traditionally been held to be incompatible with factual truth, the burden of proof rests on the compatibilist. Curiously, the only argument Boghossian offers is a mere analogy with attempts of reference fixing. This is the first point I want to belabor. The next problem I want to raise concerns the meaning and impact of "factual" and its relation to "substantive."

Consider the first point. Either there is the phenomenon of contingent a priori and the stipulationist can take it as a starting point, or there is no such phenomenon. Take the first option first, and begin by assume that there are stipulated contingently true and a priori knowable propositions, like "1m = the length of this stick." Can one extend the lesson to meaning stipulations? Even if we accept, for the sake of argument, that the meter example is a genuine example of contingent a priori, the model cannot be extended to meaning stipulation. Here is why. Boghossian simply ignores the crucial feature of contingent a priori, namely that it is contingent-

factual because it is not meaning giving, but only used for reference fixing. This is the first thing Kripke tells us in his classic explanation:

There is no conflict between that counterfactual statement (the stick might have been longer-NM) and the definition of 'one meter'... because the 'definition', properly interpreted, does not say that the phrase 'one meter' is to be synonymous (even when talking about counterfactual situations), with the phrase 'the length of S at  $t_0$ ', but rather that we have determined the reference of the phrase 'one meter' by stipulating that 'one meter' is to be a rigid designator of the length which is in fact the length of S at  $t_0$ . (1972: 275)

If the meter sentence expresses a contingent truth, it does so because it does not stipulate meaning. This is the whole point. The attempt to extend it to the opposite case does not work, since that what gives it its contingent-factual character is precisely that it is not a stipulation of meaning. Therefore, there is not a shadow of an argument from contingent a priori to factually true meaning stipulations or sentences derived from such stipulations.<sup>6</sup>

Consider now the second sub-issue. Boghossian is just helping himself to the assumption that there is a phenomenon of contingent a priori as required for the analogy. It is very problematic that a conceptualist can do this. If you believe in a rich conceptual a priori the most charitable interpretation of the situation is that there are two propositions expressed by the same sentence: one is necessary a priori ("1m = the length of the stick" in every world in which there is a stick of definite length ostended in the process of stipulation) and the other is contingent a posteriori ("In our world: 1m = the length of the actually ostended stick"). The most sophisticated account of this decomposition has been offered by the 2-dimensionalists. Here is one 2-dimensionalist way to represent it (there are others, but difference in detail is unimportant here):

"1m = the length of this stick" evaluated in

<sup>6</sup> Bob Hale and Crispin Wright noted that there is something dubious in the analogy. In their criticism of Boghossian they warn against his picture of "implicit definition as aimed at hitting on some (unique) pre-existing meaning, as if it were akin to an attempt of reference fixing (Hale, B and Wright, C. "Implicit Definition and the *A priori*" in Boghossian, P., and Peacocke C. (Eds.) (2000), *New Essays on the A priori*, Clarendon Press, Oxford, p. 290).

|                                             |             | w1 (our)<br>stick S<br>_____ | (w2)<br>stick S: _____ | (w3)<br>stick S: _____ |
|---------------------------------------------|-------------|------------------------------|------------------------|------------------------|
| context<br>of utter-<br>ance or<br>thought: | w1<br>(our) | @<br>true                    | False                  | false                  |
|                                             | (w2)        | false                        | @ true                 | false                  |
|                                             | w3)         | false                        | False                  | @ true                 |

And here is Chalmers's comment:

Instances of the contingent *a priori* have a necessary epistemic intension but a contingent subjunctive intension. One's cognitive world-model is constrained by epistemic content, not by subjunctive content; so a contingent subjunctive intension does not indicate a cognitive achievement. (Chalmers, "The Components of Content", p. 32)

In short, Boghossian is in no position to appeal to the analogy with contingent *a priori*. His only argument for factuality is a non-starter. But let me dwell upon the issue, this time turning briefly to the epistemology suggested by the analogy.

The issue of knowledge of stipulated sentences, like the schematic "#f", has been discussed a lot, especially in the relation to warrant transmission. I would like to use Boghossian's attempted parallel with the meter stick example to illustrate the difficulties with his proposal. So, take the meter-stipulator and consider the character and extent of her knowledge. Either she is perceptually acquainted with the stick, and can think of its length in terms of "the length I am acquainted with" ("or "this length" where "this" is accompanied by perceptual awareness of the length"), or she is not. On the latter reading, on which she has had no perceptual acquaintance with the length of the stick, the perceptual acquaintance does not enter stipulator's knowledge; it is as if the stipulation has been made blindly. Then she just knows that whatever the length of the stick (if the stick has a definite length with a definite way of measuring it), this length is one meter because one meter just is this length, whatever it might be. Call it the blind reading.

On another reading (which seems to be Kripke's preferred one) the stipulator is perceptually acquainted with the length of the stick. She knows which length is to become our "one meter." But in that case, it is certainly not stipulated that the stick is to have this visually presented length, since it has it anyway. So what seems to be stipulated is really that this (visually presented length) is to count as one

meter. And this is true in virtue of stipulation, is a man-made fact. Call this the sighted reading.

Compare this with what the stipulator is supposed to know after she has understood the sentence "#f"? Model one: the sighted stipulation. Now this is out of question, since then perception must be replaced by intuiting concepts and Boghossian does not accept this. He does not want the "flash grasping" and offers his account as an alternative to it. So, the sighted reading offers a very poor analogy to knowledge of analytic truths.

If you are not convinced, and think that there can be a "partly" sighted reading (one-eyed Jack, so to speak), less rich than the problematic flash grasping, but sufficiently rich to assure *a priori* knowledge, return to Horwich's existence and uniqueness problem: how does the stipulator then know that there is a unique meaning that makes "#f" *a priori* true? Well, if "#f" is sufficient to fix the meaning, the stipulator might first, come to know *a priori* that this is so, i.e. that "#f" is sufficient to fix the meaning, and second, grasp the meaning through grasping "#f." (E.g. the stipulator might come to know the meaning of "and" by grasping the into and elim. rules, and perhaps constructing the truth table for "and" out of them). But how can the stipulator come to now that "#f" is sufficient, given the problem of the bad company, like "tonk"? Boghossian reminds us that we can come to know *a priori* that the "tonk"-rules are no good. But what we do know *a priori* is that tonk-rules allow derivation of any proposition from any other. And, we know this independently from our stipulation. And, why do we then reject tonk? Presumably, because we independently reject contradiction, and so on. So, the ground of knowledge seems to be external to stipulation.

If this holds, the only remaining option is model two, the blind stipulation. The stipulator knows the meaning of "f" only under the guise of "that meaning (whatever it is) which makes "#f" true. (I stipulate that "Glub is whatever causes cancer", and thereby come to know *a priori* that cancer is caused by glub. Well, I know that glub means whatever causes cancer, period.) This blind reading is the only one even remotely plausible for Boghossian. But it is rife with difficulties. One problem is precisely the lack of acquaintance (*de re* knowledge): it is clear that the stipulator does not know which length the length of one meter is. She does not understand the "import" of her stipulation. This raises the problem of believing ill-understood sentences. One way for the stipulator to go is to tell oneself: "I will hold this sentence true come what may, or at least behave as I did, and then I shall see what it means. If I am lucky." (A bit like Pascal's advice on religious be-

lief: go on behaving as if you believed and the belief will come) The alternative is to propose two stages: first, holding true without knowing which proposition(s) the sentence expresses, then "grasping" the right proposition (e.g. introducing to a class the Schaefer stroke by intro and elimination rules). But this reintroduces the flash grasping at the second stage. (Perhaps mere familiarizing with the ill-understood sentence is sufficient, the stipulationist might claim in despair. But it is desperate indeed.) Another way to go is to follow Horwich and acquiesce in conditional character of the sentence believed: if "f" has a meaning then #f (And Horwich adds that if the antecedent is false no knowledge, if true, you have to prove that the consequent is also true). But then, do we really want to claim that our knowledge of logic (and possibly) mathematics is of this sort?

Let me now turn to the consequences of these problems. Our alleged a priori knowledge is supposed to guide us in all our cognitive efforts. But how do merely stipulated-true contents perform the guiding? Do we use them as fallible guides, to be corrected by the new discoveries or not? Call this the issue of guidance (and note that it is connected to the problem explaining how a stipulated-true sentences can express factual truth).

The option of fallible guidance would suggest initial free choice of stipulation followed by subsequent testing. The option seems congenial to Boghossian since his theory starts from "stipulating as being true." However, it reopens the Horwichian gap between holding true and being true. It is psychologically implausible. It is contrary to Boghossian himself: his justification for logic is "we can't think otherwise." (And, of course, it raises the problem of cheapness and bad company. Stipulations are cheap. (Besides TONK and BOSCH think of "maloeba", stipulated to stand for "male amoeba.") Mere choice leaves open the issue of why these and not others. A possible answer: the undesirable ones are weeded out at the testing stage. But how? Well, MALOEBA through empirical knowledge, TONK because it leads to contradiction, BOSCH because "we don't use such concepts." In summary, here is a dilemma: either (A) external guidance or (B) arbitrariness. If (B) cheapness threatens, if (A) then our alleged a priori knowledge is not explained by stipulation, but the other way around as Horwich has been first to note. We shall return to this point presently.

Alternatively, it is the actual holding true of a stipulated sentence S (or perhaps of a system of stipulations SS): the holding true makes up our understanding of the domain (e.g. logic, or arithmetic, or kinship relations). The problems are again arbitrary,

lack of explanation for the truth, and the threat of change: some system SS1 being replaced by SS2 under the impact of empirical evidence plus theorizing.

To summarize. Here is a dilemma: either we have (A) external guidance or (B) arbitrariness. If (B) cheapness threatens, if (A) then our alleged a priori knowledge is not explained by stipulation, but the other way around (Horwich). What about primitive compulsion as an alternative. It would solve the justification problem by "ought implies can" and offers some guidance, but opens the problem of being primitively compelled to hold true sentences (and to hold correct rules) one only partly understands. Worse, it offers a competing explanation: it is not stipulation per se but its irresistible force (Plus, it is hard to see how a set of sentences by being merely held true does not yield analytic propositions, but by being held true irresistibly does). There is a possibility: retreat to anthropocentrism: well, these are our concepts. Irresistibility is the mark of conceptual character, and irresistibility-for-us is the mark of ownership, of being "our own" concepts. But it is not opened to Boghossian, since it would open the door to epistemic relativism he is justly combating.

Let me now turn to a relatively new factuality problem, raised by Boghossian's recent attempt to solve the problem of cheapness and bad company. We have quoted above his diagnosis of what is wrong with problematic pejoratives: "'Boche,' then, is problematic in that possessing it requires a thinker to believe a substantive theory about the world. (web: 13). If this is the solution, the problem is: given that a priori statements should be factual, but should not be substantive, what is the difference? If there is no difference, there is no solution to the problem of bad company: every factual proposition will involve "substantive" concepts and theories, which are problematic. Boghossian needs a meaning of "substantive" that is distinctive from "factual." Call this substantive vs. factual problem.

One quick and unserious proposal is that "substantive" just means "problematic" (in respect to truth). One could take Boghossian's sentence immediately following the quoted one to express this proposal:

And it does seem intuitively right to insist that, in so far as is possible, a concept should not be set up in such a way that only those who are prepared to affirm a controversial creed are allowed to possess it. (web, p. 13)

Of course, this would beg all the questions involved. “Boche” is controversial for us, since we think that the theory it presupposes is just false about Germans, while the Germans-haters find the very same theory obviously true. That this is not limited to pejoratives, and can be raised for quite ordinary concepts, is shown by the philosophical debate recently sparked by the issue of homosexual marriage, of whether concept MARRIAGE entails that the partners to marriage are of different sexes, and prominently featuring Robert Stainton, defending the truth, and Adele Mercier the falsity of “Marriage can exist only between persons of different sex.” If avoiding such controversy were the only guideline, the stipulator would not be allowed to stipulate any sentence “#f” to be true, unless there is a prior consensus that “#f is uncontroversially true!”. So, it cannot be Boghossian’s intended meaning.

On the more serious side, I can think of four broad ways for an a priori claim to be substantive. First, epistemically, in virtue of novelty and fruitfulness of beliefs it expresses. Epistemic substance can probably be accounted for most easily in terms of the limitations of our cognitive grasp, which make even logically trivial propositions difficult and exciting for us. This doesn’t help Boghossian. If substantive character is what is bad for candidate stipulative apriority, as he claims in the paper quoted, then fruitful logical insights are in the same boat with racist theories about Boches. If “A Boche is usually cruel” is out of question for the stipulative account, so are epistemically fruitful logical truths.

Second, cognitively, by being empirical. Some concepts and their attendant concept analyzing propositions are empirical, among them some that are crucial for our epistemic lives. However, any empirical concept will have concept-analyzing propositions that state empirical truths (“A whale is a material object”). (Shall we claim that being knowable or believable a priori is not necessarily a property of conceptual truths, i.e., concept-analyzing propositions?)

Third, semantically: once their meaning is fixed, a priori claims are made true by facts independent of them, and of the human mind in general. And this is what makes them substantive. Now, this reading would be a catastrophe for Boghossian’s account, since it equates substantive character with factuality. All factual truths are substantive on this reading, so no logical or conceptual truths, no analytic truths known in virtue of meaning, can be reached by stipulation.

Finally, an a priori claim can be substantive metaphysically: some of the claims concern rich and metaphysically fundamental items, such as, for in-

stance, sets and numbers, or humanly particularly interesting items such as our minds (call this subspecies “anthropologically substantial”, if you want a special name for it). Semantic and metaphysical substance and richness are very difficult to account for in the case of a priori knowledge and it would be fine if Boghossian had an account. But “Boche” is certainly not metaphysically substantive in this sense, so by “substantive” in the text quoted he can’t mean “metaphysically substantive.”

Incredible as it may seem, the only live option is that “substantive” just means “controversial”, and this is a clear non-starter. So, either Boghossian accepts the meaning “controversial”, and deprives himself of the solution to the problem of overgeneralization i.e. of bad company, or, he has to equate substantive character with factuality, and give up his most important claim, i.e. that analytic truths can be factual. I would recommend the latter.

### 3. The Tale of the Two Logics

#### 3.1. Horwich on the a priori : “Render unto Quine what is Quine’s and unto Wittgenstein what is Wittgenstein’s.”

We now turn to Horwich’s own view as stated in his apriority paper from the very beginning of the millennium (2001). In its very dense final sections he offers his reinterpretation of Quine’s Revisability Argument, and his own proposal for preserving and explaining the apriority of some typical concept-analyzing propositions, like the following: “Spinners are not married.” “If London is larger than Paris, then Paris is not larger than London.” “If John persuaded Mary to go, then Mary intended to go.” “The belief that snow is white is true if snow is white.” “The chess bishop moves diagonally” (2000:166, also in *Meaning*, fn. 16). He also adds two propositions that are not clearly of this sort “One ought to have some concern for the welfare of other people.” and “Nothing can be both red and green (all over)” (Ibid.). I am aware that the proposal is merely a sketch, even a sketch taken out in the recent (2005) descendent of the original paper, but it is a sketch that deserves discussion.

From Quine, Horwich retains the idea that every scientific belief is revisable in light of experience, i.e. the restricted version of the general principle of revisability of every belief. He also retains a restricted version of Quine’s epistemological conclusion: none of our scientific beliefs is a priori. The reason is well known: according to the web-of-belief model, which Horwich accepts for scientific beliefs, the set of such beliefs “evolves through time subject to constraints of empirical adequacy, overall simplicity, and conservatism. This model is highly

plausible both in light of the history of science and in light of our conception of rationality...It implies that any scientific hypothesis—including a logical principle—might be abandoned on empirical grounds; for doing so might bring about an overall gain in simplicity” (2000: 166).

The recipe for retaining apriority is then disarmingly simple: some important candidates for a priori beliefs are simply not scientific beliefs, for instance the ones listed at the beginning of this section. “If John persuaded Mary to go, then Mary intended to go” expresses our linguistic intuition, not the result of linguistic science. Sentences and claims expressing such beliefs come from non-scientific “assertoric practices in which we engage” and Horwich lists ordinary life, legal contexts, and game-playing as examples. Presumably, the claim that one ought to have some concern for the welfare of other people figures in our moral assertoric practice. And, “Look, if London is larger than Paris, then Paris is not larger than London” might have come up in a very non-scientific discussion in a pub in which the speaker’s interlocutor has been stupidly first claiming its antecedent, and then the negation of its consequent. The argument is disarmingly simple. First, given that there are many assertoric practices besides science, “not everything we assert is a scientific hypothesis.” And then comes a surprising second step: “Therefore, not everything we believe is subject to the pressures of empirical adequacy and global simplicity.” Horwich seems to assume that empirical adequacy and global simplicity are relevant and cogent only for scientific hypotheses, that only the credibility of such beliefs “is hostage to empirical considerations” and this assumption, call it Restriction Assumption does the main work. The rest is easy. If the Restriction Assumption holds, “Quine’s argument provides no reason to doubt the existence of a priori knowledge outside the domain of science.” (2000: 206). Ordinary non-scientific assertoric practices will thus for Horwich save the day; it is as if a division of competences between Quine and Wittgenstein saves a priori. “Render unto Quine what is Quine’s and unto Wittgenstein what is Wittgenstein’s.” The former gets science and scientific hypotheses, the later, in a finely naturalized version, rules over ordinary assertoric practices, and peace is restored.

And here come my first two questions for Professor Horwich. Question one: do you really hold the Restriction Assumption? If not, then the division of competences between the scientific and the non-scientific is of a very limited nature and does not save the a priori. If yes, what is the rationale for it? To start with the obvious, take the non-scientific

assertoric practice of discussing witches, trying to identify some in the neighborhood, and accusing them of witchcraft. The main problem with such a practice is not its political incorrectness, but its empirical inadequacy. There is no empirical evidence for successful witchcraft. And this has killed the assertoric practice (thank God!), as Quineans would have predicted. More mundane practices, like discussing the size of towns, say London and Paris is, in the last instance, very much beholden to empirical adequacy, otherwise, I would be happy to brag that my home town, Rijeka, is larger than both, and you would have no way to criticize me. And the hypothesis that Rijeka is in fact smaller is vastly simpler than the alternative hypothesis that every time the census is made, there are sophisticated mistakes committed, that give the false picture of its relative size. And this are the main reasons why no one takes the later hypothesis seriously.

It is a bit ironical that the main philosophical support from these consideration comes from the view that simplicity and empirical adequacy are guides to truth, and that truth is a central goal of assertion. And we have it from Horwich that he finds truth highly desirable (and I have been impressed by his brilliant attempt to show that this is compatible with deflationism many years ago). For instance, here is his recent statement:

I think we should acknowledge that true belief has a non-instrumental value—a value for its own sake. In this respect, truth is a desideratum regardless of which proposition is in question—whereas it is instrumentally desirable only in certain cases. However, although the value of truth for its own sake is distinct from the instrumental value of true belief, there is nonetheless an important relationship between these qualities. It is presumably because most truths are useful in practical inference—and not merely to those individuals who discover those truths, but also to all the rest of us to whom they are communicated—that our society, simplifying for the sake of effectiveness, inculcates a general concern for truth for its own sake (2006: 351).

Now, the reason why scientists cherish simplicity and empirical adequacy is their role in guiding us to truth. Horwich himself says so:

But standard *epistemological* considerations—namely our conviction that simple theories are more likely to be true—will perfectly well suffice to explain our adoption of simple hypotheses (2000: 176).

Setting aside normative discourse, it is hard to see why other assertoric practices, having truth as their goal, would not value the same qualities. Indeed, discussing the size of towns, or witches, or most other descriptive matters, seems to be beholden to exactly the same qualities. If this hold, we come to the second question: are there any good reasons to uphold the Restriction Assumption? In other words, if we render unto Quine what is Quine's, there will be little left to render unto Wittgenstein.

Let me add a third question formulated by my student, Adrian Briciu, that goes into the opposite direction. In a later passage Horwich admits that meaning-constituting considerations are valid in other assertoric practice. Adrian asks: Why not in science? And what happens if they do? If you render unto Wittgenstein what is Wittgenstein's there will be little to render to Quine.

### 3.2. Genetic Origins of Apriority

This brings us to the crucial section of Horwich's paper, the one discussing semantic, pragmatic, and genetic origins of apriority. Semantic ones are brief: "purely meaning-constituting commitments can be nothing stronger than Carnap conditionals" (2001: 167). And I agree. The next group is pragmatic, and Horwich does not find it very promising, and here I again agree with him. The pragmatic rationale holds "if we are concerned with a domain of discourse whose purpose is not to construct the simplest overall account of what we observe. In any such domain it should be possible to find beliefs whose motivation is wholly pragmatic" (Ibid.). I wonder what these domains are, however. If you think morality is one of them, you will either have to go non-cognitivist about morality, or you will have to insulate morality from the rest of real world. If you are a non-cognitivist, the apriority of moral beliefs comes for nothing, but is also not much worth. If you are a cognitivist, but you are not interested in linking morality to natural world, the way some followers of McDowell, for instance Dancy, are not, than it is all right. I wonder, however, if Horwich would approve of such an insulation strategy.

We now pass his main point, the combination of nativism with the "ought implies can" strategy. It features a simple and elegant argument, which I would reconstruct in the following way<sup>7</sup>:

<sup>7</sup> Here is the relevant quote:

A further potential source of *a priori* knowledge is the innate structure of our minds. Suppose that each human being is born with, and stuck with, a simple language of thought (i.e. mentalese) containing, amongst other things,

1. Suppose that each human being is born with, and stuck with, a simple language of thought (i.e. mentalese) containing, amongst other things, certain symbols whose intrinsic nature is such that the principles of classical logic are obeyed. (A scientifically plausible assumption, that can be extended to arithmetic, geometry, and innate concepts in many areas)
2. This commits humans to classical logic as their natural reasoning system underpinning their rationality.
3. This commitment derives neither from experience nor is revisable in light of it.
4. This commitment is *inevitable*, humans *cannot get rid of it*. (Inevitability Claim)
5. Norms are constrained by possibilities: ought implies can, so if one cannot believe *p* one is not rationally required to believe it. (Impossibility implies lack of obligation).
6. Therefore, (under assumption that it is the case that certain beliefs are innate and irremovable) we can understand why they would be amongst those beliefs that we treat as unconditionally rational.
7. Therefore, then we have an *a priori* commitment to classical logic.
8. Therefore, the innate structure of our minds is a potential source of *a priori* knowledge.

Horwich notes that there is the common core to science and ordinary assertoric practice, namely logic, arithmetic, and geometry. He renders unto Quine what he thinks belongs to him: the logic, arithmetic, and geometry of science are a posteriori. However, in non-scientific assertoric practices "we will continue to rely on classical logic, on standard

certain symbols whose intrinsic nature is such that the principles of classical logic are obeyed. In that case we would have an *a priori* commitment to classical logic. For this commitment would neither derive from experience nor be revisable in light of it. Of course, the *a priori justification* of a belief does not consist merely in how it is actually engendered, but is a matter of the conditions in which it *ought* to be adopted and maintained. To say that a proposition is known *a priori* is to make a normative claim about it, not a descriptive claim. However, norms are constrained by possibilities: ought implies can. In particular, if it is the case that certain beliefs are innate and irremovable, then we can understand why they would be amongst those beliefs that we treat as unconditionally rational (2000: 168).



arithmetic, and on Euclidean geometry.” “They are a priori,” he adds. The result: there are two logics, an a priori one and an a posteriori one. Presumably also two epistemically disparate kinds of arithmetic, one a priori and the other a posteriori.

I very much agree with Horwich about premise 1. I also note that he needs a much wider assumption about innateness. First, he needs it for arithmetic and Euclidean geometry, and I am happy to grant it to him. But remember his list of candidate a priori propositions. “If London is larger than Paris, then Paris is not larger than London” is not merely pragmatically a priori. And it is unthinkable for us that it does not hold. So, it and its likes should go into the innate-inevitable set. (I am not sure about the rest, the John-persuaded -Mary sentence, the instance of T-schema, the chess bishop sentence, and in particular the “Nothing can be both red and green” sentence. How are we supposed to explain their claim to apriority? Horwich owes us an answer.)

The crucial problem is posed by premise 4, the Inevitability Claim. Innate beliefs seem to be inevitable in a weak sense: we find them natural, we normally don’t question them and we find it very difficult to doubt them. Naïve physics and Euclidean geometry are very good examples. But also some reasoning heuristics seem compelling to most people. Unfortunately for the argument, this weak sense is not sufficient. Premise 5, to the effect that impossibility implies lack of obligation (converse of the Ought –implies-can principle) and takes “cannot” and “inevitable” in a much stronger sense of “being humanly impossible.” However, Euclidean geometry is not any more firmly and literally believed by educated people. A good high-school student can get rid of her innate naïve physics with the help of a good teacher. Scientific revision of ordinary beliefs usually trickles down into commonsense, this is what science classes and “Scientific American” are here for. In short, if humans can revise their innate beliefs when it comes to science, then they can revise them simpliciter. If Quine can demand so much of the scientist, he can demand something similar from those acquainted with the results of science. The bipartite picture, separating science from all other assertoric practices is false.

Science driven reflective revisability gives some assurance against blind epistemic luck. It is a matter of such luck that some of our innate beliefs are true and reasoning strategies correct. But we should not rely on luck if we don’t have to. And the fact that we indeed can revise our native frames of mind is a great asset: it reduces the power of blind luck. Just consider the problem posed by innate “garbage,” the realistic possibility of their being innate material that

is cognitively pathological. Some of us believe that racist or religious beliefs can be a case in point, others that some possibly innate reasoning heuristics are one. Again, such “garbage” might be weakly inevitable, in the sense of requiring a lot of effort even to get rid of in some people for some time, but being ineradicable in general, for most people and forever. The justification bestowed by such circumstances is not what we would like to have as a rational justification for a priori beliefs.

Science driven reflective revisability therefore has an enormous normative significance. So, if humans can revise their innate beliefs when it comes to science, then they can revise them simpliciter, and if they can revise them, they should do so.

All this brings us back to the common starting point, the truth-goal. Since the relevant practices are assertoric, they produce truth-evaluable, true or false items. Truth is an important goal in production of most such items. Therefore, the relevant practices have truth as an important goal, so truth is an important common goal of science and of the relevant assertoric practices. Now, if a “theory” that has been innate is revised in and by science in search of truth, it is reasonable to inform other practices about this rejection. Further, it seems that innate beliefs are de facto revisable. And it is hard to live with double epistemic standards. So, if humans can revise their innate beliefs when it comes to science, then the revision should influence all the assertoric practices in which they play a role. Then, the beliefs would stop being a priori. The same would be reasonable in the case of logic, should the need arise.

Conclusion: since the bipartite picture is incorrect, we need a better (equally internalist) justification for our armchair practices and their ambition to produce a priori beliefs. Before accepting it, let me put a question to Professor Horwich: how would you block the main line of criticism, namely the claim that if humans can revise their innate beliefs when it comes to science, then they can revise them simpliciter, so they ought to do so?

Suppose that we have accepted the conclusion: innate beliefs are difficult to revise, but can be revised if the need arises. Second, science is much more into the kind of business that can lead to revision, ordinary assertoric practices are not. So, there is a duality, but much weaker than needed for Horwich’s strategy. We also retain Horwich’s partly internalist idea of justification with which we have been implicitly working until now. What are the alternatives for accounting for candidate a priori beliefs? We might then assume that two sorts of justification are better than two mutually isolated (kinds of) belief systems. So our candidate a priori

beliefs could enjoy *prima facie* spontaneous justification from their obviousness and immediate compellingness, i.e. from “can’t do otherwise” considerations, featuring the weak and ordinary meaning of “can’t.” In this sense, I can’t imagine logic and arithmetic to go wrong. I can’t imagine that spinsters are married (though I can think of two “classical” spinsters united in a marriage, namely a homosexual one). I can’t conceive of state of affairs where London is larger than Paris, and Paris is also larger than London. (Unless “larger” means “has more inhabitants” in the antecedent, and “has a bigger area” in the consequent of our sentence). I believe that if John persuaded Mary to go, then Mary intended to go. (Unless Mary only thought she formed an intention, but deceived herself. Some people would still say John was successful in persuasion, only Mary spoilt things by self-deception.) But my inability to conceive otherwise is still hostage to what one day we can discover. So, my relevant beliefs are weakly *a priori*, not immune to empirical revision. Where does the threat of aposteriority come in? Maybe in a second order, reflective coherentist justification which has an *a posteriori* component, namely that science has until now lived happily with classical logic and arithmetic, and that there is no known counterinstance of the non-symmetry of “larger than” and so on. I would be quite happy with such an alternative, and I think it would preserve much that is valuable in Horwich’s original proposal.

#### 4. Conclusion

Let me summarize the main points and questions for Horwich. I have tried to join him against Boghossian arguing that stipulation yields no substantial, factual and *a priori* knowledge and that stipulationism entails conventionalism. I adduced two kinds of considerations. The first kind of consideration starts from the fact that the only candidate evidence for compatibility is that reference-fixing stipulations might yield contingent-factual *a priori* knowledge. It is not usable for the purpose. First, even if such stipulations yield contingent-factual knowledge, they do it precisely because they are not meaning stipulations, their results not holding in all possible worlds. The analogy is therefore misleading and wrong-headed. Second, it is not clear that Boghossian is in the position to agree with Kripke even on contingent *a priori*. His conceptualism leaves no place for it and is much more congenial, if not forced, to the alternative account according to which each of the sentences taken by Kripke to express some one contingent *a priori* proposition in

fact expresses two distinct propositions, none of which is both contingent and *a priori*. It is incredible that Boghossian has built his factuality claim on such a meager, or even non-existing evidence.

The second consideration points to his recent work in which he warns us not to use substantive claims in stipulations. Some meanings of “substantive” bring it close to “factual,” others turn out to be irrelevant (e.g. “problematic” or “empirical” just beg the question).

If this negative result holds, Horwich is certainly right in his criticisms. Conceptualists-stipulationists then face a dilemma: either retain the view that rich conceptual contents are due to stipulation, giving up their substantial character, or retain the substantial character of conceptual contents, and jeopardize their *a priori* status, and thereby also the *a priori* status of conceptual analysis. The first horn leads to conventionalism and anti-realism about conceptual knowledge: it is just meta-conceptual knowledge of our stipulation, the way logical positivists had proposed. The second horn leads to the view that our concepts are our best attempts at encapsulating our alleged knowledge about their referents, and beliefs in conceptual truths retain the original character of the knowledge encapsulated. The final suggestion is that this second horn of the dilemma points to the correct account of concepts and conceptual knowledge.

Turning to Horwich’s own account of the *a priori*, I have first raised the issue of separability of scientific investigation and non-scientific assertoric practices in regard to norms of adequacy. I questioned Horwich’s Restriction Assumption according to which only the former are answerable to the standards of empirical adequacy and overall simplicity (which threaten *apriority* in the case of science). My questions to Horwich are: Do you really hold the Restriction Assumption? Given that most assertoric practices, having truth as their goal, value empirical adequacy and overall simplicity, are there any good reasons to uphold the Restriction Assumption? What if meaning constitutive considerations persist into science? (A. Briciu). Next, I have concentrated upon his argument that inability to think otherwise might guarantee *apriority*, and stressed the phenomenon of science driven reflective revisability of possibly innate beliefs, like those in Euclidean geometry. My last question is therefore: if humans can revise their innate beliefs when it comes to science, then they can revise them simpliciter? And if they can, doesn’t your defense of the *a priori* fail?

## References

- Boghossian, P. (1996) "Analyticity" in Hale, B and Wright, C. (Eds.), *A Companion to the Philosophy of Language*, Blackwell
- Boghossian, P., and Peacocke C. (Eds.) (2000), *New Essays on the A Priori*, Clarendon Press, Oxford.
- Boghossian, P. "Knowledge of Logic", 2000 in Boghossian and Peacocke (2000).
- Boghossian, P. 2001. "How Are Objective Epistemic Reasons Possible?" *Philosophical Studies*, 106:1–40. reprinted in Boghossian, (2008)
- Boghossian, Paul (2003). "Blind Reasoning", in *Proceeding of the Aristotelian Society Supplementary Volume*, 77: 225–48.
- Boghossian P., "Epistemic Analyticity: A Defense" in his (2008).
- Boghossian, P. (2008), *Content and Justification*, Oxford University Press.
- Chalmers, "The Components of Content", available on his website.
- Hale, B and Wright, C. (2000) "Implicit Definition and the A priori" in Boghossian, P., and Peacocke C. (Eds.).
- Horwich, Paul. (1997) *Implicit Definition, Analytic Truth, and A Priori Knowledge*, *Nous* 31: 423–440.
- Horwich P. (2000), "Stipulation, Meaning, and Apriority" in Boghossian, P., and Peacocke C. (Eds.).
- Horwich, P. (2005), *Reflections on Meaning*, Clarendon Press, Oxford.
- Horwich, P. (2006), "The Value of Truth", *Nous*, v. XL; No.2, , 347–359
- Williamson, T. (2003),, "Blind Reasoning", *Proceeding of the Aristotelian Society Supplementary Volume*, 77, 249–291.
- Wright, Crispin. 2001. "On basic logical knowledge", *Philosophical Studies*, 106:41–85.

## MEANING AND TRUTH

*Joško Žanić*

Bartolići 21, HR-10000 Zagreb  
josko\_zanic@yahoo.com

### **Abstract:**

The paper is divided into three parts. In the first part, four types of meaning theories are presented (namely, formal or truth-conditional theories, use theories, structuralism and the cognitivist and conceptualist approaches) with respect to how crucial they consider the notion of truth to be in the explication of meaning. In the second part, Conceptual Semantics, as the theory that understands the inquiry into meaning as an investigation of our conceptual structure, and doesn't use truth as a key notion, is advocated as a very promising approach. In the final part of the paper, a construal of truth as a matter of multiple fit is proposed as one that both sits well with the framework of Conceptual Semantics and also sheds some light on the ways in which our cognitive system operates with the notion of truth.

**Key words:** meaning, conceptual semantics, realism, anti-realism.

### **1. Meaning theories and truth**

Many contemporary semanticists and philosophers of language believe that meaning is essentially tied to truth – that the concept of truth will have to play a central role in any serious account of meaning. The most common way in which this idea is cashed out is by developing a *truth-conditional semantics*: a theory of meaning which assigns meaning to every sentence of a language by specifying its truth conditions, which are built up compositionally from the meanings (denotations) of the linguistic items which occur as parts of the sentence. *Locus classicus* of this approach is, of course, Davidson's (2001/1984), which draws crucially on the work of Tarski (cf. his 2001/1944). Davidson's basic idea was that to give the truth-conditions of a sentence just *is* to give its meaning, and many theorists today believe that a theory of meaning which doesn't specify truth-conditions doesn't deserve to be called a theory of meaning. An important step in the development of this program happened when Davidson's truth-conditional approach was married to Chomskyan syntax. Since formal rigour was always an important desideratum of this approach, and since sentences of natural language don't wear their logical

form on their sleeve, what was needed was a way of translating the sentences of natural language into their semantic representations that resembled an algorithm as close as possible. The semantic representations, given in the meta-language and couched in the vocabulary of formal logic, were a way to turn sentences of natural language into something that was, due to its artificiality, sufficiently clear and precise for the truth-conditions to follow immediately from it in a familiar way (by employing set-theoretic devices). What wasn't clear, however, was why a sentence should be translated into this formal language one way rather than another, i.e. how to effect this translation in a non-ad hoc way. That's where Generative Grammar came in. Chomsky and his followers had developed ways of systematically representing the syntactic structure of sentences, and this apparatus came to be used by semanticists as a way-station en route to representing the truth-conditions of a sentence (cf. Chierchia & McConnell-Ginet 2000 for a contemporary introduction). The notion of *logical form* plays an important role in this theoretical framework, but it is ambiguous between two different readings. On the one hand, logical form, abbreviated as "LF", is a level of *syntactic* representation. The derivationally derived surface structure of a sentence serves as input to an interpretive component of the grammar<sup>1</sup> which, via operations of covert movement,<sup>2</sup> yields LF as that level of syntactic representation where relations (such as scope) relevant to semantic interpretation are overtly displayed. LF is then translated into *lf*, the other kind of logical form: namely, the real *semantic representation* of a sentence. So, a sentence such as "Every man loves a woman" will first receive a tree-like LF representation (or, rather, two of them, reflecting the two different readings that are made possible by the scope-ambiguity of the sentence) which will then be

<sup>1</sup> Actually, there are two of these; the other yields phonological form.

<sup>2</sup> As opposed to overt movement which serves to generate S(surface)-structure.

turned into a formula (again, two of them) in the familiar vocabulary of formal logic, for which truth-conditions can be easily given (with respect to a model). This, then, is the way in which semantics often proceeds nowadays.

It is not, however, the *only* way in which it can proceed. Within analytic philosophy itself there is another tradition, a family of theories that usually go under the name of *use* theories, as opposed to *formal* theories presented in the preceding paragraph. Outside of the analytic tradition, there have evolved, in the course of the twentieth century, at least two other very influential families of theories of meaning: one of them goes under the name *structuralism*, and was developed mainly in France; the other is the family of *cognitivist and conceptualist* approaches, represented importantly by George Lakoff and Ray Jackendoff. In none of these other three kinds of approach is truth a central concept used to account for meaning, nor are truth-conditions what is primarily aimed at as a way to specify meaning. I intend now to say something about how these kinds of approach try to account for meaning.

Use theories (also known as *pragmatics*) draw on the work of the later Wittgenstein, Austin, Searle, Grice and Strawson. Following Wittgenstein's remarks in his *Philosophical Investigations* about meaning as use, philosophers of language began to explore the ways in which language is used in conversational situations and to investigate how meaning emerges as the result of the circumstances of this use. Speech-act theorists Austin and Searle set out to classify the different kinds of speech acts that we can perform by means of language, to elucidate the notion of illocutionary force and to describe the conventions that govern the speech acts we perform. Whereas, in these theories, the notion of *convention* played the crucial role, theories more influenced by the work of Grice focused on the notions of *intention* and *inference*. Gricean pragmatics wasn't concerned so much with the systems of conventions or rules that underlie our "doing things with words," but rather with how the hearer inferentially works out what the speaker intended to convey. More recent work in pragmatics continued this Gricean line, but changed the focus once more, by pointing out that it is not only what the speaker implies, but also what he explicitly states that involves pragmatic processes and inference on the part of the hearer. The crucial work that initiated this kind of approach was Sperber & Wilson (1995/1986), and one of the most influential theorists within this kind of framework today is Recanati (cf. his 2004). On this approach, the meaning conveyed by a speaker in a given situation goes beyond what is linguistically encoded in two ways:

first, the hearer has to work out the *explicature*, what is literally said by the speaker, but is not fully specified by the linguistic items used and must be recovered from the context. The hearer arrives at the explicature by way of *primary* pragmatic processes (cf. Recanati 2004), which include the linguistically mandated, bottom-up ones (reference assignment, disambiguation) and the more "free", top-down ones (free enrichment, loosening, transfer). After recovering the explicature, the hearer can work out the possible *implicature* by way of *secondary* pragmatic processes (which are, according to Recanati, but not Sperber & Wilson, importantly different from the primary ones).<sup>3</sup> Now, the explicature, or what is said, is considered to be a proposition, so something that can be true or false; it therefore *is* open to truth-conditional treatment. However, the brief overview given above indicates the extent to which the notion of truth and the problem of truth-conditions have been removed, within this kind of framework, from the focus of inquiry into meaning. Now, pragmatic issues can be integrated into a broadly conceived truth-conditional semantics (cf. Chierchia & McConnell-Ginet 2000: ch. 4); but they can also be integrated into a different kind of approach, which will be discussed a bit later. In any case, the approach to meaning offered by pragmatics shows one way to investigate meaning without being predominantly concerned with truth and truth-conditions.

Yet another way, and one that is as far removed from the concerns of pragmatics as possible, is structuralism (initiated by the work of de Saussure; a central work of structuralist semantics is Greimas 1986/1966). Whereas pragmatics is concerned with the dynamics of conversational exchange, structuralism wishes to discern the *system* behind the *process*: it is interested in the system that is called language and that, according to the proponents of this view, underlies, as an abstract and static entity, linguistic communication within a speech-community. In their conversational exchanges speakers will manifest this system, in a fragmentary way, and thereby provide data for the theorist; the theorist of the structuralist persuasion will, however, disregard all aspects of communication other than those that lead to a reconstruction of an abstract system of linguistic items that stand in the relations of difference and opposition to each other. Since the key concern of the structuralist approach is to delineate the linguistic

<sup>3</sup> Basically, Sperber & Wilson see both the processes that yield explicature and those that yield implicature as unconscious inference, whereas for Recanati the primary pragmatic processes are unconscious but not so much inferential, and the secondary pragmatic processes are inferential but much more conscious.

system as a self-contained entity, where the linguistic units are constituted by their relations to one another, it is not surprising that structuralism doesn't care about truth or truth-conditions, which normally involve the bringing of something else into the picture in addition to the linguistic units alone. Structuralism will care more about some kind of coherence or internal logic of a text. Although structuralism is not a productive research program any more, it has provided many valuable insights, particularly into the meaning of literary discourse, for which its particular methodology seems best suited; and it illustrates one more way in which semantics needn't make truth and truth-conditions its central concern.

Finally, there is the cognitivist approach, the originator of which is George Lakoff (cf. his 1980, 1987). This approach sees meaning as located in our conceptual system. What is of primary interest in this kind of framework are the operations of our conceptual system which are expressed in language, but are not to be identified with it. A version of this approach, called Conceptual Semantics, is represented by Ray Jackendoff (his most recent presentation of this approach in book form is his 2002) and Steven Pinker (2007), and I believe it to be the most promising semantic theory on the market. It differs from Lakoff's theory in several respects, most importantly on the issue of the importance of metaphor in thought (it denies its centrality, the affirmation of which is the hallmark of the Lakoffian approach) and on the issue of combinatoriality of language and thought and the related issue of semantic primitives (it subscribes to a search for these, which is a lost cause in Lakoff's view). Conceptual Semantics also accepts much of the general Chomskyan framework, which Lakoff rejects – in particular, it sees the attempt to characterize the innate Universal Grammar as an important and viable objective of linguistics. It agrees with Lakoff, however, in demoting truth and truth-conditions from the position of a central problem in semantics. In the next section, I will present a brief outline of this approach and try to indicate the ways in which it seems to be a more promising framework than those presented above; in the last section, I will offer a conception of truth which, I believe, fits naturally into this framework.

## 2. Conceptual semantics

Conceptual semantics understands the inquiry into meaning as an investigation of our conceptual structure, an investigation whose ultimate aim is to reveal the structure of thought itself. Its central premiss is that meaning is expressed in language, but is not part of language per se: meaning is located in our conceptual structure and merges with thoughts

arrived at by other means. Now, this conceptual structure can be split into three tiers: the *descriptive* tier, which encompasses the algebra of conceptual units; the *referential* tier, which encodes the referential claims that some of these units involve; and the *information structure* tier, which deals with the distribution of information within a sentence. The distinction between the descriptive and the referential level is of crucial importance, because it allows us to loosen the stranglehold that "the world," in the opinion of many philosophers, has on meaning. Many philosophers believe that meaning has to be "anchored" in the world, that linguistic units have to stand in some kind of real relation to things in the world; otherwise, they fear, we will never be able to break out of the prison-house of language and connect to reality. More specifically, they subscribe to externalism, the doctrine that "meaning ain't in the head," namely, that the meaning of linguistic units (and, more generally, content as such) is partly determined, and therefore controlled, by our environment and our causal links to it.<sup>4</sup> So, the terms of our language often serve as "pointers," and a part of their meaning crucially depends on what they point to in the world, and what the composition or internal structure of that entity is (as discoverable by science). This is, however, a very problematic idea, and is perhaps the result of a certain philosophical bias, namely the metaphysical need to be Linked to Reality. Now, we, of course, *are* linked to reality in the sense that we constantly causally interact with it; but the idea that we will be able to establish direct links between the units of language and things in the environment and thereby appease certain metaphysical fears seems extremely far-fetched, and this semantic project hasn't delivered particularly impressive results. It is, of course, often the case that our learning of the (meaning of the) word "horse" will involve (visual and other) contact with horses; in the perspective of Conceptual Semantics, it will help establish in the mind of the speaker a visual prototype that will form a part of its meaning. The idea, however, that we can account for the meaning of "horse" by invoking some kind of real relation that it constantly stands in with respect to horses, that we need the causal link to the Ur-horse to secure its meaning is very implausible; Putnam himself, who was one of the originators of the externalist view, later argued that these causal links will be too many, or not the

<sup>4</sup> That is one part of the externalist view of meaning, namely causal externalism; there is also social externalism, which claims that meaning depends on other speakers. I find social externalism unproblematic (if construed in a certain way), and will not discuss it here. Cf. Jackendoff (2002: 281).

right ones, or in some cases non-existent (cf. his 1981). In any case, each individual speaker will enter into causal relations with horses that are multiple and also different from those that other speakers enter into; which among the vast number of these relations are the ones that establish meaning? The meaning of each linguistic unit plays a role in the enormously complicated mental structure that allows us to think and communicate in the ways we do. But this mental structure is not linked to the environment in any direct and obvious way, and the attempt to trace individual elements of it back to their real-world counterparts and thereby secure linkage to reality may well be intractable. Also, the way we think and conceive of things is not slavishly dependent on the "real world": as Pinker (2007) shows, and as Goodman (1978) has argued long ago, the mind has great powers of construal and reconstrual. We can present (to ourselves and others) the same things or events in many different ways: we can talk of three horses in the stable, or of horsemanship thrice instantiated in the stable, or even of horse-stuff being unevenly distributed in the stable; we can talk about a man riding a horse, or about a horse carrying a man, or about a moving horse-man unit. It seems much more interesting to investigate these mental powers we have than trying to anchor our mind in reality in a simplistic way.

Here's how the descriptive tier/reference tier distinction may allow semantic analysis to proceed in a fashion more plausible than an externalistic account would allow. Take "Bucephalus", the name of an actual horse, namely the horse of Alexander the Great; take also "Pegasus", the name of the mythological flying horse. Now, on the externalist account, the semantics of these two names will differ crucially: "Bucephalus" will refer by virtue of the causal link, preserved through generations, to the real horse of that name, and this causal link will be constitutive of its meaning. The meaning of "Pegasus", on the other hand, will have to be accounted for either in terms of some kind of description, or a different kind of causal link will have to be invoked, namely one to the myth where this name first appears. But do we really want our semantic analysis to depend on historical contingency? Do we have to make the semantics of these two names crucially dependent on what did or didn't exist a few millennia ago? It seems more plausible to have a level at which these two names have the same kind of semantics, and another level where the link to "reality" is dealt with. The latter is handled in the referential tier, and the former in the descriptive tier. Now, the referential tier is also envisaged in a mentalistic fashion: it encodes, not links to the real world, but to

the world as conceptualised by the speaker. So we might say that "Bucephalus" carries an "ext-feature" at the referential level ("ext" for "external"), namely that it is marked as referring to something that belongs to the world independent of the mind of the speaker (although this world also has to be conceptualized by the speaker in order for him to be able to talk about it). "Pegasus", by contrast, will carry an "int-feature"<sup>5</sup>, indicating that its referent is not conceptualized as belonging to the mind-independent world. At the descriptive level, however, these names will be amenable to analogous analysis in the sense of semantic markers, which can be subdivided into formal, constitutive, agentive and telic ones (cf. Jackendoff 2002: 369-373, drawing on the work of Pustejovsky). So, Bucephalus will be characterized by being a horse (a formal, or taxonomic, marker), by having certain size, shape, etc. (a constitutive marker), by having come into being a certain way (an agentive marker), and by serving a certain purpose, namely carrying Alexander the Great into battle (a telic marker); and analogously, *mutatis mutandis*, for "Pegasus". Now, the semantics of names, and the problem whether they have descriptive content, is an extremely controversial issue. However, even staunch anti-descriptivists such as Devitt (cf. his 1999) admit, under the heading "the qualification problem", that some kind of descriptive content has a role to play at some level. Even if the analysis sketched above is completely misguided, the basic idea is that there is much to be said about the mental structure that underlies our use of language and that the investigation of this structure seems to be a much more interesting road of inquiry than the attempt to link individual linguistic units to "things in the world."

The inquiry into the descriptive tier has as one of its goals the uncovering of the basic elements of cognition and the innate constraints on conceptual structure. The central hypothesis is that, in order to be able to learn the meanings of words, we have to be endowed with a certain conceptual structure that is not learned and is therefore innate (i. e. genetically prespecified). The Quinean gavagai-example shows that the externally observable conditions of language use do not uniquely specify the conceptual scheme that is at the basis of this use; the number of possible alternatives is therefore too great for the child to proceed in its learning process on the basis of these observable conditions alone. It is therefore ex-

<sup>5</sup> Alternatively, we might take one of those features as default, so that its presence will be assumed in all cases unless it is cancelled by the presence of the non-default feature.

tremely plausible that the basic conceptual apparatus is already innately prespecified, and that this apparatus enables the child to learn the “right” meanings (or, indeed, to learn any meanings at all). The units of this apparatus are the “universal elements of thought” that the child operates with (unconsciously, of course) in building word-meanings. The list of these basic elements might look something like this (cf. Pinker 2007: 81):

- certain basic concepts: event, state, thing, path, place, property
- a set of basic relationships: acting, going, being, having
- a taxonomy of entites: human/nonhuman, animate/inanimate, male/female, object/ stuff, flexible/rigid
- spatial and temporal categories
- causal and teleological concepts

Now, whatever the ultimate result of this kind of investigation turns out to be, it seems to be a very exciting research program that is almost completely neglected by the truth-conditional approach, which focuses on the way that language is anchored in “the world.” Also, Conceptual Semantics seems to be able to integrate research programs pursued by other theoretical frameworks. It can talk about reference and truth, the strong points of truth-conditional semantics: how it might handle reference was sketched above, and how it might account for truth will be the subject of the next section. Furthermore, it can integrate the valuable structuralist insights into literary discourse, and precisely because reference and truth are not its basic notions. Truth-conditional semantics almost completely neglects literature because its methodology is tailored to account for the way language connects to reality and is therefore incapable of grasping a kind of use of language where that connection is abandoned; literary discourse is, however, a very important and rich way in which language can be used, and a semantic theory that is inherently incapable of dealing with it has a very serious problem, whether it is aware of it or not. Conceptual Semantics, on the other hand, has no principled problem with literature. Finally, it can also integrate pragmatics because the explicature and the implicature are obviously worked out at the level of conceptual structure, so that Conceptual Semantics seems very well suited to serve as a background theory for pragmatic investigation.

### 3. Truth as multiple fit

In Conceptual Semantics, the primary category corresponding to a sentence is an event or state, not a truth condition or a truth value. Sentences express events or states and are analyzed by employing cate-

gories listed above (and others), whereas truth is an evaluation that speakers and hearers (i. e., primarily, thinkers) ascribe to sentences<sup>6</sup> on the basis of their understanding of the sentence and their understanding of the world-as-conceptualized. Truth is not seen, therefore, as some sort of objective correspondence between mind-independent propositions and mind-independent states of affairs;<sup>7</sup> rather, truth is reconstrued as truth-effect that happens in our conceptual structure.<sup>8</sup> Now, this does not mean that truth becomes completely relative and subjective – but its non-subjectivity is understood as intersubjectivity rather than objectivity in a mind-independent sense. The metaphysical quest to preserve the mind-independence of truth requires that we postulate something out there, some kind of Transcendent Entity, that guarantees the truth (or falsity) of what we think or say, that is the truth-maker of our sentences. This attempt leads, however, into a metaphysics that seems both cumbersome and unscientific. We don't know what the world consists of; or, if we do, then it is elementary particles, or, in any case, it is whatever fundamental physics says it consists of. This does not square at all well with “facts”, the favorite truth-maker for many philosophers. In an attempt to secure the truth-makers for our truths, philosophers will portray the world as consisting of facts, but it is completely unclear how this view is to be justified. No empirical argument has been offered for the existence of facts; science doesn't seem to support such a view. A priori arguments, on the other hand, are extremely questionable when trying to establish the existence of anything (let's not forget the notorious “ontological proof” of theologically-minded philosophers). Furthermore, it seems that the only a priori consideration adduced in favor of facts is that we need something for our sentences to correspond to, whereas one would like an independent argument that purports to show that there are facts. Facts are, in fact, often modelled on sentences (or

<sup>6</sup> It is probably best to speak of truth as ascribed to statements, i.e. the thoughts expressed by sentences, rather than to sentences themselves. However, when the sentence is written down rather than uttered, the term “statement” might sound awkward. I will therefore use both terms interchangeably, but the basic idea is that truth (and falsity) pertains to thoughts, not grammatical structures.

<sup>7</sup> This is the basic idea of the correspondence theory of truth, usually the most popular one; there are at least four others, namely the coherence theory, the pragmatic theory, the redundancy theory and the semantic theory. For the most extensive contemporary overview cf. Künne (2003); some well-known earlier sources are Haack (1978) and Walker (1997).

<sup>8</sup> I will be drawing here a lot on my 2008.



propositions): Russell and Wittgenstein considered them to be linear. But it is sentences that are linear (namely, they consist of elements that succeed one another), not the world: the world seems to be simultaneous. In any case, the question “how many objects are there in this room?” seems reasonable (pending the specification of what is meant by “object”, of course); the question “how many facts are there in this room?” on the other hand, seems unintelligible.

There has to be another way to avoid the complete subjectivity of truth and that is the way of intersubjectivity. Truth isn't a matter of personal opinion, not because there are transcendent units out there that correspond to units of our thoughts/sentences, but rather because there are other minds that might also have an opinion on the matter and might challenge our opinion. We don't know what it is out there that our minds have a fragmentary grasp on, but this grasp itself will conveniently be in the same mental/linguistic format amongst thinkers/speakers, and this allows us to compare and discuss our views. This openness to challenge is what fundamentally secures the non-subjectivity of truth. Whatever it is that minds grasp (or “mirror”), different minds can obviously grasp it in different ways, and some more correctly than others (as consequences of behavior will ultimately show, but without correctness being reduced to success in behavior). Any claim to truth is therefore open to challenge by another person who claims to have a more correct view of the matter; the burden is then on them to furnish arguments in support of their claim. What lies at the basis of this game of challenge-and-answer is of course some notion of universal human rationality: namely, the idea that even if we don't agree on an issue, and even if we don't understand each other, at least the possibility of us understanding each other and furthermore agreeing is never closed off. As humans, we all share basically the same mental structure (this is, at least, a very plausible working hypothesis), and it is this common structure that is the final guarantee of the possibility of mutual understanding and of the achieving of a shared view. I believe that this is the most we can ever expect to have in the way of a guarantee of objectivity. “Reality” will have a way of playing a role, of course: the way it will come in is that different people will gather different kinds of evidence, and bring it into the discussion. But the hope that we will find some kind of ultimate reality out there for our thoughts/sentences to correspond to seems to me forlorn. And, as has been noted before, the mind is very flexible and not a slave to “reality”: it has ways of construing and reconstruing events or states, of building alternate mental models of the

same things. Has Iraq been invaded or liberated? This question is not to be settled by appeal to some ultimate Fact that will resolve the issue; the best we can do to is to offer arguments and evidence in support of one framing or the other.

Truth will, therefore, be a truth-effect that happens in our conceptual system, and is then open to challenge from other minds. It is now time to say more about what this effect consists in. I propose that truth is a matter of multiple fit. Now, the attempt to shed light on the problem of truth by employing the notion of fit can be found in the works of some important philosophers (cf. Goodman 1978, Putnam 1981), yet this strand of philosophising about truth has been rather neglected recently (even in Künne 2003). In addition to its being out of tune with what's currently fashionable, what might explain the neglect of this approach is also the fact that the notion of fit hasn't been sufficiently worked out in the writings of the philosophers that introduced it. What Goodman has to say on fit comes down to this: “[we should] subsume truth [...] under the general notion of rightness of fit” (1978: 132) and: “a version is taken to be true when it offends no unyielding beliefs and none of its own precepts” (17); as for Putnam, it's the following: “Truth is ultimate goodness of fit” (1981: 64, italicized in the original) and: “Truth ... is ... some sort of ideal coherence of our beliefs with each other and with our experiences as those experiences are themselves represented in our belief system” (49-50, italics in the original beginning with “as”).

There are two ways in which the notion of fit might be elaborated in order to make the truth-as-fit approach more appealing. On the one hand, we need to specify where exactly this fit takes place. On the other, we need to say more about what it is that our thoughts fit.

The natural way to satisfy the first demand within the framework of Conceptual Semantics is to point out (as has already been done above) that the fit takes place in our conceptual structure. Jackendoff's theory of a “common cognitive structure” that serves as the locus of convergence for meaning acquired through language, inference, background knowledge, perception and action (cf. his 2002: 273) seems to be just what we needed to elucidate the notion of fit. The fit-effect happens, therefore, in our conceptual system: it takes as input the meaning conveyed by language, the data of perception, the operations of inference and whatever relevant knowledge we have stored and delivers the “that's true” impression as output.

As for the what issue, that's where the proposal of multiple fit comes in. Philosophers most often

assume that truth-makers of our statements are only one kind of thing, and that truth itself is essentially one kind of thing. In the view proposed here, there is indeed something that is common to all true statements: namely, that they fit something (that they “click” with something in our conceptual structure). However, what they need to fit in order to be deemed true might be very different things, depending on the occasion and the type of statement in question. Here is a (probably non-exhaustive) list of the “factors of fit” that are responsible for the truth-effect of statments:

- the criterion of internal consistency
- sensory data
- data from memory (long-term and/or short-term; semantic and/or episodic)
- non-verbalized beliefs (given in conceptual structure but not expressed in language)
- other parts of discourse

Since the criterion of internal consistency is, so to speak, eternal, whereas the data of perception are always subject to change (both in the sense that things change in front of our eyes, and that these data “enter” and “exit” our short-term memory, with only some of them sticking around by being stored in long-term memory),<sup>9</sup> that leaves the data from memory, non-verbalized beliefs and other parts of discourse as relatively stable, though not immutable, parameters of evaluation. I would like to call these last three factors the sedimented background. Most statements will need to be checked against all or at least some parts of the sedimented background (i. e. all or some of the three constituents, not the whole of each of these factors, of course).

This is one way in which the notion of truth might be reconstrued in a way that makes it less metaphysical and more cognitive. Another way is to point out that there isn't just one kind of truth around, but rather a family of types of truth. The central examples of truth, the prototypes, are mathematical truth, everyday or commonsense truth (“There is a cat on the mat”), and scientific truth. But there are other, more marginal or derivative, kinds: artistic truth, metaphorical truth, ethical or moral truth, political truth. Whether a type of truth is central or derivative is a matter of empirical research, the same as whether a whale is considered by people to be a prototypical mammal – it is determined by investigation of our cognitive system, by

looking into what people most easily classify as “a truth”, and where they begin to diverge.

What is the relation of these kinds of truth and the factors of fit? Well, mathematical (and logical) truth need satisfy only the criterion of internal consistency;<sup>10</sup> as for the other kinds, all the factors come into play, their respective role depending on the features of the particular statement (its particular subject, its generality, the nature of the concepts it employs, etc.) and the context (both linguistic and non-linguistic) of the utterance. Internal consistency will figure as a less important desideratum in artistic truth than in scientific truth, and so will agreement with observation (i. e. fit with sensory data) – but a particular artistic claim to truth might engage much more of our cognitive system than a scientific claim to truth (e. g. it will engage much of our episodic and semantic memory, whereas the scientific claim will engage only some parts of semantic memory).

These, then, are some of the ways in which the notion of truth might be reconstrued in order to free it from the grip of certain metaphysical preconceptions on the one hand, and integrate it into the framework of Conceptual Semantics on the other.

## References

- Chierchia, G. and McConnell-Ginet, S. (2000) *Meaning and Grammar*. Cambridge, Mass.: MIT Press.
- Davidson, D. 2001 (1984): *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.
- Devitt, M. and K. Sterelny (1999) *Language and Reality*, 2nd edn. Oxford: Blackwell.
- Goodman, N. (1978) *Ways of Worldmaking*. Indianapolis: Hackett.
- Greimas, A. J. (1986) [1966]: *Sémantique structurale: recherche de méthode*. Paris: Presses Universitaires de France
- Haack, S. (1978) *Philosophy of Logics*. Cambridge: Cambridge University Press.
- Jackendoff, R. (2002) *Foundations of Language*. Oxford: Oxford University Press.
- Künne, W. (2003) *Conceptions of Truth*. Oxford: Oxford University Press.
- Lakoff, G. 1987: *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press.
- Lakoff, G., and Johnson, M. (1980) *Metaphors We Live By*. Chicago: University of Chicago Press.

<sup>9</sup> So there is a partial overlap between the demand of fitting perceptual data and fitting data in short-term memory; however, short-term memory comprises much more than the data of perception, e. g. recently heard sentences, recently thought thoughts, etc.

<sup>10</sup> “Other parts of discourse” might also be relevant, in the sense of the axioms and other theorems, as well as the rules of inference, a particular statement is deduced from. However, each deduction can be presented as one long conditional, so the matter comes down to internal consistency again.

- Pinker, S. (2007) *The Stuff of Thought*. London: Penguin.
- Putnam, H. (1981) *Reason, Truth and History*. Cambridge: Cambridge University Press.
- Recanati, F. (2004) *Literal Meaning*. Cambridge: Cambridge University Press.
- Sperber, D. and Wilson, D. (1995) [1986] *Relevance*. Oxford: Blackwell.
- Tarski, A. (2001) [1944] “The Semantic Conception of Truth and the Foundations of Semantics”, in: Lynch, M. P. (ed.) 2001: *The Nature of Truth*. Cambridge, MA: MIT Press.
- Walker, R. C. S. (1997) “Theories of truth”, in: Hale, B., and Wright, C. (eds.) 1997: *A Companion to the Philosophy of Language*. Oxford: Blackwell.
- Žanić, J. 2008: “Truth: a multiple-fit theory”, ms.

## ON WHAT VALUE, MY LORD? HOW VALUES INTERVENE IN HARD LEGAL CASES

*Mladen Domazet*

Institute for Social Research – Zagreb, Croatia  
domazet@idi.hr

### **Abstract:**

The paper confronts the issue of single jurisprudence facing a value (-system) pluralism, the one often arising nowadays. Starting from the Raz – B. Williams debate, it outlines a proposal close to Raz's but ontologically less demanding.

**Key words:** value, pluralism, jurisprudence, relativism.

### **Judges are bound by the values that are (exist) out there.**

One pertinent fact about our world is that it contains people of different and conflicting religious, moral and philosophical beliefs. J. Rawls claims (1993) that the facts of pluralism consist not merely in there being different and competing comprehensive conceptions of the good (moral beliefs regarding what is of value), but different yet *reasonable* comprehensive conceptions of the good. So we can't discard all but one of those opposing views as mad, bad or dangerous to know. Some may be such, but there will be many comprehensive conceptions which are perfectly reasonable, but which conflict with other, equally reasonable, conceptions. For Rawls this is a predictable outcome of the operations of reason under conditions of freedom.

Yet, in contemporary Western societies (but increasingly elsewhere in the world, too), all those people are expected to operate, under a single legal system of a state, or even international law. Legal positivism of the straightforward Austinian variety (Austin, 1832) offers a first step towards dealing with this predicament, by insisting on sharp separation of law and morals. Whatever the moral or other beliefs of any individual may be is irrelevant to legal practice that is governed by positive law (rather than supposed natural law<sup>1</sup>). Thus, though the judge's

values may not coincide with the values of the accused, both will rely only on the value-free system of rules governing procedures and sanctions related to the action considered to have been a transgression of those rules.

Ideally the said system of rules will be deductively complete, comprehensive and internally coherent. Execution of legal judgements should then be, given clear-enough evidence, almost an automatic process. Thus, Kelsen and other positivist idealists (Austin in England) hope for an entirely deductively closed system of laws that all derive from the foundational axiomatic norm. But, in practice, Hart warns, the ideal of completeness is unfulfilled and judges sometimes have to make decisions that are not automatic and deductively proscribed. Rational justifications for such judgments will often draw on some moral principles, and this may be seen to be working against the recognised pluralism in society. It may on some occasion be rationally justified to use the discretion to draw conflicting conclusions by drawing on conflicting value systems.

On another hand, there is no absolute vouchsafe that the existing system of rules itself is internally coherent (incoherence is discovered when two rules contradict each other in a single case), and when issues of incoherence arise straightforward positivism (i.e. pre-dating Hart and Dworkin) does not offer guidelines how to resolve them (i.e. what independent standard is to be used in choosing preference). If we were to try to order conflicting rules according to importance (e.g. rules governing driving and rules governing individual liberties) we would need a common scale of importance against which to judge all the individual rules. This is where morality seems to make a comeback into the legal system, and Dworkin calls for reliance on extra-legal principles that help interpret the "internal" rules. He also sees this as a serious weakness of legal positivism.

<sup>1</sup> Moreover, the natural law (Kelsen, 1967) lacks the element of force, i.e. has to be adopted by humans into

some positive law in order to be implementable, and thus instrumentally meaningful.

ism.<sup>2</sup> Moreover, Dworkin (1977) argues that there are extra-legal principles, which will be called values below, that must be binding for the judges when applying (supposed) discretion in hard cases. And what values are they to draw upon in cases requiring explicit respect for pluralistic character of the society today?

So the problem of different value systems co-existing under the roof a single legal system cannot be avoided by escape to legal positivism. If the conflicting value systems, as ascribed to the contemporary societies above, were to be equated with simple social relativism (the values are dictated by community proclamation) then any moral system judges may draw on in discretionary decisions will be an act of oppression of one social group against another. It is therefore desirable to respect pluralism whilst striving to secure a distance from social relativism.<sup>3</sup> This is the aim of Raz's (2003) lecture. Raz's proposed solution rests on a "*metaphysical thesis*" that values exist independently (modulo emergence conditions) from social groups espousing them. Thus the objective existence of values is a precondition for ordering them in a hierarchy of importance.<sup>4</sup> Such a hierarchy would then allow drawing upon those values in legal cases without committing a political oppression of one social group over another, as to adjudicate in such matters

would be rationally justified and not arbitrarily or politically imposed. Raz claims that his social dependence theory (SDT) can account both for objectivity (allowing absolute ordering within a genre when called for) and fluidity (allowing for present state of pluralism) of values.

### Existence of values is superfluous

Williams (2003) criticizes Raz mainly for unnecessarily introducing a much stronger metaphysical thesis than is needed. Williams picks forcefully at Raz's arguments for the existence of values, but admits that the thesis itself is useful for bringing to the fore the historical nature of our value systems. He, in fact, thinks that Raz can achieve the stated aim "without asking questions about the conditions under which various values exist" (Williams, 2003: 114).

Williams says that he does not understand what it means for values to exist, and that Raz's thesis does not answer this as it gives the conditions for existence but presupposes that everyone knows what such existence means. His conditions specify what it means for a value to come into existence (an *emergence* condition), namely that there should be a social practice sustaining it (and he acknowledges Raz is very detailed and explicit on how this is to be achieved), but not what it is for a value to continue existing even without the social practice sustaining it (a *continuation* condition). Without some such continuation condition, Raz's thesis merely starts off at the same place as relativism does, namely that values arise or are in the first instance embodied by the social practices sustaining them in the right way, but then unjustifiably projects the metaphysical claim that they continue existing even when no longer practically supported by the relevant social practices. Moreover, Raz suggests that once the values come into existence through a sustaining social practice, that existence is projected across cultures and even backwards in time, to a time before the sustaining social practice arose.

Williams' criticism of Raz's thesis with respect to political values is of special interest here (political values being more closely bound with law than aesthetic ones): can we say that certain values exist even in the societies that do not know of the social practices sustaining them? If we replace "societies" with 'cultural groups within a society under a jurisdiction of a given legal system' it becomes obvious why the question is important for the issues of judicial discretion. Can such discretion be applied based on a system of values of a dominant cultural group (being the one the judge belongs to) alone? This is obviously a claim that Raz cannot support given that

<sup>2</sup> An important disclaimer may be called for here. From the above discussion it may seem that most, or even substantial proportion of legal cases are difficult to decide (given all the facts). This is not the case, and most cases are dull and obvious (meriting the automatic treatment the positivists may have hoped for, above). But when it comes to of principles behind the legal system, as well as the thin line between the legal system and the wider social machinations, the hard cases raise the most interesting issues for understanding of the law (Goldman, 1999: 278).

<sup>3</sup> Relativism, though, as all authors in the debate admit, is a murky ground with many sufficiently distinct versions present (it was once said that there are as many relativisms as there are relativists). Perhaps it is important to bear in mind, pre-empting Williams' criticism below that it is only the negative characteristics of relativism (once we have singled those out clearly) that have to be avoided, whereas the concept as a whole is more of a scarecrow than a real danger.

<sup>4</sup> Admittedly, Raz is not aiming to propose some absolutist picture, that merely pays lip-service to pluralism, whilst presenting a strict hierarchy of all value. He uses a genre-based picture to allow for incommensurability of some values in some situations. But, with respect to the genre (and this could be a courtroom of a pluralistic society), some comparisons can be made even of values that appear to come from the conflicting comprehensive conceptions of the good.

he wants to distance himself from relativism. Williams forces Raz's position into a knot: on one hand the values that are "brought into being" by one cultural group "bear on everything" outside that cultural group and contemporaneous with it, as well as what comes before and after it. Thus, those values "behave" as if they always existed and always will. But then it becomes an evaluative judgement against the cultural groups that don't endorse those values that they somehow fail to recognise their worth. Williams finds it easiest to illustrate this with a historical point: if we assume that most of the contemporary liberal political values came into being in the eighteenth century, but then spread across time, we must say it was a failing on behalf of ancient or medieval societies not to have endorsed values that existed, i.e. were available to them.

... what was wrong with the pre-modern world, that it did not recognize these values? Why did the existence of these values, which had always been there, only burst on the world in the eighteenth century? (Williams, 2003: 113)

Williams thinks the problem can be solved by dropping talk of the existence of values, and plainly admitting the relativist nature of all values as simply a product of cultural practice that may or may not spread across cultures (both spatially and temporally), but can only do so if adopted as practice by other cultural groups. And this appears to be in spirit of Hart's warning that "if we have learned anything from the history of morals it is that the thing to do with moral quandary is not to hide it" (Hart, 1977: 33). The problem of specific interest here, though, is how to rationally justify judges applying discretionary powers based on the values of their cultural group over members of a pluralistic society that may be from another social group. If they were to admit that they are simply using their power to uphold values of their own cultural group, this would certainly peel away strongly from the ideals of legal positivism and rule of law ("...to the end it may be a government of laws and not of men" (Massachusetts Constitution, Part The First, art. XXX (1780))).

It is often the case in philosophy that matters can be obscured, problems only seemingly resolved by unjustifiably stepping into the realm of metaphysics, by simply postulating a metaphysical solution. There is no room here, nor ambition, to solve the issues of moral realism, though Raz and Williams admit that they differ from the outset, in their views on the topic. We can merely repeat the motivation stated above, that a realist solution would in most cases make the judiciary discretion an easier task, by allowing judges to draw on something existing, though possibly intangible. Certainly, if existing it is

open to everyone to appropriate and their failure to do so may be taken as a moral (and pace positivism in hard enough cases this may coincide with legal) act on their behalf. But, Williams' criticism of Raz explication of how values come into existence and are connected to the social and historical circumstances cannot be dismissed or ignored by a sturdy belief that values exist.

### **Williams' relativist strategy and its problems**

A mere recognition, says Williams, of the pluralism in a given society today (such as expressed in Rawls' claims we started with) is sufficient to raise awareness of the pluralistic nature of values that judges may wish to draw upon. The ordering of the conflicting values that agrees with the fundamental postulates of the society within which the recognising takes place should follow such explicit recognition. But it would not, in and of itself, simply justify raising the principles upholding pluralism above all others, for that would be turning a simple thesis about values into another political or ethical ideal (a thing to be evaluated using the thesis about values). Thus, Williams calls for explicit admittance of when reliance on some social conditions is exercised, rather than trying to fudge matters with metaphysical posturing. Let's be relativist and unashamed.

But the problem is that without the metaphysical thesis, Raz's model can be reduced to the admittance of the power struggle as the fundamental norm upon which the legal system rests. Namely, when Pippin (2003), in criticizing Raz, claims "that reason is incapable of ever resolving the dispute in favour of one side or the other (that the matter is therefore essentially a political contestation, a struggle for power)" (Pippin, 2003: 101), Raz agrees and says that political philosophy cannot solve the deep social divisions and problems. What he does not call upon, and in our context this seems of paramount importance (given that he can, which remains open, overcome Williams' arguments against the existence of value), is the importance of the supposed existence of value in assisting the judges in making discretionary decision, in drawing upon the extra-legal principles. Most notably, it seems that without further complications to the metaphysical thesis, the possibility of which is not excluded but is better avoided here, openly conflicting values cannot exist for they are negations of each other (for the sake of simplification we should ignore Raz's genre-based value conflicts). Thus, an act either bears moral value or it does not, that is what makes the judges' job much easier.

Sadly, though for the present purposes instructively, Raz all but abandons the metaphysical thesis

for the instances of values most under duress here. In his replies to criticism (Raz, 2003a) he says that the problems that Williams brings up (among which is that of 'liberal political values') were not explicitly treated in his original lectures and deal with borderline values that may require additional specification and do not straightforwardly fit the metaphysical thesis. When forced to face the music on the issue of political values, most notably how they can "bear on everything without restriction", Raz denies that they do so. He says those values are highly genre- (or kind-) specific and are only applicable in certain political contexts. He therefore agrees with Williams that it is senseless to debate liberal issues with King Arthur or medieval Roman popes (Raz, 2003a: 152). Political values have to be recognised as such, and we should be wary of a repeating human mistake to convert some of the political values into more universal moral values. Liberal political values are political, however universal their advocates may think them to be. Thus, liberal values are simply not of the kind we often take them to be (and of the kind Williams understands them to be in his criticism).<sup>5</sup>

In perhaps the most pressing issue for pluralism in contemporary society, that of political values, Raz and Williams effectively agree (whether they happen to use the same terminology or not): unashamed relativism is the best that philosophy can provide. Yet that presents a problem that shouldn't be dismissed too lightly, and that brings us back to the strong motivation for the (now discarded) metaphysical thesis. For if social relativism is the model (the view that Raz aims to reject in 2003), and the power struggle is the norm, then the whole point of legal conduct might be called into question. Given suitable premises (most notably those pertaining to social and historical context), other forms of power struggle may also be rationally justified (especially if an early positivistic stance of authoritative though openly immoral positive laws is accepted), though directly conflicting with respect for the legal system.

In an admittedly oversimplified, but hopefully illustrative example, why respect any ruling by the judge, be it based on legal proscription or moral justification, if it is merely a product of a successful

power struggle of one social group against others?<sup>6</sup> On the other hand it does not straightforwardly legitimise terrorism or similar forms of power struggle, for those may present a transgression against even more universal values than were ever under discussion in a specific judicial ruling. We should be wary of using borderline cases to legitimise a jump to general conclusions. Again, there is no room here to investigate this matter further, and the main reason for bringing it up is to refresh the motivation for a metaphysical thesis.

Where does the above leave the judge? In explicit failure of simple legal positivism the judge was forced to draw on some system of values, external to the law, in passing discretionary ruling on a "borderline" case. This would be most easily achieved by drawing on a system of existing values (in some realm), fixing the failure of legal positivism by effective transferral to ethical positivism (the values are what is out there and are in no need of additional interpretation). The model most likely to respect the pluralism of contemporary society that the judge is to make her ruling in was Raz's metaphysical thesis above (or some such along the lines of social emergence of values, but application unrestricted by social context). Otherwise, the metaphysics of values collapses into relativism—which may have a metaphysical dimension, though would have to solve the problem of the concurrent existence of mutual negations) which would have problems similar to the ones stated above, or some form of naturalism (the values that exist are the ones that are natural, or decreed by a supreme being, and have always existed) which would have a hard time showing desired respect for pluralism. Yet, in the case of political values, the most likely ones to be a part of out "borderline" case,<sup>7</sup> Raz and Williams agree: the explicit metaphysical thesis fails. Given that it does, and given that it was the most likely one to meet the motivation and the demand for pluralism, what options are left to the judge?

### Epistemology vs. metaphysics

<sup>5</sup> This is in line with Raz more general claim in the replies (Raz, 2003a), that everything must be first assigned a kind and then evaluated as an instance of its kind (a take on genre-specificity). It is senseless to judge how good architecture a film is. Though such a view is leaning dangerously close to Korsgaard's Aristotelianism (Korsgaard, 2003) that Williams criticizes (Williams, 2003: 115).

<sup>6</sup> The imperative to respect the judge's ruling may come through the use of force, which is another important ingredient of the positivist view of a legal system, but it is also yet another instance of a power struggle.

<sup>7</sup> Important rulings may be called for in hard cases not explicitly connected to liberal political values, such as might be those of aesthetics (what is a work of art and what is not), but when they become a matter of adjudication between interest groups within a society (and not one of straightforward expertise) they can then be subsumed under general political values case.

When choosing values against which the transgressor has “sinned,” in cases where there is no positive rule straightforwardly applicable to the case at hand and universally understandable to be such, the issue before the judge is effectively one of finding values to replace it. And the transgression against values, as well as against positive laws, is one of conscious and willing disregard for behavioural prohibitions they state. With laws it is assumed that they are all known to the transgressors and at best a small mitigation in a sentence passed<sup>8</sup> can be provided by the fact that the transgressor simply had no inkling that his actions violate a positive law. Likewise with values, the judge in using discretion in problematic cases has to stipulate that the transgressor wilfully chose to act “badly,” to go against his knowledge of what is of value. And this would most easily be achieved in that antithesis of positivism – naturalism; where the judge could claim that the value was either ingrained in the transgressor as a human being and that he or she chose to ignore its counsel, or that it was somehow naturally (even super-naturally) mandated and thus available and applicable to everyone. But, pluralism and pluralistic societies have long ago given up on such naïve naturalistic proscriptions both of law and value.

The next best step, then is to claim that though the value was not ingrained in the individual transgressor, it existed, i.e. was out there for him or her to become cognizant of it and to subsequently show respect for it. But even this is philosophically going a step too far, for as explained above the effective transgression is in knowing and ignoring. And to argue how one must come to know, might in this situation be seen as too much philosophical baggage. Not that such baggage would not solve a host of problems. But values, unlike laws, are in pluralistic societies not written in ink or stone, and thus placed before everyone. Even those that claim to respect them would be hard pressed to name them all, not to mention to show how they fit into a reasonably coherent whole. Having them existing in whatever realm seems to alleviate this burden.

Yet, we have seen it is hard to philosophically justify this assumption, and even if we could do it, our judge in passing a sentence would still have to show that not only did the value(s) he draws upon exist, but that it was cognizable to the transgressor as well. So the crux is that the transgressor could have known about a given value and its place in the

society. And in a pluralistic society we might have to face a situation where the transgressor could have lived in a cultural group that did not value what the judge values, or may have even valued the opposite. In a parallel scenario with laws (cf. Hart, 1977: 32–33) the judges are advised to step out of the legal jurisdiction (competitive legal systems, one of the transgressor and the other of the judge) and explicate the moral quandary they find themselves in, explicitly drawing on a system of values external to both legal systems but somehow presumed universal. They are advised to choose the lesser of two evils, “.... with the consciousness that they are what they are” (Hart, 1977: 33).<sup>9</sup>

But is there a further stepping out to be done in the moral case, what quandary can we explicate in this case? The further step, the quandary made explicit, in such case seems to be epistemological. Any member of the society, no matter what cultural group he or she belongs to, must be conscious of value pluralism inherent in contemporary society<sup>10</sup> and the limitations that places on his or her own value system. The judge should then draw on a system of values, genre-specific as they may be, that is upheld by the democratic and academic leadership of the society (the sort of “moral professionals”) as a representation of the hierarchy of values as ordered in the self-consciously pluralist societies. These are the values, to a slightly lesser degree such is their hierarchy as well, that the transgressor either knew or could have known about, and as such a binding for him and the judge.

What is the difference here from Raz’s original metaphysical thesis (Raz, 2003) and Williams’ argument for unashamed relativism (Williams, 2003) above? Obviously there is no explicit talk of the values existing, and what the democratic and academic leadership upholds may be rather vague, in a sense of not being written in stone (and sometimes not even in ink) and explicated to the level of fine detail. On the other hand it does take us a step back (though, admittedly not a leap) from the explicit admittance that the value system applied is merely a temporary resolution of a power struggle in a society. The transgressor is sentenced not because the judge has the power to do so, based on the power

<sup>8</sup> Of course, this only applies to standard grown-up individuals of average (or above) health, and not to minors, people raised away from the society (“Robinsons” and feral children) or people with impaired cognitive powers.

<sup>9</sup> With foresight to the issues of contemporary pluralism Hart says that failing to openly admit this moral quandary “... will encourage the romantic optimism that all values we cherish ultimately fit into a single system, that no one of them has to be sacrificed or compromised to accommodate another” (ibid.).

<sup>10</sup> Though it is a further task, and pressing one, for such societies to make their members conscious of inherent value pluralism and to uphold it openly in public actions.



that allows her chosen value system to reign supreme, but because the transgressor chose to act against a value (thus to act “badly”) known to be held in high esteem in the given society.

How this is to be achieved in practice is what Raz’s genre-enhanced thesis does well, minus the talk of existence of values. Once a value “emerges”<sup>11</sup> through a sustaining social practice it bears on everyone that could have possibly known about it (rather than “on everything without restriction”). Like everyone who is told about them can discuss the beauty of unicorns, though they needn’t exist. And Raz, again, is explicit about how non-trivial claim this is, for values involve complex notions that can be combined and recombined in practice and on paper. Just as Williams advises, large sections of Raz’s detailed explication can be adopted for practical positioning of values in a society, largely excluding the talk of existence or taking it to as an instance of misleading grammar. But only omniscience calls for metaphysics (Kusch, 2004: 575) and that is never what we are dealing with in these situations.

As Williams points out, though, sadly, declines to explicate fully (Williams, 2003: 118), the remaining details of Raz’s thesis are more than helpful in drawing out the rationality of judicial discretion such that respects the pluralism in society. Thus many of the values drawn upon are genre-specific and only apply to the instances of the respective genre, which must be evident in the given case. Moreover, new values arise as a generalisation of more specific ones and may subsequently take order of precedence above a whole range of (culture-) specific ones. Contradictions between culture-specific values in pluralist societies needn’t arise if the practice of judging something as good is performed in the following stages:

First, we identify a kind to which [the thing or the act] belongs, a kind that by its nature or construction is governed by a particular value (that is, by standards of excellence for being of that kind); and, second, we judge the item under consideration good (or bad) to the extent that it is good (or bad) of its kind.

<sup>11</sup> Raz himself admits that the grammar of his thesis can be misleading, and that in some instances he talks of the existence of value even when there is no clear metaphysical theory for all that is subsumed under the term value. Williams then picks at the specific problematic instances of it and shows where such grammar may be exceptionally misleading, but Raz still believes that does not detract from the whole enterprise. Likewise, for brevity here it is convenient to talk of the emergence of value as if something existing in its own realm, despite the overall argument being pitched against such thesis.

This allows us to recognize the existence of values with apparently contradictory criteria (Raz, 2003a: 138).

Yet this does not call into question the whole point of legal conduct, as is suggested might be the case with straightforward social relativism above, as

judges’ reasoning needn’t be seen as incapable of ever resolving the dispute between conflicting value systems, thus admitting openly that the matter is essentially a political contestation, a struggle for power. All that remains now is to clearly delineate, for the purposes of fair sentencing, what the judge should know about the transgressor’s value system, as well as what every potential transgressor should know about the judge’s.

## References

- Austin, J. (1832), *The Provinces of Jurisprudence Determined*, London: J. Murray
- Dworkin, R. M. (1977), “Is law a system of rules?” in R. M. Dworkin (ed.) 1977: 38–65
- Dworkin, R. M. (ed.) (1977), *The Philosophy of Law*, Oxford: OUP
- Goldman, A. (1999), *Knowledge in a Social World*. Oxford: OUP
- Hart, H. L. A. (1977), “Positivism and the separation of law and morals,” in R. M. Dworkin (ed.) 1977: 17–37. Originally published in *Harv. L. Rev.* 71 (1958).
- Kelsen, H. (1967), *Pure Theory of Law*, Berkeley: UCP
- Korsgaard, C. M. (2003), “The Dependence of Value on Humanity,” in J. Raz (ed.) 2003: 63–85
- Pippin, R. (2003), “The Conditions of Value,” in J. Raz (ed.) 2003: 86–105
- Rawls, J. (1993), *Political Liberalism*, New York: Columbia UP
- Raz, J. (2003), “The Thesis,” in J. Raz (ed.) 2003: 15–36, also available on [http://www.tannerlectures.utah.edu/lectures/Raz\\_02.pdf](http://www.tannerlectures.utah.edu/lectures/Raz_02.pdf)
- Raz, J. (2003a), “More on Explaining Value: Replies and Comparisons,” in J. Raz (ed.) 2003: 121–156
- Raz, J. (ed.) (2003), *The Practice of Value*, Oxford: OUP
- Williams, B. (2003), “Relativism, History and the Existence of Values,” in J. Raz (ed.) 2003: 106–118

## SHOULD WE CONSULT KANT WHEN ASSESSING AGENT'S MORAL RESPONSIBILITY FOR HARM?

*Friderik Klampfer*

University of Maribor  
friderik.klampfer@uni-mb.si

### **Abstract:**

The focus of the paper is the conditions under which an agent can be justifiably held responsible or liable for the harmful consequences of his or her actions. Kant has famously argued that as long as the agent fulfills his or her moral duty, he or she cannot be blamed for any potential harm that might result from his or her action, no matter how foreseeable these may (have) be(en). I call this the Duty-Absolves-Thesis or DA. I begin by stating the thesis in a more precise form and then go on to assess, one by one, several possible justifications for it: that (i) it wasn't the view Kant himself actually held or was committed to; (ii) there is nothing strange about the DA, either theoretically or intuitively; (iii) the DA is more plausible as an account of legal (either criminal or tort) liability; (iv) the DA becomes perfectly plausible when conceived as a thesis about what insulates the agent from either remedial moral responsibility or the demands of compensatory justice; (v) the rationale for the DA is to protect our moral assessment of agents and their actions from the threat of moral luck. I show, using the famous Inquiring Murderer example, all these (and some other) justificatory attempts unsuccessful. I conclude that besides being counter-intuitive, the DA-thesis also lacks firm theoretical grounding and should therefore be rejected as (part of) an account of outcome moral responsibility.

**Key words:** Kant, benevolent lie, moral responsibility, duty, harm, wrong, moral luck.

### **Introduction**

Are any of our moral duties absolute? Is, for example, our duty to speak the truth (as we know it) unconditional and/or exceptionless? Whoever feels compelled to answer this in the affirmative, should pause to reflect on the following example: Would it really be a crime to tell a falsehood to a murderer who asked us whether our friend, of whom he was in pursuit, had not taken refuge in our house? Benjamin Constant who introduced this story as a counter-example to any moral theory which treats moral duties as unconditional and admitting of no exception, had no doubts about it – telling the truth is a duty, but only towards those who have a right to the

truth. A right that the Inquiring murderer (IM), he thought, has clearly forfeited. To which Kant has famously replied:

Truth in utterances that cannot be avoided is the formal duty of a man to everyone, however great the disadvantage that may arise from it to him or any other; and although by making a false statement I do no wrong to him who unjustly compels me to speak, yet I do wrong to men in general in the most essential point of duty, so that it may be called a lie (though not in the jurist's sense), that is, so far as in me lies I cause that declarations in general find no credit, and hence that all rights founded on contract should lose their force; and this is a wrong which is done to mankind. ...To be *truthful* (honest) in all declarations is therefore a sacred unconditional command of reason, and not to be limited by any expediency.

Not even the "expediency" of saving an innocent human life or preventing a hideous crime? Not even that, was Kant uncompromising. The small vignette on the impermissibility of a benevolent lie has acquired an almost legendary status by now. With its frank, almost oblivious acknowledgment of Constant's reductio, and the brief, rather enigmatic discussion that followed it, it has provided a battlefield for generations of moral philosophers. While Kantians are struggling to explain away the unease felt by most impartial and philosophically uneducated observers over what looks not just as an needlessly rigorous and heartless, but also clearly counter-intuitive conclusion, his (mostly, but not exclusively consequentialist) critics see in his pronouncements either a straightforward reductio of his moral theory or at least something very close to it.

In the present paper I will not, however, seek to settle the longstanding dispute between Kantians and their consequentialist opponents. My ambitions lie

elsewhere. I propose to focus on a single controversial idea, admittedly just one of many,<sup>1</sup> about when it is legitimate to hold agent responsible or liable for harmful consequences of his or her actions. Kant seems to suggest that as long as one fulfills his or her moral duty, he or she cannot be blamed for any potential harm that his or her duty-fulfilling action may bring about, no matter how foreseeable these

may (have) be(en).<sup>2</sup> Call this the Duty-Absolves-Thesis or DA for short.

I will begin by stating the DA-thesis in a more precise form. I will then go on to assess, one by one, several possible justifications for it. I will show all attempts unsuccessful. I will conclude that besides being counter-intuitive, the thesis also lacks firm theoretical grounding and should therefore be rejected as (part of) an account of the so-called outcome moral responsibility.

Kant's Duty-Absolves-thesis (DA)

(DA): One can be held morally and legally responsible for the bad (harmful) consequences of his/her actions, only if these are wrong in the first place, that is if their commission or omission constitutes a violation of one of agent's moral or legal duties. No one, though, can be held responsible for the bad consequences of his or her obligatory/dutiful acts.

Strictly speaking, DA is a conjunction of two logically independent theses: (a) that you are not responsible for any bad consequences of your duty-fulfilling conduct whatsoever (Duty Absolves or DA proper), and (b) that you are responsible for all the bad consequences of your duty-violating conduct ("wrongness implicates"). I will treat DA in a package, as a single thesis, so some of the arguments I will advance will challenge DA proper (i.e. try to show that even though the agent did his moral duty, he may still be responsible for the harm he thereby caused), while others will be meant to disprove the "wrongness implicates" part of it (i.e. draw attention to cases where it would be unfair or inappropriate to hold me responsible for some harm, even though this harm clearly resulted from my wrongdoing). I am quite aware that the two don't stand and fall together. But since they are both motivated by the same or similar deontological considerations, undermining just one of them will unavoidably have implications for the other as well.

One further remark. By subscribing to DA, Kant adds another anti-consequentialist twist to his theory. Not only are wrong actions wrong irrespective of their bad consequences and despite their poten-

<sup>1</sup> Despite its brevity, the short essay abounds by controversial declarations. Here is just a selection of Kant's most counter-intuitive claims:

(i) It would be impermissible/wrong to lie to the Inquiring Murderer, even if this were the only way of saving my friend's life because

(ii) telling a lie to the Inquiring Murderer constitutes a violation of our duty of veracity, which is both an absolute legal duty (to humanity) and a perfect moral duty (to ourselves).

(iii) By telling the truth I may have, as it turns out, harmed my friend, but I did him no injustice.

(iv) As a matter of fact, I didn't even harm him – for harm was caused by accident (*casus*), not by something I did.

(v) Even if I have harmed him, I cannot be held responsible for the harm he suffered, namely his violent and untimely death, for my decision to tell the truth was not really free; suppose I cannot avoid answering the Inquiring Murderer's question, and suppose my choice is restricted to two options only, telling the truth or lying, then given that I am unconditionally bound by the duty of veracity, I am not really free to choose and do otherwise.

(v) Harmfulness is merely an accidental feature of truth-telling; given that in the case under consideration the harmful effects of telling the truth escape our control, no one can blame/impute them on me – bad luck is responsible for them, not me.

(vi) As for the so-called benevolent/charitable lie, almost anything (the slightest coincidence) can render it criminal; but if so, its legal status is a matter of pure accident; whatever only accidentally complies with the law, however, is justifiably considered unjust (from the point of view of external laws).

(vii) If my lie, my violation of the universal and absolute duty of veracity, doesn't prevent the murder, but, due to some perverse and unlikely twist of events/reversal of fortune, actually helps the Inquiring Murderer hunt down and kill my friend, then I am going to be justly accused as the cause of his death, even if I played my part in the Murderer's plan against my best intentions and knowledge. Whereas if I tell the truth and, as a foreseeable consequence of that, the Murderer locates and kills my friend, I bear neither legal nor moral responsibility for his death.

(viii) Whoever lies, whether from noble/admirable motives or not, must therefore bear responsibility for the bad consequences of his actions and omissions and accept the punishment imposed by the court.

<sup>2</sup> Or in the IM case: no one can blame me for the violent and untimely death of my friend, if all I did was fulfill my moral duty of veracity by telling the truth about his whereabouts to the inquiring murderer, even if it was pretty clear to me that the latter will use the information that I provided to find and kill my innocent friend, that is, even though by speaking the truth I consciously (but not willingly) made myself an accomplice in a violent crime.

tially good consequences, and right actions right regardless of their good effects and despite their potentially disastrous consequences, but also if a right action actually produces the most disastrous consequences, the agent is not to blame for any of them, not even for those he clearly foresaw and could have effectively avoided. Similarly, if a wrong action brings about some great good, the agent deserves no credit or praise for it either. My point will be that there is no theoretical need for such a correction – judgments about responsibility (or hypological judgments, as Zimmerman (1993) prefers to label them) should be governed by principles that are neither clearly consequentialist nor clearly deontological in their nature.

### Preliminary stage-setting

Let me start with a concession to Kant. DA presupposes a strict separation between the right and the good. The morally obligatory course of action in S need not be the one which brings about most good (i.e. is best, or most beneficial, from some impartial point of view) in S. In fact, it doesn't need to bring about any good whatsoever, or prevent any evil. On this picture, morality can occasionally, albeit rarely, demand that we take the most rather than the least harmful course of action open to us.

I will not dwell on this, however. Let's simply assume, for the sake of the argument, that often the only way to accomplish some great benefits, or prevent some great harms, is by violating one or many of our moral obligations or constraints. This still doesn't completely remove the initial oddity of DA: Why believe that a good, virtuous will gets you off the responsibility hook? Or, to put the puzzle in the words of Paul Kutz: "A virtuous will seems to insulate from responsibility, while only the vicious will implicates. What Primo Levi called "the grey zone" of collaboration with evil is artificially transformed into the sharp light of duty and the darkness of its violation" (Kutz 2000; 3).

Secondly, I will, as far as I can, try to keep two things apart: my assessment of Kant's judgment in the particular case under consideration (namely the Inquiring Murderer case) on the one hand and my verdict on his general account of moral responsibility for resultant harm (the truth of the DA) on the other. It is one thing to ask whether I could be, as Kant contends, justly accused as the cause of my friend's death if I lied, but not also if I told the truth to the Inquiring Murderer, and quite another to try to answer a more general, principled question about whether fulfillment of duty really absolves one from

all responsibility for the harm that the dutiful action might have caused to an innocent victim.

It is easy to see that a negative answer to the first question will not necessarily settle the second. For I might come to the (correct) conclusion that I should be held responsible for my friend's death in the IM case because by bluntly telling the harmful truth to the IM I will have failed to discharge another equally pressing moral duty, the duty to assist people in dire need.<sup>3</sup> Once we assume that one and the same action-token, my truthful reply to the IM's inquiry, can at the same time instantiate two moral properties, "violation of (one) duty" and "fulfillment of (another) duty," the comparative moral advantage of truth-telling over lying disappears and the liar and the honest guy are set morally on a par with respect to their responsibility for the tragic outcome. So a friend of Kant's can avoid his counterintuitive conclusion in the IM case without also giving up his further idea that we can be held responsible, morally and legally, for the harmful consequences of our actions and omissions only insofar as these constitute violations of some of our pre-existent moral duties. All she needs to do is acknowledge that the conflict between the duties which the agent faces in the IM case cannot be resolved without sacrificing one of the conflicting duties, either the duty of veracity or the duty of assistance. And that for this reason the agent will be morally responsible (in some yet to be defined sense) for her friend's death

<sup>3</sup> An appeal to the duty of assistance is probably the most common Kantian solution to the IM case. Such a solution comes with a cost, however. For in the case of a genuine conflict between duties the only way to fulfill one binding duty is by violating another, equally binding duty. But this would imply that one and the same harmful action-token can be both dutiful and wrong, and it would seem to follow from DA that in such cases we both are and are not responsible for the harm we may cause. So the things look even grimmer for Kant. Kant himself, of course, does not allow for a genuine conflict between duties and is therefore, at least on the face of it, immune against this sort of objection. I say "on the face of it," because his position comes at a considerable cost. As long as he denies the possibility of a genuine conflict between Rossian all-things considered, resultant duties, I have no quarrel with his position. With respect to Rossian *prima facie* duties, on the other hand, his options are fairly restricted; to deny the very possibility of a genuine clash between two or more *prima facie* duties is both implausible and unmotivated; if, on the other hand, he wants to allow for such conflicts, then he should also make room for some moral residue in the form of due compensation, apology or regret/remorse. Yet in his discussion on the Inquiring Murderer case Kant is suspiciously silent on this.

no matter what she does – for she will have violated some duty or another in any case, no matter which course of action she happens to choose. In the light of this, I will do my best to advance arguments questioning the truth of the DA directly, rather than undermining Kant’s particular judgment in the IM case. But some of the arguments presented will support a different conclusion, namely that the truth-teller cannot escape responsibility for her friend’s death by simply doing her duty. Since their force will not depend on a further premise that by fulfilling my duty of veracity I necessarily violate another duty of mine, that of assistance, these arguments will, however, effectively establish the same general conclusion – that since duty does not always absolve of responsibility (it certainly doesn’t in the IM case), the DA-thesis (at least in its present, non-qualified form) must be rejected.

### A *prima facie* case against the DA

According to Kant, then, at least legally, but for all we can say, morally as well, we are not equally responsible for all the (unintended) consequences of our actions; in the case of actions, harmful to others, we bear no responsibility whatsoever for any harm that may result from them, including the most predictable one, as long as our action (or omission) counts as a fulfillment of some (legal or moral) duty of ours (coincidence is to be blamed for harm, not us); whereas in the case of violations of duty the agent takes the blame for all, even the most unpredictable, indirect and distant consequences of his or her actions and omissions.

Kant’s replacement of certain widely, both intuitively and theoretically, accepted distinctions (such as whether harm was intended or not, directly caused or merely not prevented, actively pursued or merely allowed, and so on) with a new criterion (whether harm is a result of a fulfillment or a violation of our moral and/or legal duties) is somewhat surprising. The revision he proposes is, if nothing else, hard to reconcile with two firm and relatively common intuitions: that responsibility should (a) track control and (b) reflect causal contribution. The first idea is common enough – the more control (or causal influence) we have over something the more responsible we are for it; the more we lack control – as is certainly the case with distant and unforeseen (unpredictable) consequences of our actions – the less responsibility we (can be held to) bear for it. Similarly, we are used to treat agent’s responsibility for certain harm as a function of her causal contribu-

tion to that harm<sup>4</sup> – the more intervening events there are in a causal chain leading from my decision or action to the occurrence of harm, and the higher the proportion of other people’s voluntary acts that mediate between them, the less accountable I am for the harm. The sole focus on the moral status of action that Kant proposes, in contrast, discounts both types of concerns as morally irrelevant.

Besides, AD looks theoretically unmotivated and insofar arbitrary. Admittedly, Kant does explicitly allow for the possibility that the action with the most harmful consequences, the most dreadful outcome, is the only right and hence obligatory one in the given situation, while the action with the most generally beneficial effects is the wrong one to do. Still, the idea of absolving the agent of all the responsibility for the bad consequences of her right actions and withholding her any credit for the good consequences of her wrong actions will strike many as tendentious, if not outright ridiculous. For Kant could have insisted that among the three basic options, namely doing right, preventing evil and doing good, doing the right thing enjoys some sort of lexical priority over the other two, but nevertheless allow that on those (rare?) occasions when doing right not only interferes with doing good, but is actively opposed to it (i.e. when my performing the morally prescribed or commanded action is not only not to anyone’s benefit but is actually harmful to some), I may well owe an apology or some sort of compensation to at least those “victims” of my choice to do the right thing who have done nothing to deserve the harm.

Suppose, then, that being moral, doing what morality demands, or doing the right thing, as the Inquiring Murderer case painfully reminds us, sometimes carries not just high personal, but also moral cost.<sup>5</sup> That occasionally, we are going to either di-

<sup>4</sup> For a persuasive defence of a causality requirement for both moral responsibility and accomplice liability see Moore (1999) and Moore (2007).

<sup>5</sup> It is often overlooked that consequentialism and deontologism are on a par in this respect. Just as there is collateral moral damage when we focus on maximizing the good, there is such a thing as a collateral moral damage of committing oneself hundred percent to doing the right thing. What’s different is the particular form that collateral moral damage takes in each case. If we decide to pursue the consequentialist agenda, we will be occasionally forced to commit gross injustice (think of the paradigmatic objection against consequentialism, the permissibility of sentencing an innocent person to jail or death). If, on the other hand, we are committed to doing the right thing, we will sometimes do or allow serious undeserved harm to innocent people (which might also be plausibly

rectly harm some innocent people, or at least fail to prevent them from being seriously harmed if we decide to meet our moral obligations no matter the consequences. Suppose, in other words, that there is such a thing as “moral collateral damage.” If so, then one of the questions we need to answer is this: Are innocent victims of “collateral moral damage” owed an apology or compensation in case they are directly or indirectly harmed because someone was determined to do the right thing, no matter the consequences in human pain and suffering? And if so, who is supposed to provide such compensation?

Proponents of DA are going to be forced to deny the victims of collateral moral damage any valid ground for complaint, or at least any valid moral claim against the moral agent himself. The proper addressee for their claim, they will have to insist, is the evil-doer (i.e. the murderer), not the good, law-abiding guy (i.e. the truth-teller). Yet there is something morally unsettling about this kind of hand-waving and hand-washing. The conscientious agent was, after all, complicit in a criminal act, or at least what he did freely and in full awareness of likely consequences, was not just instrumental, but crucial to the latter's success, so how can he simply walk away morally pure and clean, not a single stain on his moral robe?<sup>6</sup>

### Ways of getting Kant off the hook

Pointing to the initial oddity or awkwardness of DA is not the same as providing a conclusive proof of its falsity, of course. Perhaps DA can be supported by not plainly evident, but nevertheless powerful considerations. The following list contains those among them that I find most promising.

---

considered a form of injustice). Any moral theory, if it is to be complete, should therefore provide some clue about how to deal with such cases, a sort of basic account of corrective or compensatory justice. Some of the objections I will raise in the paper amount to a charge that Kant's account of moral responsibility for harm not only doesn't properly address this issue, but fails short of providing an even *prima facie* plausible starting point for an account of corrective justice for victims of collateral moral damage.

<sup>6</sup> He has, after all, by telling the deadly truth, if not made himself an accomplice in crime, then at least interfered with his friend's core protected interest in avoiding personal injury and death, and for this harm alone the tort law shall find him liable and impose a duty of compensation on him. I discuss the idea of harm that calls for compensation, or of injustice that needs to be corrected, in more detail later on.

(i) DA is just plain false – fortunately, it is not the view Kant himself actually held or was committed to;

(ii) the charge of DA's oddity is misdirected – DA contradicts neither certain well-established moral principles nor our firm moral intuitions;

(iii) DA offers an account of legal (either criminal or tort) liability, not of moral responsibility; with its scope properly constrained, it becomes much more plausible;

(iv) DA should be understood as a thesis about the conditions under which it is legitimate to impose remedial responsibility on the agent, or expect her to fulfill the demands of compensatory or corrective justice;

(v) DA is very restrictive indeed, but for a good theoretical reason – it is meant to rule out the damaging effects of luck on our moral assessment of agents and/or their actions.

Kant himself suggested three other possible defense strategies which I will take up in the Appendix. That

(vi) what accounts for the difference, in moral responsibility for the outcomes of our actions, between cases of dutiful and wrongful actions is the fact that only the latter are genuinely free – and since responsibility entails freedom of choice, we cannot be responsible either (directly) for our dutiful actions or (indirectly) for their contingent harmful effects;

(vii) I am not responsible for my friend's predicament for the simple fact that he was harmed by accident and not by me;

(viii) even though I clearly harm my friend if I tell to the murderer the truth about his whereabouts, I don't thereby wrong him, nor do I do him any injustice – for there is no such thing as a (claim-)right of people that others lie for their benefit.

### Let's discuss these suggestions one by one.

(i) Evidently false, but not truly Kant's

The first defense strategy is to deny that Kant ever subscribed to such a view. The problem with this strategy is that it is not difficult to find other passages in Kant's works which express pretty much the same idea as DA.

Kant, first of all, implicitly endorses DA, or something very close to it, in the following paragraph in the essay on the right to a benevolent lie:

This benevolent lie however, may by accident (*casus*) become punishable even by civil laws; and that which escapes liability to punishment only by accident may be condemned as a wrong even by external laws. For instance, if you by a

lie have hindered a man who is even now planning a murder, you are legally responsible for all the consequences. But if you have strictly adhered to the truth, public justice can find no fault with you, be the unforeseen consequence what it may. It is possible that whilst you have honestly answered Yes to the murderer's question, whether his intended victim is in the house, the latter may have gone out unobserved, and so not have come in the way of the murderer, and the deed therefore have not been done; whereas, if you lied and said he was not in the house, and he had really gone out (though unknown to you) so that the murderer met him as he went, and executed his purpose on him, then you might with justice be accused as the cause of his death. For, if you had spoken the truth as well as you knew it, perhaps the murderer while seeking for his enemy in the house might have been caught by neighbors coming up and the deed been prevented. Whoever then tells a lie, however good his intentions may be, must answer for the consequences of it, even before the civil tribunal, and must pay the penalty for them, however unforeseen they may have been; because truthfulness is a duty that must be regarded as the basis of all duties founded on contract, the laws of which would be rendered uncertain and useless if even the least exception to them were admitted. To be truthful (honest) in all deliberations, therefore, is a sacred and absolutely commanding decree of reason, limited by no expediency (Kant 1797).

Is this an unfortunate, one-time slip of tongue? For skeptics, here is another telling quotation, this time from his Lectures on Ethics:

If we do either more or less than is required of us we can be held accountable for the consequences, but not otherwise – not if we do only what is required, neither more nor less. If all the good that we do is just what is required and no more, the consequences of our actions cannot be adjudged to our credit.... At the same time, if we do just what we ought and no less, any untoward consequences are in no way our fault and cannot be debited to us; we cannot be held responsible for them.<sup>7</sup>

Kant is admirably clear here, I think (which is not always the case). He leaves no doubt – as long as we do exactly what is required (i.e. our duty), we cannot be held responsible for the consequences of our actions. Outcome responsibility only comes with doing either less (i.e. violating one's duties) or more (i.e. going beyond the call of duty) than one is required to do. The "he didn't mean it" strategy is therefore hopeless and we better look around for a more promising one.

(ii) Only apparently unorthodox and counter-intuitive

We can divide this strategy into two separate, independent lines of defense: to take the first is to show that by endorsing the DA we are not forced to give up any well-supported moral principles or doctrines; the second is to accommodate the DA to fit our core intuitive judgments.

(a) doesn't contradict any well-established moral theory or principle

The most prominent moral principle which the DA seems to fly in the face of, is the principle, or doctrine of double effect (DDE). DDE is a theory about when we may do something that has both good (beneficial) and bad (harmful) consequences. It prohibits causing harm intentionally (as a means to an end or as an end in itself), but as long as we either cause harm as a side-effect of our pursuit of some other, morally permissible goal(s), or merely allow it to occur, the DDE gives us green light to proceed. At first sight, then, DDE and DA single out different sets of morally relevant properties which renders them theoretical rivals, not companions. And insofar as DDE is a well-entrenched doctrine, incompatibility with it is certainly bad news.

A closer look, however, reveals a fatal flaw in the above reasoning. DDE is an account of what determines the *moral status* of a specific subset of right and wrong actions, namely those with a double effect. As such, it is silent on a related, but separate issue of moral responsibility for the bad consequences of our actions, be they intended or not. It doesn't even pretend to answer the question of when, and to what extent, the agent is morally responsible for the harm resulting from her action. DA, on the other hand, is not a thesis about the features in virtue of which right actions are right and wrong actions wrong. As an account of the conditions under which the agent is morally responsible

---

dutiful actions, but the following passage makes one wonder: "On the other hand, *if we do either more or less than we need*, the consequences can be imputed to us, either as merit or demerit" (Kant (1963); 59, my emphasis).

---

<sup>7</sup> What is left unclear is what stance Kant took on the issue of responsibility for the consequences of our supererogatory actions. Hill (2000) treats them on a par with

for resulting harm, it doesn't even seem to have any direct implications for the moral status of actions. So given that DA is about something else, namely moral responsibility for harmful consequences, and not their impact on the moral status of the action, which is DDE's main concern, the two cannot really contradict each other.

(b) DA intuitively vindicated after all?

Perhaps DA is not so much at odds with the mainstream moral theory as it is counter-intuitive. Hauser and his colleagues (Hauser et al 2007) have collected a mountain of empirical evidence indicating that people tend to judge the variants of Trolley cases in accordance with the Doctrine of Double Effect (even though most of them are incapable of formulating it). Even assuming that there is, for reasons stated above, no direct clash between the well-documented Trolley case intuitions and the implications of the DA, the latter still seems to largely depart from our common-sense, pre-theoretical view on moral responsibility. For most people apparently believe that we are morally responsible for all and only intended and easily foreseeable consequences of their voluntary actions (and omissions?) and nothing else (Kekes 2002).

There is nothing arbitrary about this principle. A rationale for it is deceptively simple. If I decide to do or omit something when nothing forces or prevents me, if I understand both the decision and the surrounding circumstances, and if a normally intelligent person could be expected to see that the action/omission is likely to bring about certain specific results, then (and only then) it is justified to praise or blame, reward or punish, approve or condemn me for my action and its consequences. DA, in contrast, absolves the agent of all the responsibility for the consequences of his right actions and holds him accountable for all the consequences of his wrong actions. One of them, common sense or DA, must give way.

Is DA then counter-intuitive after all? Such a verdict would be premature. A proponent of DA can still hope to find intuitive support for it in a different body of empirical research. Fortunately, some empirical findings (Nelkin 2007) do seem to detect the kind of bias (or asymmetry) in people's attributions of responsibility which promises to at least explain, if not justify, the moral distinctions Kant makes.

I have in mind two such phenomena. The first is the so-called Intention and Action Asymmetry (IAA) and the second the so-called Side-Effect Asymmetry (SEA). IAA refers to our negative bias

towards bad or evil intention.<sup>8</sup> SEA, in contrast, is a label for our negative bias towards bad foreseen side-effects.<sup>9</sup> In a nutshell: as scientists have discov-

<sup>8</sup> Malle and Bennett ran a series of experiments to test the hypothesis that there is a higher degree of blameworthiness for bad intentions relative to bad actions than praiseworthiness for good intentions relative to good actions. They invited subjects to judge how responsible (deserving or blameworthy) agents were for each of the following actions or intentions.

Positive pair:

|                                                   |       |
|---------------------------------------------------|-------|
| [action] helped a neighbor fix a roof             | (10+) |
| [intention] intends to help a neighbor fix a roof | (3+)  |

Negative pair:

|                                                           |       |
|-----------------------------------------------------------|-------|
| [action] sold cocaine to his teenage cousin               | (10-) |
| [intention] intends to sell cocaine to his teenage cousin | (6-)  |

It turns out that in general intentions get less praise and blame than their corresponding actions, but there is twice as much discounting of praise when it comes to intentions as discounting of blame. For example, forming the intention to help a neighbor fix his roof engenders half as much praise, relative to actually helping him, as the blame engendered by forming the intention to sell cocaine, relative to actually selling it.

<sup>9</sup> The Side Effect Asymmetry was observed in judgments of responsibility for foreseen side-effects. It turns out that people judge agents as having a high degree of blameworthiness for bad unforeseen side effects, and offer hardly any praise for good unforeseen side effects. Joshua Knobe gave subjects two vignettes that confirmed this hypothesis:

Harm Condition

The vice-president of a company went to the chairman of the board and said, "We are thinking of starting a new program. It will help us increase profits, but it will also harm the environment."

The chairman of the board answered, "I don't care at all about harming the environment. I just want to make as much profit as I can. Let's start the new program."

They started the new program. Sure enough, the environment was harmed.

Help Condition

The vice-president of a company went to the chairman of the board and said, "We are thinking of starting a new program. It will help us increase profits, and it will also help the environment." The chairman of the board answered, "I don't care at all about helping the environment. I just want to make as much profit as I can. Let's start the new program."

They started the new program. Sure enough, the environment was helped.

Most subjects said the vice-president deserved blame in the Harm Condition, while few said he deserved praise in the Help Condition. When it comes to side-effects, then, we have a new kind of asymmetry. Doing good things (as unintended side-effects) does not engender praise; in contrast, doing bad things (as unintended side-effects) does seem to call for sanctions.



ered, we are more likely to blame people for bad intentions than to praise them for their good intentions, and we tend to blame them more readily for bad foreseen side-effects of their actions than praise them for the good foreseen side-effects thereof. In other words, we seem to be rather “generous” in assigning blame to people for their bad intentions and bad unforeseen side-effects of their actions, but quite “mean” when it comes to praising them for their good intentions and good unforeseen side-effects of their actions. No wonder, then, the proponent of DA can triumphatically exclaim, that in the Inquiring Murderer case Kant found the combination of alleged bad intention (vicious will) and bad side-effects (friend’s death), exemplified in the harmful lie, much worse than the combination of a good intention (virtuous will) and bad side-effects (the friend’s death), as exemplified in harmful truth-telling.<sup>10</sup>

Seemingly, then, we are less tolerant of bad intentions (relative to bad actions) and of bad foreseen side-effects. Ideally, our moral reaction would be proportionate to its object. Instead, our moral judgment and the accompanying moral emotions reveal negative bias – bad intentions and bad foreseen side-effects elicit negative emotions that are much stronger than the positive emotions elicited by comparably good intentions and good foreseen side-effects. The disapproval that the observed or assumed bad intention typically invokes in us is disproportionately strong compared to the approval which accompanies the recognition of good intention. Intention and side-effects simply have a much greater impact on our moral assessments of agents and actions in cases of wrongdoing than in cases of duty-fulfillment.

What are we to think about all this? Is it of any direct relevance to the issue under consideration and if so, what implications does it have for the truth or falsity of DA? I suggest the answer is “None.” The rebuttal can go in two directions. One is to discredit moral intuitions that people have about such hypothetical cases as messy and unreliable. Ordinary moral intuitions have been repeatedly shown vulnerable to the distorting influences of a host of psychological phenomena. One of the best documented is the so-called framing effect, the impact that the choice of words, or the way we describe hypothetical cases, even their order of presentation, have on people’s intuitive judgments. Moral intuitions are malleable and as such easily manipulated. Take the example that was presented to the subjects in one of

the above experiments. We could, it seems, explain the observed difference, in people’s judgment, between the bad intention relative to bad action on the one hand and the good intention relative to good action on the other, by relative triviality of the good deed and its corresponding intention (helping a neighbor fix the roof) compared to the seriousness of the bad deed and the corresponding intention (selling cocaine to your teenage cousin). If this speculation is basically correct, we could easily bring the score for the positive action/intention pair closer to the one assigned to the negative action/intention pair by simply changing the latter to “jumping in the river and saving a child from drowning.”<sup>11</sup> This is just one example of “the noise effect,” the difficulty of identifying, among many, that very feature (or features) of the situation which shapes, often unconsciously, people’s moral judgment. For these and other reasons, we should be careful not to take intuitions, reported or measured, at their face value and draw immediate conclusions about the validity of general moral principles (such as the DA) which they allegedly vindicate.

The alternative approach is more conciliatory. We begin by crediting the intuitions of ordinary people with some epistemic or evidential value and then look for good pragmatic, social, educational, and similar reasons that could perhaps help explain the observed asymmetry. This approach is more promising on the conception of responsibility-attribution as a social practice, for as such it may well serve other, non-moral purposes, or pursue other, non-moral goals. Perhaps there is a rationale (although I myself fail to see it) for wanting to encourage good deeds in people more than merely good intentions, and to discourage people more effectively from forming bad intentions than from performing bad actions. From a purely theoretical point of view, the assessment of the truth of such assertions, however, such a differing attitude to good and bad intentions for, and outcomes of, our actions looks arbitrary and unjustified.

The appeal to psychological phenomena like IAA and SEA in settling normative disputes, even if legitimate, cannot be straightforward, then. My main worry with this line of defense, however, lies elsewhere. It stems from the observation that the distinctions that people rely upon in their moral judgment about Help and Harm, and about Roof and Cocaine, don’t easily translate over to the IM case. Here is why. Kant, we saw, thought that if we lie, i.e. violate our duty and our friend is killed as a result, we are to blame for his death, whereas if we tell the truth, i.e.

<sup>10</sup> And the same no doubt goes for the combination of a bad intention and good side-effects in the life-saving lie).

<sup>11</sup> I owe this suggestion to my colleague Nenad Mišćević.

fulfill our duty, and our friend is killed as a result of this we don't deserve any blame for his tragic death. In the original IM case, then, the alternative courses of action differ in their respective moral status (telling the truth is right and lying is wrong), but share their side-effects – in Truth, the IM finds our friend where we said, truthfully, he would be; in Lie, the IM finds our friend where we said, untruthfully, that he went – in both cases, he is killed on the spot. In contrast, in Help and Harm, both programs are morally neutral and to that extent morally equivalent, so the entire moral difference between them, if there is any, must come from the impact, good or bad, that they have on the environment.

So what we would need to ask, instead, is how deserving the agent is for the side-effects of his actions in the following two cases:

You tell the truth for the sake of doing your duty, no matter the consequences (which renders your friend's fate/destiny a side-effect of your telling the truth) and in one case, (a) your friend dies, while in the other (b) your friend survives (but only thanks to a fortunate twist of events).

What Kant would have to say, I suppose, is that if you are not deserving of any blame when your friend dies, you are also not deserving of any praise when he lives. In contrast to this, the observed side-effect asymmetry implies that you deserve blame for your friend's death (a bad side-effect), but no or very little praise for his unexpected survival (a good side-effect). Admittedly, the problem with the above example is that the unintended side-effects are not equally foreseeable in (a) and (b). Despite this, the asymmetry intuition threatens to undermine Kant's conclusion. For whatever he may want to say about our moral desert (or lack thereof) in case where our friend magically survives, common morality should, in the line of SEA, judge us blameworthy for the loss of our friend's life both in Truth and Lie.<sup>12</sup>

All this makes one wonder whether asymmetry findings are applicable to the IM case at all, either directly or indirectly. Before we can draw any conclusions about how intuitive or counter-intuitive Kant's judgments are we first need to compare actions which differ, morally, in either agent's intentions, or in their actions' unintended, side-effects.

An example of the first would be the pair deadly Truth vs. deadly Lie. If ordinary people were to acquit the truth-teller, but condemn the liar in such circumstances, we could track the difference in their

moral judgment down to the moral difference in agents' respective intentions – the first was acting on a good intention, the second on a bad one. There are two major problems with this suggestion, however. The first is that most people's intuitions don't really support Kant's conclusion – people either find both truth-teller and liar equally guilty or, what is more common, tend to condemn the truth-teller more than the liar. The second is that very few people are willing to join Kant in judging the benevolent liar's intention to deceive the would-be-murderer bad or wrong. Whether there is a good reason for their reluctance or not,<sup>13</sup> it effectively renders truth-telling and lying morally indistinguishable (at least from the point of view of intention) and the findings of either IAA or SEA insofar inapplicable.

Do comparisons of the second kind, between an action which produces a good side-effect and the one that produces a bad side-effect, lend themselves any better for the purpose of assessing DA from the point of view of IAA and SEA? The following two actions differ, morally, (among other things) in their side-effects:

I tell the truth (with the intention of fulfilling my duty) and my friend lives  
vs.

I lie (with the intention of sending the IM on the wrong track) and my friend dies.

Here the consequences are clearly not intended in either of the two cases – in the first, my friend's survival is a side-effect of my duty-fulfillment, while in the second, my friend's death is not just not the intended outcome, it is the opposite of it. So if the SEA were to support not only the DA as a general principle of responsibility attribution, but also Kant's particular judgment in the IM case, people should find the truth-teller only mildly deserving for his friend's survival in the first case and the liar

<sup>12</sup> On the assumption, of course, that if the agent deserves blame for the bad *unforeseen* side-effects of her action, then she must deserve at least equal, if not even more, blame for their bad *foreseen* side-effects.

<sup>13</sup> This will depend on how plausible we find the following principle: that for every intentional action A and intention I, if action A is morally wrong, then intention I must be morally wrong as well. (That the intention of a morally wrong action is itself necessarily wrong.) I myself am rather sceptical. Whatever initial appeal this principle might have, the abundance of examples of well-motivated duty-violations will raise serious doubts about its truth. We could perhaps try to save the above principle by transforming it into something like "For every intentional action A and intention I, if A is morally wrong, then I must be morally wrong as well – provided that A is intended, as a means to an end or as an end in itself, under that very, i.e. moral, description." That, however, would rob it of all interesting content by bringing it dangerously close to the empty (tautological?) "It is wrong to intend to do, and to intentionally do, that which you have every reason to believe is wrong."

highly blameworthy for his friend's death in the second. This prediction, however, fails to materialize when put to test, since most people<sup>14</sup> want to exonerate the liar of any moral responsibility for his friend's death in Deadly Lie and deny the truth-teller any desert in the surprisingly good outcome in Beneficial Truth.

But then it is hard to see how we could make the negative bias, as disclosed in thought experiments on SEA and IAA, bear, either directly or indirectly, on our assessment of either the truth of Kant's verdict on the IM case or, even more importantly, of the DA.

(iii) Meant to account not for moral, but legal responsibility

The legal terminology that Kant employs in the text could tempt one to a somewhat different line of defense. Perhaps what Kant wanted to offer was an account of legal and not moral responsibility for harm. Is this a promising way out of the trap for Kant? Here are some reasons for thinking that switching from moral to legal responsibility won't help save either Kant or the DA-thesis.

The first comes from the context of use. Kant is trying to substantiate his differential moral attitude towards the law-abiding truth-teller and the benevolent liar with an appeal to responsibility that we would want to see imposed on each of them in case of a potentially disastrous outcome of their actions. Would such an appeal really carry much moral weight, if what he invoked were more or less contingent provisions concerning legal responsibility, as found in the existing criminal or tort law? I doubt it.

Secondly, even if his controversial remarks were about legal and not moral responsibility, the implicit account can be naturally extended to the latter.<sup>15</sup> Kant himself draws no sharp lines between the moral and the legal realm. According to his famous formulation an action is legal as long as it conforms with the moral law, and moral if conformity to the latter is properly secured – that is, if the action originates in, or is motivated or brought about by, the sense of duty or respect for the moral law. Admittedly, he does divide his theory of obligation into two parts, *Rechtslehre* and *Tugendlehre*, but the

differences between the two don't run very deep – legal obligations regulate conduct, thereby constraining our freedom of action, while moral obligations regulate feelings, intentions, desires, attitudes, emotions, aims and so forth, i.e. they constrain our inner freedom.

Thirdly, even if the issue under discussion is the agent's responsibility before the law and those in charge of its implementation rather than before the court of our inner conscience, the law that Kant uses as a standard of judgment here is the morally ideal law, i.e. the law which makes use of the means of coercion to ensure external, behavioral compliance with the demands of morality, and not the more or less contingent positive, existing law. But if this is correct, then we can expect the legal debate pretty much collapse into the moral one.

Fourthly, since neither the truth-teller nor the liar can be charged as a principal with committing a complete crime, the only kind of criminal responsibility they may be assigned is that of an accomplice in crime. Yet for this suggestion to work in Kant's and DA's favor, the liar will have to prove more vulnerable to this charge – an implication that only few will be ready to swallow! In fact, exactly the opposite might be true – that on any plausible account of criminal responsibility and complicity in criminal acts<sup>16</sup> the truth-teller, given that by definition he is aware of the murderer's intentions and can predict that the truth that he is going to tell about his friend's whereabouts will help the murderer carry out his vile plan, will be found guilty of complicity in murder, while the liar who has done his best to spoil the murderer's plan (and failed due to no fault of his own) will be acquitted.

This, however, need not be so. Exactly what relationship the accomplice (both with respect to his state of mind, *mens rea*, and his actions, *actus reus*) must bear to the result prohibited by law (in our case, murder) is a complicated issue. According to a popular idea, you need to 'aid and abet' the other to be guilty as an accomplice in his crime. More precisely, "to be an accomplice, my act must have something to do with why, how, or with what ease the legally prohibited result was brought about by someone else" (Moore 2007; 401). At first, this might look like offering a sufficient ground to implicate the truth-teller in the murder, but it doesn't. For in order to be found guilty of complicity in crime you must be proven to have had a purpose in, not

<sup>14</sup> Admittedly, this estimation is not based on a large representative sample, only on my first- and second-year students' reactions.

<sup>15</sup> This seems to be the point of agreement between the authors so diverse as Thomas Hill jr. (Hill 2000), James Rachels (Rachels 1997; 153) and Alenka Zupančič (Zupančič 1993; 53). Herbert Paton (Paton 1986), on the other hand, simply assumes the whole dispute is about our moral, not legal, duties.

<sup>16</sup> For some accounts of the former see Honore & Hart (1962), Moore (1999) and (2007) and Duff (2006), for an extensive treatment of the latter Kutz (2000).

just the knowledge of, the principal's criminal action.

So given that the truth-teller only has knowledge, but lacks purpose,<sup>17</sup> that he is not acting on any intention that he would share with the IM, I need to qualify my conclusion accordingly. I want to say, then, that the liar and the truth-teller don't really differ, morally, from the point of view of conditions that are individually necessary and jointly sufficient for complicity in crime. So while Kant might be right to insist that the truth-teller should not be treated as an accomplice in murder, he is also wrong insofar as he wants to implicate the liar in it. From the point of view of the conception of accomplice-responsibility which pervades the theory and practice of criminal law, they are legally on a par. But then it is certainly unjustified to find one of them, the liar, criminally liable as an accomplice to murder, while letting the other one, the truth-teller, get away unscathed.

Finally and most importantly, replacing moral with criminal responsibility is not going to help either Kant or the DA, since the idea of sending people to jail for trying to save an innocent victim, who happens to be their friend, by lying, i.e. by breaching their moral duty of veracity, is equally, if not even more, abhorrent than the idea of blaming them for their death. We are forced to conclude, then, that DA hardly fares any better when re-interpreted as an account of criminal responsibility.

(iv) A plausible account of legal liability for harm?

Maybe what Kant had in mind, however, was tort liability for harm and not criminal responsibility (i.e. the notion of liability found in tort and contract law, but not necessarily in criminal law). According to this proposal, Kant's intention was to offer an account of when the law should impose on the agent at least some of the costs of amending the harm that befell someone else (at least partly) because of what the agent did (voluntarily).

Does this interpretation make DA any more plausible? Well, there is at least a *prima facie* case for it. As Moore (1999; 1) points out, in the Anglo-American tort and contract law, "we are each liable only for those harms we have caused by the actions that breach our legal duties." Marina Oshana (Oshana 2007; 75) concurs: "Traditionally, when we ascribe liability for harm we assume the harm resulted from faulty behavior (active or omitted) on the actor's part and that the actor is open to judgment and response on our part for this reason. That is, we assume what might be called contributory

fault on the part of the actor." And so does Judith Thomson (1990; 192): "Under traditional tort law, a plaintiff had to show three things in order to win his case: that he suffered a harm or loss, that an act or omission of the defendant's caused that harm or loss, and that the defendant was at fault in so acting or refraining from acting" (my emphases).

This, however, may be a more or less contingent feature of a prevalent 'economic' conception<sup>18</sup> of liability for harm underlying the existing theory and practice of tort law. According to this conception, the main point of tort law is to achieve an *efficient* (rather than *fair* or *just*) allocation of resources – by giving an incentive to discover the true costs of an enterprise or activity in terms of its harmful effects on others and making sure that the agents pay them, the tort law achieves a more correct pricing of the goods and services produced by that enterprise and thereby assists the market in its basic, allocating role.

Given a close connection between law and morality that Kant insists on, however, a rather different, justice-based (Perry 2001; 59) conception of tort law may be more appropriate. According to this account, the tort law should be designed to best accomplish the goals of (either distributive or corrective) justice. The idea behind this is simple and appealing – we all have primary moral duties not to hurt or harm others; whenever we culpably violate such primary moral duties, we impose on ourselves a secondary moral duty to correct the wrong or injustice we have done. Tort liability rules are nothing but the enforcement of such antecedently existing moral duties of corrective or compensatory justice.

The central idea of tort law, then, is that some sort of compensation (ranging from apology to financial compensation) is owed for harm suffered by another person through our own fault. So in order to show the agent in the IM case liable for the harm that her friend suffered because she supplied the IM with a crucial information about her location, we need to identify some wrong or injustice that the agent thereby committed and that calls for a remedy. If the friend was harmed through no fault of the agent, the latter can't be found liable for harm and Kant will have won this particular stage. The crucial issue, then, is the following: Is it really true that despite telling the truth which I know will almost certainly be instrumental in helping the murderer kill an innocent person, I do no wrong or injustice to the latter which would call for a correction or a remedy? That my behavior, since dutiful, neither gives any ground for my friend's complaint nor violates any of

<sup>17</sup> While the liar lacks both.

<sup>18</sup> I borrow this term from Perry (2001).

her moral or legal rights? “No” is a very tempting reply to this question. But the issue is only apparently simple, so I will return to it with a due attention in the Appendix.

(v) True perhaps of a sub-species of moral responsibility?

The notion of responsibility is notoriously ambiguous. The statement “A is morally responsible for h” may express very different propositions, or give expression to very different ideas. When we say of someone that he or she is responsible for h, we may want to say that

- (a) he or she has played some (minor or major) causal role in the production of h (causal/authorial or A-responsibility);
- (b) that, other things being equal, the benefits and burdens resulting from her action should fall to him or her (outcome or O-responsibility);
- (c) that the agent deserves some credit or blame for his action or omission (and/or the consequences thereof) (character-reflecting or C-responsibility);
- (d) that it is the agent’s duty to put situation h which is somehow morally defective or non-ideal right again (that it is, for example, his remedial responsibility to end someone’s suffering).

Here is another suggestion, then. Once we interpret Kant’s remarks charitably, the way we should, we will have no trouble seeing that they must be about either outcome or remedial responsibility. So when Kant contends of the truthful, dutiful agent that he cannot be held responsible for the death of his friend, what he means is that the friend’s death cannot be ascribed to him either in the sense that he deserves blame for what he did (or the tragic state of affairs he has helped to bring about), or in the sense that his action gives rise to a legal or moral demand, on the part of a sufferer or plaintiff, for compensation, apology, remorse, and so on. The fact that a given harmful action was dutiful (or, according to Hill, supererogatory), in other words, completely absolves the agent from both his outcome and remedial responsibility for the harm that he may have caused by his action.

Taking into account the vast array of distinctive uses of responsibility ascriptions, we can then try to give the DA a more precise meaning. Here is one possibility:

(DA\*) wrongfulness is a necessary condition of agent’s remedial or compensatory responsibility for

the potentially harmful or adverse consequences of his actions.

In other words, of all the harmful actions, only the wrongful ones place the agent under duty to amend the harm, compensate, or apologize to, those harmed by them, and the like. When his actions are in accordance with duty, compensation, apology, remorse, and the like may still be appropriate, or called for, but then the moral force of this demand will stem from some principle other than that of corrective or compensatory justice (say, the principle of charity or benevolence).<sup>19</sup>

I am not convinced that anything of importance is gained by replacing the original DA with DA\*, however. For this strategy to work, the proponent of DA\* would have to deny either that the truth-teller is outcome-responsible for his friend’s death (while insisting that the liar is so responsible) or that outcome-responsibility (ever, or at least in the IM case) entails remedial-responsibility. Let’s address these two issues in reverse order.

Notice, first, that the notions of outcome- and remedial responsibility are both normative. They are both governed by considerations of justice and express our interest in fair/just allocation of benefits and burdens resulting from people’s choices and actions. The only difference is in how the respective judgment proceeds. In determining outcome responsibility, we begin with an agent whose action produces beneficial or harmful consequences and ask which of these consequences can be credited or debited to him. When we try to assign remedial responsibility, on the other hand, we begin with a state of affairs in need of remedy (in IM: “an innocent person was murdered”) and then ask whether there is anyone whose responsibility it is to put that state of affairs right (to amend the harm, or correct the injustice done). So even though the two concepts of responsibility are clearly distinct, they are also very much interdependent.<sup>20</sup>

<sup>19</sup> This reading finds additional support in Hill’s rendition of Kant’s original puzzle: “to ask of an agent whether he is responsible for the bad consequences of his deed is (for Kant) just to ask whether by his deed he has incurred an obligation to try to compensate for damages, rectify the situation, or accept other appropriate costs in response.” (Hill 2000; 160) Hill seems to agree, then, that the responsibility for harm that Kant is interested in is remedial responsibility and that the issue of who should bear it needs to be settled, ultimately, by reference to outcome responsibility.

<sup>20</sup> In Klampfer (2004) I have offered the following argument for such a claim: “Ascriptions of these two types of responsibility are governed by similar considerations of justice, the only difference being in the point at which the

Typically, judgments about R- (the same goes for C-) responsibility are going to be parasitical on judgments about O-responsibility. In order to decide who should sometimes (in the absence of other morally relevant features such as pressing needs or special claims) enjoy benefits or suffer harms we need to know who is responsible for creating them in the first place – who to credit or blame for it.

The most natural candidate for the person whose duty it is to amend harm is (s)he who is O-responsible for it. But does it need to be so? It doesn't. Miller (2007) admits that by default it is the person who is O-responsible for some occurring harm who should bear R-responsibility for it. The two can sometimes come apart, though. We may be forced to attribute R-responsibility to someone on grounds other than his O-responsibility, if the latter is difficult to determine or when we want to encourage people to take particular care in something (strict liability traffic laws are a good example). In the IM case there is no visible difference in O-responsibility between the harmful benevolent liar and the harmful dutiful truth-teller. Yet despite this Kant would have us believe that only the liar, but not also the truth-teller is R-responsible for his friend's death.

Is there any reason for considering the IM case an exception to the rule which makes R-responsibility a more or less direct function of O-responsibility? I can see no such reason. Miller allows for various other grounds of R-responsibility besides O-responsibility: moral (what I call character-reflecting/revealing or C-) responsibility, causal (or A-) responsibility, benefit, capacity and community. (ibid.; 100-104) One of these could, then, at least in principle, explain the difference in Kant's treatment of harmful lies (violations of duty) and harmful truths (dutiful actions).

The problem, however, is that none of the above looks anything like a promising candidate for this

role – the benevolent liar and the principled truth-teller don't really differ (contra Kant) in their moral character, none of them benefits (more) from the situation (than the other), their capacity to amend the harm is equal(ly small), while their membership in the community plays no role whatsoever in the given situation. But if so, then R-responsibility must (at least as the IM case goes) be determined by O-responsibility alone and the two cannot really come apart in the way Kant implies.

What about the other option? Do the prospects of absolving the truth-teller of O-responsibility for his friend's death look any better? I am equally skeptical about that. Since determining truth-teller's O-responsibility for his friend's death would be difficult, if not impossible without a preceding general account of when an agent is O-responsible, which is both tricky and goes beyond the scope of this paper, I propose to ask instead, whether it is plausible to find the liar, but not also the truth-teller, O-responsible. The answer, I believe, is clearly negative. My friend's death in both variants of the Inquiring murderer case, the truthful and the untruthful, is unintended, unforeseeable,<sup>21</sup> and causally linked to what I did only indirectly, through someone else's decision and action – this is what ultimately matters, morally, and what both explains and justifies, contra Kant, our natural tendency to ascribe (or deny) (R-) responsibility to both liar and truth-teller. The same conclusion can also be vindicated in a different way. Moral responsibility, whether for action or omission, is a function of three parameters: ability, opportunity and awareness. When judged by these standards, Kant's position loses any remaining appeal. For assuming that the liar and the truth-teller are on a par with respect to the first two, the ability and the opportunity (they are both capable of doing what is needed to reverse the course of events and they both get a chance to do it), the awareness-condition is clearly only met, or met to a sufficient degree, by the truth-teller – he's got all the reason to believe that her friend is still hiding in the attic, whereas the liar has no reason to believe that her friend has left his initial hiding-place in the meantime and is now on his way to the train station. So if the notion of O-

---

judgment is asked – with O-responsibility we are interested in what would be a just or fair allocation of those benefits and costs that flow from agent's action before they are distributed, whereas with R-responsibility the starting point is the state of an unjust distribution of costs and benefits that urgently needs to be remedied. But since causal considerations play an essential (or even paramount) role in determining desert the pattern of which, in turn, forms the basis for redistribution of costs and benefits, it is safe to assume that the very same causal considerations will play an equally important role in the context where we try to determine the pattern of just or fair distribution for the first time. This common feature of O- and R-responsibility thus explains both their similarity and the role that causation plays in both."

---

<sup>21</sup> In fact, it is not equally unforeseeable. Unless one is ready to assume, along with Kant, that we can never, not even with the smallest degree of certainty, foresee even the most immediate and short-term consequences of any decision or action, it is quite clear that the tragic outcome is much more a result of bad luck, and hence much harder to predict, in *Deadly Lie* than in *Deadly Truth*. Given this, it would be much more natural to ascribe R-responsibility for my friend's death to me, or to ascribe more of it, in *Deadly Truth* than in *Deadly Lie*.

responsibility is, as is commonly assumed, gradual, I will bear some O-responsibility for my friend's death in case I choose to fulfill my obligation by telling the lethal truth. Whereas when I try to save him by lying and the unfortunate train of events renders my utterance true, my O-responsibility for his fate will be minimal or none.

(vi) Theoretical necessity – protecting moral responsibility against the adverse effects of (good or bad) moral luck

Kant was very cautious to rule out, as much as he could, the impact of luck on our moral judgment. That's why he tied the moral value of the agent's character to her motives (rather than her conduct) and the moral value of her actions to her intentions (rather than their more or less contingent outcome). According to Julia Driver (ms), luck was of Kant's central concern: "the Kantian system is actually constructed so as to avoid the impact of moral luck on moral worth. It is a theoretical strength of the Kantian position that it insulates moral worth from luck."<sup>22</sup>

Here are two well-known passages, one from *Groundwork*, where he tries to secure the goodness of the good will against the "unfortunate fate or the niggardly provision of stepmotherly nature," and the other from the *Benevolent lie* pamphlet, where he contrasts doing (in)justice ((*Un*)*Recht*), after it was made immune against the vagaries of luck, with benefiting and harming (*Schaden*), which cannot be done or omitted at will, or ever brought under effective control.

A good will is good not because of what it effects or accomplishes, nor because of its fitness to attain some proposed end; it is good only through its willing, i.e. it is good in itself.... Even if, by some especially unfortunate fate or by the niggardly provision of stepmotherly nature, this will should be wholly lacking in the power to accomplish its purpose; if with the greatest effort it should yet achieve nothing, and only the good will should remain (not, to be sure, as a mere wish but as the summoning of all the means in our power), yet would it, like a jewel, still shine by its own light as something which has its full value in itself. Its usefulness

or fruitlessness can neither augment nor diminish this value. (my emphasis)

And a bit later, building on his previous insight about the threat of luck:

An action done from duty has its moral worth not in the purpose to be attained by it, but in the maxim according to which the action is determined. The moral worth depends, therefore, not on the realization of the object of the action, but merely on the principle of volition according to which the action has been done.

All fine and well. We don't want to see our own moral worth as a person, or the moral worth of our actions determined by luck, whether good or bad. We can even sympathize with Kant's efforts to make the moral status of actions (i.e. our moral duties) as independent as possible of their good or bad effects, of whether they succeed or fail to attain the end pursued in action. But why on earth would he want to protect judgments about moral responsibility *for the consequences of our actions* against the effects of moral luck as well? Why try to extend protection against the damaging effects of luck from the notions of moral worth and moral duties to the notion of moral responsibility for the consequences of one's actions?

Such an (over-)ambitious project could be motivated by something like the following reasoning. No one can be held responsible for that which is beyond his or her control (the so-called controlling condition); the consequences of our actions are beyond our control, i.e. it is ultimately not in our power alone to produce or prevent them; hence we cannot be held responsible for the consequences of our actions, whether good or bad. But then no one can ever deserve praise for what he managed, or blame for what he failed, to bring about.

But if all this is true and control over whether X obtains or not is necessary for the O-responsibility for (the obtaining of) X, then the very notion of outcome-responsibility, or responsibility for the consequences of our actions, is, Kant should have concluded, confused, if not outright inconsistent, and we would do better without it. What we can no longer do, however, is pretend, as Kant seems to, that in contrast to the bad consequences of his dutiful actions (duty-fulfillment), the agent can nevertheless be O-responsible for the bad consequences of his or her wrongful (duty-violating) actions.

This line of argumentation is not really available to Kant, then, I'm afraid. For he denies O-responsibility for harm only in the case of dutiful

<sup>22</sup> For a more detailed discussion on this topic, one that goes beyond the scope and ambitions of the present paper, see Moore (1990). For an excellent account of many subtle distinctions that underlie moral assessment of agents, as well as various types of agent assessment, in Kant see Jokić (2002).

actions.<sup>23</sup> When harm ensues from a violation of duty, the agent bears full responsibility for it, regardless of how remote, indirect, mediated by others' decisions and actions, unintended or unforeseeable it was. But surely the consequences of our wrongful actions are just as elusive and dependent on factors beyond our control as the consequences of dutiful actions? And if so, are they not entitled to the same treatment then? But the whole idea of justifying different O-responsibility ascriptions as a remedy for luck is hopeless, anyway, because the consequences of our actions are always going to be subject to the whims of luck. The advocate of luck will thus need to abolish the very notion of moral responsibility for consequences (if not as a contradiction in terms, then at least as a hopeless mess) and give up our practice of ascribing it.

Here is another possibility. What perhaps disturbed Kant more than anything else was the possibility that whether what we do is right or wrong may ultimately depend on luck, in the sense that if we are lucky, then what we do will turn out to be right, and if we are not, it will turn out to be wrong (Moore (1990)). This is an unfortunate implication, it seems, of those normative, most prominently consequentialist, accounts that treat rightness and wrongness of actions (or what makes actions right or wrong) as a function of the overall goodness and badness of their consequences, since whether my action will produce more good than bad or the other way around is surely not (at least not always) up to me.

So when Kant emphatically denies that whether or not it is right to lie to the Inquiring Murderer about our friend's whereabouts ultimately depends on whether, as a result of our pronouncement our friend is going to live or die, we can, I think, agree with that, even if somewhat reluctantly. For we have accepted the widely deontological framework, within which Kant is reasoning. Whoever shares Kant's deontological intuitions will gladly admit that if lying is wrong in the IM case, then it must be for some reason other than the fact that it will have terrible consequences for your friend. The same goes for telling the truth – whether it is the right, dutiful thing to do in given circumstances or not cannot depend on its outcome, on whether it will harm or benefit our friend, if the latter clearly escapes our control.<sup>24</sup>

<sup>23</sup> And perhaps supererogatory as well. But even if Hill (2000) is right about this, he will need very different, perhaps even pragmatic reasons, to substantiate such an extension of the DA.

<sup>24</sup> Perhaps we have to do with two related worries here instead of one. The first arises from our commitment to something like the control condition on responsibility – if

But the legitimate worry about the moral status of our actions being determined by sheer luck shouldn't be confused with a somewhat different idea that since we can only be (held) morally responsible for what is under our control and the survival or the perishing of our friend is not under our (complete) control, we cannot be morally responsible for his death. For first of all, we have some, arguably quite enough, control over what course the disturbing events are going to take (we can influence it, for it is not causally pre-determined), even if due to the fact that the final outcome will result from decisions and actions of several agents which we cannot fully foresee or calculate, we cannot guarantee it or vouch for it. We may not enjoy a complete control over the course of events (for this we would need to have restricted control over every event on which this particular chain of events is contingent, and no one has such control over anything), but who cares – a restricted control (i.e. the ability to prevent a particular occurrence of the event) will do perfectly fine.

Secondly, we may accept the control condition on responsibility, but nevertheless insist that we can be responsible for our friend's violent death if the piece of true/accurate information that we supplied was instrumental to it. For we can agree that whether I am a good or a bad person for having done X (telling a truth or a lie to the IM) shouldn't depend on the outcome of X (on whether my friend will live or die as a consequence of my telling the truth or my

---

the bad effects are beyond our control, if ultimately they are a matter of chance, then they cannot be imputed on us, or taken to count against us in "the court of morals." The second stems from certain hidden deontological assumptions – if the outcomes of our actions are largely determined by chance, then they are more or less contingent, external features of our actions, but since the right- and the wrong-making properties must be intrinsic, the rightness or wrongness of what we do cannot depend on its actual outcome. The role of chance is thus twofold – if we stress its "out of control" aspect, it jeopardizes our practice of praising and blaming agents for what they did or failed to do and the notion of responsibility that underlies it; if, on the other hand, we stress its "contingency" aspect, then we have no choice but to morally discount the outcomes of our actions and treat them, contrary to our firm intuition or better judgment, as morally irrelevant or impotent. Outcome or resultant luck then seems to have two far-reaching theoretical implications – it forces us to give up the idea of outcome responsibility and to say goodbye to consequentialism. But of course we can also preserve both and reject the control condition on responsibility instead, or the idea that rightness either itself is an intrinsic property of actions, or at least supervenes on such properties.



lying), if the latter is largely or even entirely determined by luck. What we shouldn't simply assume, however, is that if the actual causal chain of events that led to our friend's death was to some degree subject to chance, then we cannot be blame- or praiseworthy, deserving of blame or praise for what befell him. The violent death of our friend may not have been something we intended when we spoke the truth, but it was still predictable. And why would it be unfair to blame agents for the unintended, yet clearly foreseeable consequences of their actions? We do it all the time – we blame smokers for their damaged health, car racers for endangering the life of other participants in traffic, overeaters for their high cholesterol and blood pressure, car manufacturers for their products' lethal defects, and so on and on.

Another possibility worth considering is that Kant's main concern was not so much with the threat that moral luck poses to the moral status of our actions, as with its threat to our (equal) moral worth (as persons). For if our moral worth depends on whether we are a good person or a bad person and this in turn depends on whether we do, or fail to do, the right thing, meet or fail to meet the requirements of morality, then our moral worth is built on a rather shaky and unstable ground. Maybe Kant's main reason for denying agent's responsibility for his friend's violent death in the IM case was that he thought accepting responsibility for harm would entail some sort of recognition of guilt and guilt would put a stain in the agent's so-far impeccable moral record – and all this for having done his duty! If we ascribe responsibility to the truth-teller for his friend's violent death, wouldn't this naturally invite another negative moral evaluation, namely that he must be a morally worse person for his involvement in a violent crime? But there is surely something paradoxical about the idea that meeting your moral obligations can reflect badly on you, that you can (turn out to) be a morally worse person simply for having done your moral duty.

This sort of worry is misplaced, however. It assumes too close a connection between two kinds of agent evaluations: judgments of responsibility ("A is responsible for having done X") and judgments of moral worth ("A is a better/worse person for having done X").<sup>25</sup> The truth, in contrast, is that the two are very much independent from each other. To see this, one only needs to remember a whole range of moral propositions that can be expressed by means of responsibility ascriptions. By saying "A is responsible for X," we sometimes ascribe mere causal or author-

ial responsibility to A ("it was A who did X"). Or we may want to ascribe the so-called outcome responsibility instead – when we say of someone that he or she is responsible for something in this sense what we mean is that, other things being equal, the resulting benefits and burdens should fall to him or her.<sup>26</sup> "A's responsibility for X" may also refer to what we might call "remedial responsibility" – in that case saying that A is responsible for X is equivalent to saying that it is A's job, or obligation, to put something that has gone awry right again (end someone's suffering, for example, or compensate the victim of some form of injustice).

Kant's worry would be legitimate if "A is responsible for X" meant "A deserves praise or blame for having done X." Such a 'meritocratic' understanding, or use, of responsibility ascriptions is surely common enough. We do often invoke the notion of responsibility to express our assessment of agent's moral merit or desert. If "I am (at least partly) responsible for my friend's violent death (in virtue of having told the truth about his whereabouts to the IM that was instrumental in his murder)" could only be interpreted as "I deserve blame for speaking the truth," then this might indeed be difficult, if not impossible, to square with the idea that telling the truth was the only morally right thing to do in the situation. For how can I possibly deserve blame either for doing the right thing or for thereby accidentally causing or allowing harm? We would have to give up, or so it seems, either the idea that it was my duty to tell the truth, or the intuition that I am responsible for my friend's violent death after all.

The problem is that the above "meritocratic" reading of this particular responsibility ascription is not the only one on offer. Alternative, equally plausible interpretations are available under which the above responsibility ascription will come out true and not false. For instance, it is beyond dispute, I think, that I bear some causal or authorial responsibility for my friend's death if my sincere pronouncement led the murderer to his hiding place. Furthermore, on any plausible construal of outcome-responsibility I would also bear some outcome- and, consequently, remedial-responsibility for his tragic death. By this I mean that it would be fair to allocate to me some burden (most of which should be borne by the murderer himself) of remedying what I earlier called moral collateral damage, the harm that (both)

<sup>25</sup> I borrow this distinction from Greco (2006).

<sup>26</sup> This notion is normative (governed by considerations of justice) and gives expression to our interest in fair or just allocation of benefits and burdens resulting from people's choices and actions.

my friend (and his family) suffered as a side-effect of my focusing on moral duty, success, purity and integrity. So even if I'm not blameworthy for my truthfulness, even if I deserve no blame (or punishment) whatsoever for my friend's death, there are other interpretations under which it would still be appropriate to find me responsible for his tragic death.

So what? If Kant was primarily interested in denying the agent's "responsibility" for her friend's death in the IM case in its "meritocratic" sense, why would it matter that the agent may nevertheless be responsible for it in some other sense, one that gives rise to a duty of apology and compensation? This objection has some force. Fortunately, Kant's reasons for denying agent's culpability in IM are relatively weak. Kant is probably right to insist that if we tell the truth because it is our duty to do so, and our friend is killed (partly) as a result of our moral choice, we shouldn't be considered a morally worse person for having contributed causally to her death – though we have certainly proved to be a rather terrible friend! So we also shouldn't think worse of ourselves as moral agents, or see our moral integrity compromised. But is it really so implausible to insist that some agent-regret, some moral residue would be in order? That we should feel bad about choosing to be truthful in the situation where truthfulness and blind adherence to our moral duty came at such a high moral price? An agent who felt no regret over such a painful, even fatal choice would strike us as morally defect, even corrupt. For however we may construct the IM case, whether as a conflict between two equally binding duties or merely as a forced choice between fulfilling a moral duty on the one hand and preventing undeserved harm on the other, the agent deserves deepest sympathies for the moral predicament in which he found himself. But if so, then there is both room and reason for regret, remorse, perhaps even feelings of guilt on her part.<sup>27</sup>

Finally, and perhaps most important, the tragic outcome in Truth doesn't even qualify as an example of bad luck. Kant's IM examples are supposed to be instances of what Thomas Nagel (Nagel 1993) calls "outcome luck." The term refers to those factors outside of agent's control which crucially determine the outcomes of his actions. The natural im-

pression that luck might be playing a similar role in the IM case is mistaken, however.

As noticed before, Kant has built an argument against holding the truth-teller responsible for the death of his friend on the premise that the harm that befell the latter as a result of the blind adherence of the former to truth(fullness) was contingent – that truth-telling is not harmful in itself. From which Kant inferred that the harm of my friend's premature death cannot be ascribed to us, the truth-teller. But isn't the same true of lying? Namely that since harm is also a merely contingent feature of any particular instance of lying,<sup>28</sup> the liar also can't be blamed for harm that the chance or accident brings about behind his back and against his explicit will/intention? But if the effects of truth-telling are no more subject to luck than the effects of lying, then the truth-teller and the liar shouldn't really differ much in terms of their O-responsibility, should they?

But forget this unfortunate complication for a moment. Let's grant to Kant that whatever harm may result from our telling the truth on a particular occasion, or to a particular person, it will be a contingent feature of our action. Still, being contingent (i.e. not necessary) is far from being purely accidental or subject to pure chance. A simple test with the most plausible accounts of luck<sup>29</sup> shows that while the bad effects of lying to the IM may qualify as bad luck, the bad effects of truth-telling which Kant was most eager to insulate from the adverse effects of luck, do not.

Intuitively, the meaning of "luck" is determined by its proximity to the notions like chance, accident, (mis)fortune etc. The most popular definition of luck is in terms of lack of control: "What(ever) is not within our control, is a matter of luck." This, however, is both too vague and too broad. The sunrise, for example, is not within our control (we can do nothing to either bring it about or prevent it), but we wouldn't call it a matter of luck (Pritchard 2006).

Here, then, is a better proposal: "We call something a matter of (good or bad) luck when 1) it is of interest or importance to us, 2) it was not under our control, and 3) we had no reason to expect its occurrence. So I am lucky if I find a twenty-pound note on the street, because it is of interest to me, I did not put it there or arrange its being there in some other

<sup>27</sup> This conclusion shouldn't depend on whether we find agent-regret and guilt appropriate even when the agent did nothing wrong, that is on who we choose to side with on a wider issue of moral dilemmas. For an account of moral dilemmas that is sensitive both to their complex nature and to often contradictory aspects of our moral phenomenology, see Williams (1981).

<sup>28</sup> That much is at least true of the harm to our friend in the IM case, even if we accept that by lying we always and necessarily commit injustice, if to no one else than to humanity.

<sup>29</sup> The most popular accounts of luck are those in terms of what (a) is beyond agent's control, (b) is unreasonable to expect and (c) occurs in the actual world, but fails to occur in the relevant subset of nearest possible worlds.

way, and I did not expect to find it at all, because usually when I walk along the street there are no banknotes to be seen at all” (Schinkel 2008). This gets us a bit closer, but is still too imprecise for our purpose. After all, what it is reasonable to expect, often depends on contextual parameters that have more to do with the state of the agent’s mind (and what he knows about the world) than with the nature of the given event.<sup>30</sup>

Driver (ms) and Pritchard (2006) propose and defend an alternative, modal account of the nature of luck. According to this account, “a lucky event is an event that occurs in the actual world but does not occur in most of the nearest possible worlds to the actual world where the relevant initial conditions for that event are the same as in the actual world.”

Let’s apply this standard to the IM case, then. The decision box below presents four possible outcomes given two basic courses of action available to us, and two sets of circumstances which are going to determine the outcomes of our choices.

Assuming the IM will trust my words

| Action                  | Circumstances              | My friend remains hidden in the attic | My friend runs to the train station |
|-------------------------|----------------------------|---------------------------------------|-------------------------------------|
| I am truthful (TRUTH)   | (O1) My friend is murdered | (O2) My friend survives               |                                     |
| I am not truthful (LIE) | (O3) My friend survives    | (O4) My friend is murdered            |                                     |

In order to determine whether any of the four possible outcomes, O1-O4, is due to luck, good or bad, we need to look, the modal account says, to those nearest possible worlds in which the relevant initial conditions are the same. If the event under consideration only takes place in the actual world, but is suspiciously missing from all, or most, relevant nearest possible worlds, then and only then was

its occurrence in the actual world a matter of luck.<sup>31</sup> But which are the relevant initial conditions that we need to keep fixed on our journey through parallel worlds? I stipulate that it must be things such as the agent’s, the victim’s and the murderer’s characters, intentions, determinations, plans, values, and the like (some of which we will need to add to Constant’s and Kant’s rather colorless and blank original description). For example, a sudden change in my friend’s character or his relation to me – a sudden surge of deep mistrust in me, a helpless panic attack, and the like – certainly belongs to a farther possible world than his remaining confident in me. In a similar vein, a sudden change of the inquiring murderer’s character – say, an instant decision to drop the plan for trivial reasons, or an unexpected, out of blue

<sup>31</sup> Kant seems to equate the class of events whose occurrence is due to pure chance or luck with the class of all non-necessary or contingent events. For him, whatever does not occur by necessity, occurs by chance. And whatever occurs by chance has the chance, or accident, as its cause, and cannot be attributed to anyone’s free agency. If my telling the truth doesn’t necessarily harm my friend (which it certainly doesn’t, since there are possible worlds in which he remains unharmed – where he runs away, or where IM changes his mind and spares his life), then harm, when(ever) it actually occurs as a result of my true utterance, must be due to chance and therefore not my own doing (i.e. something that could be attributed to me as its cause). But such use of “luck” and “chance” is both awkward, from the point of view of its ordinary meaning, and too crude to serve any interesting philosophical purpose. For, as Aristotle correctly observed two millennia ago, while “some things always come to pass in the same way, others do so for the most part. *It is clearly of neither of these that chance is said to be the cause, nor can the ‘effect of chance’ be identified with any of the things that come to pass by necessity and always, or for the most part.* But as there is a third class of events besides these two-events which all say are “by chance”-it is plain that there is such a thing as chance and spontaneity; for we know that things of this kind are due to chance and that things due to chance are of this kind” (*Physics*, Book 2; Chapter 5; my emphasis). So not only things that “always and by necessity come to pass in the same way,” but also those “things that come to pass in the same way for *the most part*” we are reluctant, and justifiably so, to attribute to chance and luck. Since Aristotle has both common sense and ordinary language use on his side here, I will adopt his view on chance and luck and reject Kant’s unnecessarily permissive one. Accordingly, I will continue to treat events which we would expect to occur in circumstances C, because they (statistically) normally occur in C, as not being a matter of chance or luck, even if they don’t always or necessarily occur in C. I’m grateful to my colleague Nenad Mišćević for drawing my attention to Aristotle and for helping me navigate through tricky modal issues in this and following paragraphs.

<sup>30</sup> According to Schinkel’s reasonable-expectation-account, my friend’s death in Truth cannot be ascribed to, or blamed on, bad luck, since it was reasonable to expect that my friend will be killed as a result of my true reply (I had every reason to believe that he was hiding in my attic and that if I tell the IM where he was he would look for him there); in contrast, his death in Lie, does qualify as an effect of bad luck, for given what it was reasonable for me to believe about my friend’s location (on the basis of his trust in me), it was not reasonable to expect him to be on his way to the train station (where I sent the IM, thinking that I’m sending him away) and so it was a matter of pure chance that the IM found him and killed him.

moral transformation – would trivialize the distinction between events that are due to luck and those that are the result of someone's agency. Finally, let's restrict the use of "unlucky event" to only those outcomes that are bad as *a result of luck*.

So equipped, we can start our quest. There are just three variables that we can manipulate in order to depart from our actual world.<sup>32</sup> The most promising candidate for a (un)lucky event is that outcome which we find most unexpected and surprising given what we know about our main characters and given the usual expectations people have about the ways of the world. Such are undoubtedly the two outcomes in the second column, O2 and O4. If my friend trusts me (as he surely must do given that I have built my rescue plan on this assumption),<sup>33</sup> then in most of the nearest possible worlds to the actual one in which his trust in me remains unshaken (call them "trusty-worlds" or "t-worlds") he will have remained hidden in the attic. Next, focus upon the subset of t-worlds, in which I have revealed his hiding-place to the IM. In most of such worlds he will be hunted down and killed (outcome O1).<sup>34</sup> Now look at the other subset of t-worlds, those in which I have sent the IM away under the false pretence. In most, if not all, of these worlds, my friend is going to survive this particular murder attempt (outcome O3). The connection between what I do and the outcome of my action is then strongest, and so least surprising, in the first column. Accordingly, we will come across outcomes O1 and O3 in most of the nearest possible worlds in which the agents themselves have undergone little change of character (compared to particular choices they make and the external circumstances that these choices create).<sup>35</sup> The outcomes in the second col-

umn, O2 and O4, on the other hand, will occur more frequently in more distant possible worlds, those inhabited by more dissimilar counterparts of me, my friend and the IM, but are going to be missing in the nearest possible worlds to the actual one.

What we get in the end, then, is the following picture:<sup>36</sup> survival in Truth (O2) is a matter of good luck; death in Lie (O4) is a matter of bad luck; but death in Truth (O1) and survival in Lie (O3) are not a matter of luck at all. So if luck really rules out responsibility (which I have my doubts about), then we can only be responsible for the death of our friend in Truth and for his survival in Lie. What we cannot be blamed for, however, is his death in Lie. So the conclusion we get is exactly opposite to Kant's own – luck or chance may be an obstacle to our responsibility for our friend's death in Lie, but not also in Truth.

### Conclusion

In the paper I have critically scrutinized Kant's implicit proposal that we should not hold agents responsible for the harm that might result, directly or indirectly, from their scrupulous performance of dutiful/obligatory actions. According to Kant, one bears moral and legal responsibility for the bad (harmful) consequences of his/her actions only if the latter were wrong in the first place, that is if their commission or omission constituted a violation of one of agent's moral or legal duties.

The above principle may sound plausible at first. Reflection, however, shows it to generate some highly counter-intuitive implications. Despite this, I have tried to give Kant a fair trial. I carefully assessed the (often meager) evidence that he himself presented in favor of a suggested moral asymmetry, and listened to reasons that one could advance in its support. In the end I remained unconvinced, however, and hopefully so did the jury of readers. It is hard to find compelling theoretical reasons for a moral distinction that Kant introduces into the domain of outcome moral responsibility, between the

<sup>32</sup> And only two ways, I should add, of manipulating them, if we are to stay in the nearest possible worlds to the actual one (or the relevant subset thereof) – we can either slightly change several variables or we can more drastically change just one of them. To keep the picture simple, I will assume that the possible worlds which differ from the actual world in less respects, but to a greater extent, are nearer to the actual world than those possible worlds that differ from it in more respects, but to a lesser extent. Nothing, however, hinges on this assumption, as far as I can tell.

<sup>33</sup> Mind that my decision to speak the truth is, in virtue of its motivating source, revealing of me alone and says nothing about either the IM or my friend's character.

<sup>34</sup> I say "most" rather than "all" in order to take into account the possibility that he may have left the attic for other reasons than the loss of trust in me, for example, because he has chickened out.

<sup>35</sup> This disassociation of agent's (assumed) character and the choices she makes is one of the reasons why we are reluctant to ascribe those choices and actions to her (as an

agent), or why we keep searching for other causes for, and explanations of, her behaviour. Behavioural explanations in terms of (good or bad) luck or chance can then be classified as a subspecies of those alternative, non-agential explanations. See Sher (2002).

<sup>36</sup> Outcomes 1 and 4 are identical and this for an obvious reason – because the information is true – but they differ morally (in 1 it is meant to be true, in 4 it is true accidentally, contrary to my intention). Outcomes 2 and 3 are identical and the reason for this is again obvious – because the information supplied is false – but they differ morally (in 3 it is meant to be false, while in 2 it is false accidentally, contrary to my best intention).

bad consequences of right and the (equally) bad consequences of wrong actions.

What's the damage? Given its counterintuitive character and absent any compelling theoretical reasons (it goes both against people's firm intuitions and the axioms of a well-established Doctrine of Double Effect), such an account arbitrarily restricts moral and legal responsibility to the category of wrongful actions and omissions; it draws a morally important distinction where there is none, namely between wrongful actions/omissions on the one hand and supererogatory and obligatory/dutiful on the other. Secondly, within the category of wrongful actions/omissions, Kant's account of moral/legal responsibility for outcomes of our actions/omissions is too permissive – it overlooks, or deliberately ignores, morally relevant distinctions between direct and indirect, intentional and unintentional, foreseeable and unforeseeable consequences, as well as between harm brought about by our actions and harm that we merely failed to prevent. And the third and last point: Kant fails to provide a unifying, non-arbitrary principle that would govern the attribution of moral/legal responsibility for the outcomes of our decisions and actions; the reasons he gives (as well as those we can reconstruct) for slicing the moral territory the way he does fall spectacularly short of a convincing case.

## Appendix

Kant writes:

'the French philosopher' confounds the action by which one does harm (*nocet*) to another by telling the truth, the admission of which he cannot avoid, with the action by which he does him wrong (*laedit*). It was merely an *accident* (*casus*) that the truth of the statement did harm to the inhabitant of the house; it was not a free *deed* (in the juridical sense). For to admit his right to require another to tell a lie for his benefit would be to admit a claim opposed to all law. Every man has not only a right, but the strictest duty to truthfulness in statements which he cannot avoid, whether they do harm to himself or others. He himself, properly speaking, does not *do* harm to him who suffers thereby; but this harm is *caused* by accident. For the man is not free to choose, since (if he must speak at all) veracity is an unconditional duty.

The "German philosopher" will therefore not adopt as his principle the proposition (p. 124): "It is a duty to speak the truth, but only to him who has a *right to the truth*" first on account of the obscurity of the expression, for truth is not a possession, the right

to which can be granted to one, and refused to another; and next and chiefly, because the duty of veracity (of which alone we are speaking here) makes no distinction between persons towards whom we have this duty, and towards whom we may be free from it; but is an *unconditional duty* which holds in all circumstances."

The above passage suggests three other possible lines of defense for DA:

My truth-telling was

(vi) duty-bound and hence not genuinely free

(vii) not really harmful to my friend

(viii) even if harmful, not wrongful or unjust

Let's address them in turn.

(vi) Duty-bound, hence not truly free

This line of defense (also found in Lectures on Ethics, p. 59-60) is a clear non-starter. Let me start with a minor objection. The argument from the lack of freedom is difficult to square with many passages in Kant's writings on ethics, where acting out of duty is praised as a mark of free/autonomous and rational agents and contrasted with heteronomous conduct, i.e. conduct motivated by mere inclination or instinct. As a matter of fact, many commentators ascribe Kant's erroneous treatment of the Inquiring Murderer case to his emphasis that my relation to the victim is one of friendship, hence my only, or at least predominant, motive for helping her must be a morally worthless one of affection or love for her. Given Kant's equation between moral and free action, we would expect him to deny freedom of action to the liar, who has after all acted out of affection, but not to the truth-teller, who did what he did out of respect for moral law and in opposition to his natural inclination, or urge, to privilege, and give priority to, his friend and his needs.<sup>37</sup>

Is the argument from the lack of freedom of choice any more plausible outside the narrow context of Kant's moral philosophy? In trying to rebut it, the opponent of DA has two options. She can deny that by being duty-bound to veracity we are not really free to choose a lie, or not free in a relevant sense. Or she can instead accept Kant's premise that we are not free to choose whether to comply with the moral law or not and then deny, in deference to

<sup>37</sup> Here is another worry. It might seem as if Kant is building his defense of the harmful truth-teller on two conflicting assumptions: that speaking the truth was unavoidable or necessary, and that the harm that befell my friend as a consequence of it was accidental. Kant can give neither of these assumptions away. He needs necessity to rule out responsibility, but he also needs chance to rule out harming. This tension is not irresolvable, however, as will soon become apparent.

Harry Frankfurt, that ability to do otherwise is necessary for moral responsibility, i.e. reject the principle of alternate possibilities (or some similar version of it, say van Inwagen's "principle of possible prevention").

In other words, for his appeal to necessity to work, Kant needs to establish the following inferential sequence: obligatory → unfree → not-responsible. The opponent, however, can question both steps, that is the first (if obligatory, then not free) and the second conditional (if not free, then not-responsible).

Let's start with the first. Why believe that moral duty eliminates freedom of choice and of action? Kant's argument for the first inference builds upon the so-called "alternative possibility requirement for responsibility." In order to be guilty, or deserving of blame for X, this principle says, it must have been possible for you not to do (i.e. to omit doing) X. But if you could not have avoided doing X, then it is unfair to blame you for having done it. The first, immediate problem with this defense of the DA is that the necessity that we face in the IM case is not a metaphysical one that typically concerns incompatibilists in the debate about whether there can be genuine freedom of choice and of action in a deterministic universe. If we face any necessity at all (which I myself doubt), it is of a very different kind, namely practical, or normative, or rational. Kant famously equates the "dutiful" with the "practically necessary,"<sup>38</sup> but surely practical necessity, whatever it may be, does not entail metaphysical necessity (i.e. the inability to choose and/or do otherwise) and it is only the latter that threatens to rule out moral (and legal) responsibility. But if it is metaphysically possible for us to withhold truth to the IM, albeit at the pain of being practically irrational or immoral (judged by Kant's standards alone), then we are, or at least in principle can be, responsible for having told the truth to him.

Let me put the same point in a different wording. The kind of freedom that the obligation deprives the agent of is not the same as freedom that an agent must lack in order to shrug moral responsibility off his shoulders. All that the dutiful agent lacks is other permissible, i.e. *equally rational* options, not genuine alternatives or *metaphysically possible* courses of action (including abstinence from the one that the duty prescribes). Just because morality presents an action as obligatory, or *practically necessary*, it does not thereby remove all other *metaphysically viable*

options from the menu, leaving the agent with nothing to choose from. Yet this is the only sense of the "lack of freedom" (namely lack of choice) that matters for the ascription or denial of responsibility. For Kant to plausibly deny the agent's responsibility for the violent death of his friend in the IM case, he would need to convince us that the agent lacks genuine freedom of choice in the situation, not just that she misses other, equally or more appealing (or even rational) options. But since Kant himself admits that lying is a genuine possibility in those circumstances (for how else could the agent violate his absolute duty of veracity in Lie?), the agent is undoubtedly free in the relevant sense and so cannot avoid responsibility that comes with the exercise of his (capacity for) free choice.

The second inference, from lack of freedom (ability to do otherwise) to lack of responsibility, is more intuitive, but no less controversial. Whether moral responsibility requires freedom of choice and action (i.e. the ability to do otherwise) or not, has been the topic of much heated metaphysical debate between compatibilists and incompatibilists. I cannot even begin to trade on this minefield. Suffice it to say that Frankfurt-style examples, however disputed they may be, do seem to suggest that as long as the agent herself initiates the action, as long as the action belongs, in some relevant sense, to her, as long as it is her own and not someone else's, she may be justifiably held responsible for it (and its consequences), regardless of whether she was genuinely free to do or omit the action under consideration.<sup>39</sup> Hence even if we granted Kant the premise that duty implies lack of freedom, this premise would still fail to establish the conclusion that Kant needs about the absolving power of duty-fulfillment.

#### (vii) No harm done

Can the prosecution rest its case then? Before we bury Kant, two more solutions he offers in passing need to be considered and rebutted – that I did no harm and that even if what I did harmed my friend, I committed no wrong or injustice by telling the truth.

Let's begin with the idea that I did not harm my friend. Kant seems to suggest that the ensuing harm to my friend was caused by accident and not by my truth-telling. According to this interpretation, then, we should be reluctant to ascribe to the truth-teller the responsibility for the harm that her friend suf-

<sup>38</sup> This is, after all, one of his reasons, or the main reason, for denying the possibility of a genuine moral dilemma or conflict of duties.

<sup>39</sup> See Frankfurt (1969) for some examples of actions for which we tend to hold the agent morally responsible despite the fact that he couldn't have done otherwise. For a nice overview of the current state of the debate on free will and determinism, see Šuster (2007).

ferred in a form of a violent death, because we cannot identify any free deed of the former such that the resulting harm could be attributed to it – but if so, harm must have occurred as a matter of pure accident instead.

Schwarz (1970) is one of the few who finds this transmission of responsibility congenial:

A crucial distinction is introduced here by Kant, namely, between the act of making an unavoidable truthful statement from which ensues harm to another person, and the act by which someone wrongs another person (the difference between the Latin *noceo* and *laedo*). The effect of the first kind, says Kant, must be attributed to accident and cannot be attributed to a free deed of the one who is compelled to make a statement. For, ‘if it can’t be helped that he must speak,’ he has no choice under the law of freedom but to be truthful. If harm ensues from his truthfulness, it is not he who causes it; any harm is here an event of chance caused by accident.

I think that this distinction is well founded. If the consequences of a truthful statement cannot be thought to be determined by it, they cannot be imputed to the agent. Kant, therefore, speaks also of the “unforeseen” consequence. This practical-formal, thus legal, argument in no way precludes our duty to take into consideration the possible consequences of our words so far as we can foresee them. But the indeterminate or “broad” ethical obligation to protect the life and to further the happiness of others to the best of our ability cannot displace the strict duty of truthfulness, and we cannot hold ourselves accountable for whatever uncertain effects may follow when we are compelled to speak and therein adhere to the truth. We may think also of cases of legitimate compulsion, for instance, of being subpoenaed to testify. One may become undeserving or meritless, without becoming “guilty” (ibid. 64)

Can we make sense of this kind of excuse? Would it, or should it, convince my friend? Here is how one could try to develop this idea. My friend suffered harm (after all, he lost his life), but either no harm was done to him (no one harmed him) or at least it wasn’t me who harmed him. The first suggestion, I believe, can be quickly dismissed. My friend was a victim of a murder, after all, not an earthquake or a deadly disease. So not only was he harmed, he was harmed by someone rather than something. The second interpretation hardly fares any better. It is fueled by the intuition, I suppose, that if I did the right thing by telling the truth, then all the harm that befell my friend (at least partly) as a result of me providing accurate information must

have stemmed from the murderer’s evil intention and action with a major contribution of accident or bad luck. Yet, consider the causal chain again. My friend lost his life as a result of being, let’s assume, shot to death (without justification), and the shooting was carried out successfully with my assistance, because of something I did deliberately and in full awareness of not just possible, but likely, consequences, namely reveal my friend’s hiding place. Now on any plausible account of harm(ing), if something I did put another person in an objectively bad situation (by thwarting one of her core interests) or in a comparatively worse situation than the one he would have been in had I not acted the way I did, i.e. had I not told the truth, I harmed that person. So there can hardly be any doubts about whether I harmed my friend or not. The fact that I pursued no evil intention/goal in my action, that in doing what I chose to do I was ignorant or negligent of its possible detrimental effects, cannot alter this basic fact. What it might, and should, do is mitigate the degree of my moral and legal responsibility for the harm done. But it cannot absolve me of all the responsibility for his death, for I have knowingly (though unwillingly) become complicit in a criminal conspiracy to end his life.<sup>40</sup>

Kant seems to assume, mistakenly, that if harm was merely accidental to truth-telling then it cannot be imputed on the agent as its cause. Here the underlying thought seems to be that if what the agent does is not harmful in itself, intrinsically, essentially or in virtue of what it is, its very nature, then she (or her free deed) cannot be considered the cause of the harm, rather we should ascribe the harm to chance or accident instead. How plausible is this as a general principle? Even if we accept the questionable binary opposition (if harmfulness is not an essential property of X, then it must be a purely accidental one), the fact that the ensuing harm is a contingent feature

<sup>40</sup> The charge of harm(ing) to my friend finds further support in the account of intentional harm that underlies the theory and the practice of tort law. Tort law protects people’s core interests against harm intentionally or unintentionally caused by the activities of some third party. When is harm intended and when not? “A defendant is said to intend harm if she either desires to bring harm about, or she is substantially certain that harm will ensue as a (possibly undesired) side-effect of her activity” (Perry 2001; 58, my emphasis). So not only can I not possibly escape liability for harm that befell my friend (for I am going to be liable for it even if I can prove that I did not intend his harm), my telling the truth to the IM will constitute intentional harm, since it will be hard to deny that I was “substantially certain that harm to my friend will ensue as a side effect of my telling the truth about my friend’s whereabouts to the IM.”

of the action performed does not really prove that the action was not one of harming.<sup>41</sup> Both ordinary language and moral and legal theory allow for the possibility that harm can be caused (or allowed) without intention and/or foresight. But if you can harm someone non-deliberately, unknowingly, even contrary to your best intentions (at least the tort law is, as we have seen, built on this premise), why deny that you harm someone by doing X if your doing X puts him in an objectively bad, or comparatively worse, position, or thwarts his interests and you can clearly foresee this consequence at the time of your choice?

A different line of rebuttal is also worth exploring, insofar as it promises to bypass objections against the charge of harming. Let's assume, uncontroversially, that to harm is to cause harm. All it takes to show that I harmed my friend by telling the truth to the IM would then be that my telling the truth caused my friend to die. This might not look very promising at first. For surely it was the IM's shooting that caused his death and not my words to the IM. Besides, didn't Kant deny that the truth-teller could be considered the cause of harm on the basis that his action was only contingently harmful, i.e. not sufficient for harm? These worries should not discourage us, however. It suffices to remind oneself that an event, A, can be the cause of another event, B, without being sufficient for its occurrence, to see that Kant's inference from "it is not necessary that if X, then H (harm to B)" to "X was not the cause of H" is invalid. Furthermore, as Moore rightly points out, "some accomplices are causes of the harms they aid another to cause, whereas other accomplices either make no causal contribution to a harm by their aid, or they make quite small causal contributions to that harm" (Moore (2007); 421). So I can be guilty of complicity in crime even if the causal contribution I personally made to the harm was relatively small or none. But if so, then the question "Which of these classes do truth-telling and lying to the IM belong when they implicate me in murder – do I aid the IM to cause my friend's violent death by causing it myself, or is my causal contribution to the latter small or even negligible?" becomes less pressing to address.

Luckily, then, whether I caused the harm to, or harmed, my friend by telling the truth (or by my failed attempt to lie) or not is not crucial here. It would be an important issue to settle if one believed that every instance of wrongdoing is an instance of

harming (that "wrongdoing" and "harming" are co-extensional terms). For only then would it be necessary (even though not yet sufficient) to establish that I harmed my friend in order to prove that what I did was wrong (since on this account, there can be no wrong without harm). But this is not what Kant's critics primarily object to here. I have, after all, granted Kant the (admittedly controversial) assumption that telling the truth to the IM was the right thing to do. So no amount of harm to my friend that I can be justifiably identified as the cause for, can establish that I acted wrongly in speaking truthfully. Rather, the crucial issue is whether, if my telling the truth was part of the causal chain that led to my friend's violent death and if I could foresee this (or had every reason to believe, at the time of the utterance, that it would lead to it), I bear some responsibility for it or not.

The question, then, is not whether I harmed my friend or not by doing what I did, but whether or not I am partly responsible for the harm he had suffered as a result of my action, regardless of whether the harm was directly caused or merely not averted. As long as the reader shares not just my own, but a widespread, sentiment/intuition that one can be morally responsible not just for the harm directly caused, but also for the harm that one could have prevented, but did not, my objection retains its full force. That's why it doesn't really matter whether we classify my truth-telling to the IM as causing (i.e. harming) or merely allowing harm to my friend. For we can insist that even though I did not directly harm my friend, I still bear some responsibility for his tragic fate, for I have willingly and knowingly omitted an alternative course of action (telling a lie) that could have (and almost certainly would have) saved him. Hence little, if anything, is gained by resisting the conclusion that by revealing my friend's hiding-place, I have directly harmed him. For whether I did or not, it is indisputable that I at least allowed the worst possible harm to befall him. And harmful omission (of assistance) is certainly as good a ground for assigning moral responsibility (or tort liability) as any other.

(viii) Harmful, but not wrong or unjust (harm done, but not wrongly or unjustly)

So denying that I harmed my friend by revealing his hiding place is clearly not going to work. Fortunately, there is a way out of this for Kant. Assuming I did harm my friend by telling the truth about his whereabouts, I can still get myself off the remedial or corrective responsibility hook as long as causing him harm in these circumstances was neither wrongful nor unjust. For on the plausible assumption that you can only owe apology or compensation to

<sup>41</sup> What could, I wonder, possibly motivate the view that, true, the given action was harmful to someone, yet it was not of a harming kind?



someone if you have wronged her or did her injustice,<sup>42</sup> as long as we can convincingly show that the truth-teller has neither wronged her friend nor committed any injustice to her, she won't be responsible for her death, not even in a minimal, remedial sense. And since we did agree with Kant at the outset that telling the truth was morally right, and since being morally right entails not being morally wrong, the prospects start to look brighter for Kant: it truly does seem implausible to argue now that we have wronged our friend or done her injustice by doing something that was not wrong to do.

Let's reconstruct Kant's argument then:

(1) No action-token can be both morally right and morally wrong.

(2) Telling the truth to the IM is the morally right thing to do in the IM case.

Hence, from (1) and (2)

(3) Telling the truth to the IM is not morally wrong in the IM case.

(4) For an action to be morally wrong, it must be an instance of either wronging someone or committing an injustice against someone.

Or

(4\*) Wronging and committing injustice are jointly necessary and separately sufficient for wrongdoing.

Hence, from (3) and (4)

(5) I have neither wronged my friend by telling the truth to the IM nor have I thereby committed any injustice to him.

But since it is also true that

(7) Without wrongdoing (wronging?) there can be no remedial responsibility, and without injustice there can be no obligations of corrective or compensatory justice.

Therefore

(8) I cannot be responsible for my friend's death, not even in a minimal, remedial sense, nor can I be under any obligation to correct any injustice, or put any wrong that she has suffered right again.

<sup>42</sup> In other words, that both outcome and remedial responsibility entail wronging and that corrective justice entails doing injustice.

Kant's critic again has two options here:<sup>43</sup> first, she can argue that you can be responsible (in one of the many senses of this word) for harm which someone suffered through your action or omission, even though you did no wrong or injustice to him by causing or failing to prevent this harm. Or she can take the easier route and question Kant's claim that by revealing your friend's hiding-place to the IM you neither wronged him nor did him any injustice. The critic, then, can challenge the truth of either premise (7) or premise (5) of the above argument.<sup>44</sup>

Let's begin with the premise (7). Is it really true that I can be under no obligation to remedy the harm that I caused, or helped cause, or failed to prevent, as long as by causing or allowing harm to a particular person I neither wronged her nor did her any injustice? Our intuitions about the IM case are not so straightforward, I'm afraid. On the one hand, we are not perfectly immune to the appeal of saying that if someone was wronged, but it wasn't you who

<sup>43</sup> Actually three, for one could question premise (1) as well. Surely, one and the same action-token can be both *prima facie* right and all things considered wrong, or vice versa. It is perfectly plausible to say of truth-telling to the IM, for example, that it is *prima facie* morally wrong (because it causes undeserved harm), but all things considered morally right (because the fact that it fulfils the duty of veracity trumps any possible welfarist considerations). In fact, it is even more plausible to claim, as many of Kant's followers did, that it is *prima facie* morally right, but all things considered morally wrong. I will not pursue this line of rebuttal any further, however, because it renders my initial concession to Kant insincere.

<sup>44</sup> I am simplifying a bit here. For if Parfit's analysis of actions which cause people to exist in a less than ideal conditions (Parfit 1984) is correct, then there might well be such a thing as a victimless or impersonal wrong(doin). This might look like an emergency exit for the critic of Kant, for he may now argue that even if I neither wronged nor committed any injustice to my friend by telling the truth to the IM, my conduct was nevertheless in a way morally faulty, making blame and guilt (or some sort of moral unease) quite appropriate. I don't think Kant's critics should accept this offer, however. Admittedly, Kant himself seems to embrace the idea of impersonal wrong(doin) or injustice, when he condemns lying to the IM as a wrong or injustice committed against humanity at large (but no one in particular). The suggestion that if I tell the truth to the IM, I may be guilty of an impersonal wrong or injustice, is nevertheless problematic. It fails to accommodate a firm intuition that whatever duties we may owe as a consequence of our harmful action, we will have owed at least some of them *to our friend*, or at least some of them will arise from what we did *to him*. As long as we classify the harmful truth as an *impersonal* wrong or injustice, this personal aspect remains a kind of an unresolved mystery.

wronged him, you bear no direct responsibility to remedy the situation; also, if someone was a victim of injustice, but it wasn't you who committed it, you cannot be put under an obligation to either correct the injustice or compensate the victim for it. Imposing such a burden would be simply unfair. This set of intuitions is balanced by another, however, suggesting that we should at least feel bad about our role in the whole process, or have doubts about the way our commitment to moral obligations was abused and twisted to serve an evil purpose. Also, if someone was wronged and/or suffered injustice and we have done something which we knew will facilitate the action that constituted the wrong and injustice done to him, pointing to the fact that we did not personally perpetuate the wronging and the injustice, but merely assisted it or stood by, will not even provide a convincing excuse, let alone a plausible justification. Even if our (causal) contribution were merely to the harmful aspect of a wrongdoing (such as the murder of our friend), but not also to its wronging or injustice side (which I find hard, if not impossible, to keep apart), most of us still finds some moral residue in the form of agent-regret, apology, even compensation appropriate, and would look at a content truth-teller with the same suspicion that Bernard Williams felt towards the truck driver who feels no remorse or regret over the death of a young girl that he blamelessly ran over.

Another point. The question of how exactly "harming," "wronging" and "doing injustice" are related to each other is notoriously difficult and far too complex to address properly in such a short space. For the purpose of this examination, I will therefore simply assume, uncontroversially, I think, that every instance of wronging is an instance of doing injustice and vice versa, that whenever I commit injustice, I wrong someone.<sup>45</sup> In contrast, harming, even intentional, doesn't seem to be conceptually linked to wronging and doing injustice. It is, then, at least in principle, possible to harm someone without thereby committing any wrong or injustice to him. Punishment, for example, is defined as intentional infliction of harm on someone for the purpose of restoring damaged moral universe, preventing future crimes or remoulding the criminal's character and personality. Yet only few of us would be willing to object to the very idea of a morally legitimate punishment.

This, however, need not be anything other than a mere conceptual concession on the part of Kant's critic. For Kant's strategy to work, he would need to convince us that doing harm was neither wrong nor unjust *in the case at hand*. This brings us to the second issue. I suppose everybody agrees that my friend was not just harmed but wronged and a victim of injustice by being murdered. The question we need to ask then is whether he was not just harmed, or allowed harm, but wronged and done injustice by me revealing his hiding-place to the IM. What reasons does Kant give for believing that I have done no wrong or injustice to my friend even though I knowingly provided the IM with a crucial piece of information? Well, as far as I can tell, there is just one. Kant seems to presuppose that in order to commit injustice to you, I must violate some rights of yours or other. And the only right he can think of is the right that others lie for your benefit. Since there neither is nor can ever be any such right, I cannot have violated it and my friend has no ground for a complaint that I did him injustice.

Such a conclusion is clearly premature, however. Even if Kant is correct to insist that no one can legitimately demand that others sacrifice their moral duties for his or her benefit or convenience, we can think of other potential rights that I will either infringe or violate by telling the truth to the IM. How about the right to assistance in a legitimate self-defense, the right to life and bodily integrity (interpreted broadly, as a right to the provision of life-saving and integrity-preserving measures), the right to the aid, the right not to be a victim of crime, and so on?

Kant is blind to all these possibilities. The only right he can think of is the odd-sounding "right that others violate their moral duties for our convenience," the right so construed that no one would want to grant it to any person. Kant's blindness might have been of a theoretical origin. Perhaps he believed claims about rights and privileges to be reduceable to corresponding claims about the permissibility or impermissibility of actions. According to this "simple" view, as Thomson (1986) calls it, "You may permissibly do X" entails "No one (including me) has a claim-right against you that you not do X." Now if permissibility of X entails absence of claim-rights with respect to non-X, then this must be even more true of an obligation to X – so that if I ought to tell the truth to the IM, then my friend cannot have any claim-rights against me that I omit telling the truth which I could violate by telling the truth. Yet this "simple view" is, as Thomson and Feinberg show, mistaken. There is nothing obstruse about believing both that it was my overall duty to

<sup>45</sup> Remember, however, that on this account neither wronging nor treating unjustly need entail harming – wrongdoing and injustice can be perfectly harmless, but no less wrong for this reason.

tell the truth and that I have thereby infringed upon, or violated, one or several of my friend's right(s). It simply doesn't follow from "I did what in the given situation was the right thing to do, all things considered (i.e. a Rossian 'duty proper') that "I didn't wrong anyone" or that "I didn't commit any injustice" or that "I didn't infringe or violate any(one's) rights."

To conclude. I find both assumptions – that rights-violations and infringements are the sole source of injustice, and that the only right that my friend could legitimately invoke in the case at hand is the implausible claim-right that others violate their duties for his benefit – hard to swallow. Kant has failed to convince us both that duty absolves of all the responsibility for harm, and that duty-violation morally implicates, both generally and in the particular, IM case.

## References

- Aristotle, *Physics*. (1987) In: *A New Aristotle Reader* (ed. by J.L. Ackrill), Princeton, NJ: Princeton University Press, 81–131.
- Driver, Julia (ms). "Luck and fortune in moral evaluation" (unpublished manuscript)
- Duff, R.A. (2006). "Answering for crime." *Proceedings of the Aristotelian Society*, 85–111.
- Frankfurt, Harry (1969). "Alternate possibilities and moral responsibility." *Journal of Philosophy*, 45; 829–839.
- Hill, Thomas E. jr. (2000). "Kant on responsibility for consequences." In: *Respect, Pluralism, and Justice*, Oxford: Oxford University Press; 155–172.
- Honoré, T. & Hart, H.L.A. (1962), *Causation in Law*, Oxford: Oxford University Press.
- Jokić, A. (2002), "Supererogation and moral luck: two problems for Kant, one solution", *The Journal of Value Inquiry*, 36; 221–233.
- Kant, Immanuel (1797). "On the alleged right to a benevolent lie" (several English translations available online).
- Kant, Immanuel (1963). *Lectures on Ethics*. Transl. by Louis Infield, Indianapolis, Cambridge: Hackett Publishing Company.
- Klampfer, Friderik (2004). "Moral responsibility for unprevented harm," *Acta Analytica* 19/33; 119–162.
- Kekes, John (2002), "On the supposed obligation to relieve famine," *Philosophy* 77; 503–517.
- Korsgaard, Christine M. (1996). "The right to lie: Kant on dealing with evil." In: *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press; 133–158.
- Lazar Seth R. M. (2008), "Corrective Justice and the Possibility of Rectification," *Ethical Theory and Moral Practice*, Accepted Jan. 2008.
- Miller, David (2007). *National Responsibility and Global Justice*. Oxford: Oxford University Press.
- Moore, A. W. (1990). "A Kantian View of Moral Luck." *Philosophy* 65; 297–321.

## **Book reviews**

**Marin Aiftincă (2008). *The Philosophy of Culture*, Romanian Academy Publishing House, Bucharest, 2008, 164 pages**

**[Reviewed by Alexandru Boboc, Bucharest University, steluta@acad.ro]**

Within the sound traditions of the Romanian philosophic train of thought and by his own systematic comprehension, the author deals with the basic ideation of the philosophical culture. "Starting no doubt with familiar things, the philosophic idea tried, within its own power, to suggest new ways of interpretation for the cultural phenomenon and conceptual reconstruction" (10).<sup>1</sup>

This confession voices the idea and the program for research in a modern theoretical-methodological reconstruction that forms, with discernment and nuance, the tradition and the more significant recent steps leading to a process that gathers in a conceptual universe reflection on man, values, culture and inter-cultural communication nowadays. The presence of value is considered to be decisive: "To talk about culture outside value is to depict the day without daylight. As a day without the heavenly life-giving body light is beyond any conceptualization, real culture cannot exist without the presence of value as well... the two entities, the culture and the value, cannot be parted" (38).

According to this understanding, the structure of the ideation with reference to a modern comprehension of culture has a point of reference in the world of values, which the author follows historically (within the vast axiological concepts) and systematically (within a personal re-expressiveness) well sustained on a conceptual and argumentative basis.

Significant are also the preoccupation with the "conceptual approaches" (even beyond this chapter III) and a sui generis phenomenology belonging to the culture of philosophy, based on which the theme is systematically pursued in successive phases: man-nature-culture; world of values - world of culture, culture and society; culture and civilization; unity and plurality of culture; tradition, value and intercultural communication; culture within the frames of globalization.

In the Introduction, the idea of this comprehension gives ways to this step: "Man and culture are the inseparable constituents of a bi-univocal relation that involve with priority philosophical reflection. Any matter that engages in a direct or indirect manner man's existence and destiny immediately addresses culture. The tensions in the world caused by the globalization process, military and religious influences determined, lately, the pressing resort to

culture, in search of arguments and solutions for the protection of dignity and cultural identity, for the promotion of the intercultural communication as a source of lasting creativity and development. All the same for the return of the man to his self and for moving his thought towards transcendence, by a steady effort into a deeper understanding of the sense of human being. Almost all of the scientific national and international reunions that approach, in one way or another, the theme of globalization, democracy or hint at sundry area projects call attention for the cultural concern" (page 13).

The motivation of the orientation towards culture in the time of scientific and technological ascent has not only theoretical-methodological grounds for reconstruction, but also incentives from social reality, illustrating somehow once more the interaction between theory and practice, but (this turn) in fixing the communicational ground (largely detailed theme under chapter VII: Tradition, Value and Intercultural Communication). "A theme of stringent actuality, in the sphere of those that may be called provocations of the contemporary world addressed to the society as a whole and to the philosophy, is the pluralism of culture and the intercultural communication under the conditions of globalization", that drops necessity on the preoccupation for "solutions for overstepping the difficulties and intercultural blocking, in order to rearrange the cultural universe that was meant for us to live in" (page 115).

In fact, the author's steadfast preoccupation goes in this very direction, theoretically approaching, but with applicative openings, the themes: pluralism and intercultural communication; the traditions and values, under wordings like: "Tradition, an active spiritual energy"; "cultural tradition, a way of values embodiment"; "traditions and values - premises of interculturality."

What needs to be followed, beyond the objective difficulties of the intercultural communication, as well as the attitudes that "extrapolate the difficulties", is "the constructive solution, able to answer the provocations of this historical time", the philosophy asserting here "one of its major functions", outlining "the conditions of the possibility of culture evolution" (page 118).

To be noted that the theorization of this opening is given by the chapter VII (Unity and Plurality of

Culture), that makes an account in clean well-defined terms on a bottom ideation in a philosophy of culture: universal and specific in the culture, the concept of national culture. Because “we have to do not with one, but with a multitude of cultures, made up along a tensioned historic process, in a large diversity of historical and environmental conditions” (page 111). As consequence, “the frequent correlate of the universal culture is the national or local one”; “within its structure, the national culture gathers elements of human generality, and also traditions, personal customs, dominant themes and ideations in the epoch, techniques for creation and research specific or borrowed, building up an organic unity, that allows the foundation of new original values that represent the contribution to the enrichment of the universal culture’s thesaurus” (pages 111, 112). In a brief statement: “Through what stands for local, personal and national issues, the culture penetrates the universal one. In this sense, the local, national get a special value. The universal in culture appears and is achieved through the particular” (pages 112, 113).

The more necessary the dealing with these themes is, the more lives the present world in the époque of globalization and difficulties that are engaged by “the projection of globalization on culture” (page 130). Here the attention is called by the “globalization and postmodernism” theme, with the remark that: “the globalization interweaves with the postmodernism (...) Pleading the right of freedom, difference and pluralism, the postmodernism shows no respect for the tradition, spirituality, national values that warranted so far the social equilibrium and individual’s security” (pages 133, 134).

On such background also the idea of “global culture” is projected, initiated in fact as “a homogeneous reality”, that “modifies the axiological scale, placing on the highest step the utilitarian values” (page 135). That is why it is necessary to redefine the culture – value – “global culture” proportions, since adding the term “culture” to its predicate “global”, this produces “a change in content that indicates a quantitative meaning, prevalently totalizer” (page 139). Within this context a ground analysis is unfolded for what is called “global culture” (pages 140-143), referring to the phenomenon of losing the value dimension (especially the spiritual values), to wit “the existing unbalance in the system of values”, following the overbid of the utilitarian values.

Naturally the theme of cultural identity is being hinted at, by a sustained pleading for “culture in its absolute meaning of notion”, specifying: “The culture and cultural identity make up a correlation according

with the pattern one and multiple, absolute and relative”; considering that: temporal and spatial conditionings that mark the achievement of culture”, “it is far from us not to recognize that there is multitude of cultures, constituted along a tensioned historic process, in a vast diversity of historical and environmental conditions”; “the spiritual force shapes the culture of each nation, imprinting to it emphases that define its identity”; because the nation “sublimates” the creative energies into original values”, we must talk first “about a national cultural identity and only after that, by filiation, in wide sense, about regional cultural or, by compression, individual and group identity” (pages 144-146).

By confronting the globalist theories, and also the ones that overbid the specialty, ones assert the idea of a dynamic conception about culture, in which the universal and the specific is harmonized, conditioning an opening within the appearance of communication between cultures. Because “any tradition and, essentially, any culture keeps its vigour and identity as long as it keeps on getting re-created, in consensus with the exigencies of modernity”, that constitutes the base of a real dialogue, tested by rules that offer “the chance for asserting cultural identity” (page 149).

Let us note that the above-mentioned improvements concerning the unity and plurality of culture, with focus on understanding the position of the national culture and cultural identity under the conditions of interaction and communication of cultures in the globalization era, find an start for the study of the fundamental ideation of a philosophy of culture: the value – reason for culture; the relations between nature and culture; culture and society; culture and civilization. In essence, we can dissociate a theoretical-methodological preparation, with stress laid on the axiological side, in which the author reveals his availabilities for research arisen from the strong intellection in the field, in point of both systematical constructions and confrontation of concepts on culture, exhibiting the value of tradition and following at the same time the value dimension of the historic context of the establishment of one modern philosophy of culture.

The specialty literature, used with nuance and discernment spirit, especially the systematic side, the participation to modern reconstruction of the domain confers theoretical value to this book, besides applicative valences. The clarity of style and the remarkable degree of conceptualization and ideation that sustain its constructive procedure ranges it with full right in the line of reference traditions of the participation of Romanian thought to the set up of the philosophy of culture.

**Psillos, Stathis (2007). *Philosophy of Science A-Z*. Edinburgh: University of Edinburgh Press.**  
**[Reviewed by Lilia Gurova, New Bulgarian University, Sofia, Bulgaria,**  
**e-mail: lilia.gurova@gmail.com]**

Those who have their interests in philosophy of science cannot complain that there are not enough up-to-date reference books in the field. Several introductions ([Rosenberg, 2000], [Okasha, 2001], [Losee, 2001], [Ladyman, 2002], [Hitchcock, 2004]), two companions ([Newton-Smith, 2000], [Psillos & Curd, 2008]), an encyclopedia ([Sarkar & Pfeifer, 2006]) and an impressive 16-volume Handbook in philosophy of science<sup>1</sup> have been published during the last 10 years to respond to the interest of those who would like to learn more about the central concepts and issues in this philosophical discipline. What's the place of Stathis Psillos' *Philosophy of Science A-Z* in this apparent abundance of guides?

The fact that attracts the attention first is that despite its rather modest size (less than 300 pages) Psillos' "A-Z" philosophy of science book contains nearly 390 alphabetically ordered entries. This is, for example, four times more than the number of the entries of Newton-Smith's Companion (81 entries). And it is nearly three times more than the number of the entries of the two-volume *Philosophy of Science Encyclopedia of Routledge* ([Sarkar & Pfeifer, 2006], 132 entries), which has been represented upon its publication as "the first in-depth reference in the field."<sup>2</sup> Of course, the texts included in the "A-Z" book are significantly shorter than the papers written for the Routledge Encyclopedia and in this sense Psillos' book is more like a dictionary.<sup>3</sup> But the modest size of the entry texts does not take away from the fact that this reference book provides an impressive network of entries to the main terms and names which represents philosophy of science in a

large historical and a contemporary perspective. Although approximately short in size, each entry is supplied with multiple bold-faced cross-references as well as with references to books for further reading.<sup>4</sup> If the reader decides to follow the paths through the text, pointed to by the indicated cross-references, she/he will have the pleasure of accessing to a surprisingly rich and at the same time a concise and comprehensible picture of the explored issue and its closest context. Let me support this by a short demonstration of how exactly the cross-reference system works.

Suppose you are not a professional philosopher but one who nevertheless keeps some sense of curiosity in philosophical matters. And let us suppose that one of the questions which has occupied your mind for some time is "the nature of mathematical objects": how do those things, which professional mathematicians recognize as their subject-matter, exist? You expect that there are maybe different answers to this question and discussions about them and you are keen to learn more about this stuff. You try first to look for information in the authoritative two-volume *Philosophy of Science Encyclopedia of Routledge* [Sarkar & Pfeifer, 2006] but you do not find any relevant entry there.<sup>5</sup> Let's see how the "A-Z" dictionary can help you in your search. Indeed, there too you do not see any entry for "mathematical objects" but you find one for "numbers." And since numbers, in your perspective, are good instances of mathematical objects, you decide to explore the entry for numbers. At the beginning you are disappointed for a while that there is not a paper assigned to this entry but only references to other entries such as "abstraction principles," "fictionalism, mathematical," "Frege," and "Platonism, mathematical." But after that you say, well, let me see what I can get from this cross-references. You start your inquiry with the paper discussing "abstraction principles." Reading this paper you learn that the notion of ab-

<sup>1</sup> The 16-volume *Handbook of the Philosophy of Science* has been published by Elsevier, under the general edition of D. Gabbay, P. Thagard, and J. Woods. The different volumes devoted to different subfields of philosophy of science have their own editors: the volume on general philosophy of science, for example, has been edited by T. Kuipers [Kuipers, 2007].

<sup>2</sup> See the publisher's announcement at <http://www.routledge-nv.com/ref/philosophy/science/>

<sup>3</sup> And in fact its author calls it a "dictionary" in the *Introduction and Acknowledgements* part of the book [Psillos, 2007; x-xi].

<sup>4</sup> The list of all books suggested for further reading contains about 300 titles.

<sup>5</sup> Strange enough indeed, in the two-volume set of [Sarkar & Pfeifer, 2006] there are no entries related to philosophy of mathematics.

straction principles has been “introduced by **Frege** in an attempt to explain our capacity to refer to **abstract entities**” [Psillos, 2007; 6].<sup>6</sup> The way the abstraction principles work is illustrated by two examples and one of these examples, now you are happy, is about numbers. From this example you learn that “the concept of number (and numbers as abstract entities) can be introduced by a similar abstraction principle, namely: (N=) The number which belongs to the concept *F* is the same as the number which belongs to the concept *G* if and only if concept *F* can be in one–one correspondence with concept *G*. The notion of one–one correspondence is a logical relation and does not presuppose the concept of number. Hence, the right-hand side of (N=) does not assert something that is based on intuition or on empirical fact. Still, (N=) states necessary and sufficient conditions for two numbers being the same; hence, we are offered identity-conditions for the abstract entity *number*” [Psillos, 2007; 6-7]. This sounds tough and not quite clear, you think (remember, you have never read Frege before). Yes, there is a book [Fine, 2002] recommended for further reading but you decide that it is too early for you to study primary sources. You find more reasonable to keep looking through the “A-Z” dictionary, led this time by what you have already learnt about Frege’s view of mathematical objects: that (1) they are abstract entities; and (2) they are somehow related to concepts, which are also abstract entities. Thus, the notion of abstract entity appears to you as crucial for the understanding of mathematical objects “qua abstract entities,” so you decide to explore the entry for “abstract entities.”

From the text assigned to this entry you learn that abstract entities are those entities which “do not exist in **space** and **time**” and which are “causally inert” [Psillos, 2007; 5]. This definition of abstract entities puzzles you immediately: if abstract entities do not exist in space and time, and if they are causally inert (i.e. they do not cause anything), *where, how exactly*, do they exist? And *how do we know* that they exist at all? Continuing reading the paper you learn that the difficulty in answering these two questions is the crucial problem for those who believe in the existence of abstract entities and in particular for those who believe that numbers are such entities. Now you realize that it is time to look for alternative views about the nature of numbers. Going back to the cross-references suggested under the entry for “numbers,” you decide to check this time

the entry on “mathematical fictionalism.” Thus you learn that according to mathematical fictionalists, “numbers and other mathematical ‘entities’” are nothing but “useful fictions” [Psillos, 2007; 93]. You learn also that “fictionalism is a kind of **instrumentalism** about mathematics” and that as a sub-species of **nominalism** it opposes to mathematical **Platonism**. But if numbers and other mathematical entities are mere fictions, you ask, how could they be “useful” in the study of the real world, which is not fictitious? This question is enough to inspire you to look for other alternative views. Next in the initial list of cross-references is the entry for Frege. There you learn that Frege defends his view of numbers as abstract entities against Kantian view of numbers as intuitions, against Mill’s view of numbers as generalized properties of physical objects, and against Berkeley’s view of numbers as subjective ideas. Now you feel that you have enough alternatives for the nature of numbers: abstract entities, fictions, intuitions, properties of physical objects, or subjective ideas. In order to decide on any of these alternatives, however, you need to continue your study, and maybe you should go to the original texts of their proponents.

Let’s stop our imaginary story here: it is demonstrated already that the cross-reference system of the “A-Z” dictionary works. It allows readers of all educational backgrounds to orient themselves through the subject that interests them.

The prospective readers of the A-Z book might be interested in how large the area of the subjects covered by this dictionary is. The knowledge about the total number of entries is not enough to answer this question. We should know how many categories of subjects are covered and which these categories are.

A careful look through the list of entries reveals that the great number of them fall into one of the following categories:

(1) concepts which are part of the specific vocabulary of the philosophy of science (for example: ad-hoc hypotheses, covering-law model, underdetermination);

(2) concepts which are part of the general scientific language and in the same time of interest for philosophers of science (examples: theories, models, experiments);

(3) general philosophical concepts (a priori/a posteriori, knowledge, truth, reality);

(4) important philosophical doctrines (realism, empiricism, instrumentalism, Bayesianism);

<sup>6</sup> The bold-face everywhere in the cited texts means that there are entries in the dictionary for the bolded terms and names.



(5) paradigmatic scientific theories (theory of relativity, quantum mechanics, Euclidean geometry, non-Euclidean geometries);

(6) principal figures in history of science (Copernicus, Galileo, Newton, Darwin, Einstein, Bohr);

(7) principal figures in history of philosophy (Aristotle, Ockham, Bacon, Descartes, Kant);

(8) principal figures in philosophy of science (Carnap, Popper, Kuhn, Lakatos).

Of course, there are entries ("ethics of science" is an example) which do not belong to any of the categories (1)–(8) but their existence only strengthen the impression that the covered area of topics is indeed extremely large and that makes a search through the cross-reference really fruitful.

Two groups of entries deserve to be particularly stressed. The first is the category of the concepts, which belong to the specific vocabulary of philosophy of science. It is no surprise that this category is the most extensive one. But the fact that makes it a true merit for the dictionary is that here one can find entries (and thus inquire about notions) which are not present in any other reference book (to mention only the entries for "bootstrapping," "ceteris paribus laws," "consilience of induction," "laws of thinghood," "Markov condition," "novel prediction" etc.).

The second group of entries, which deserve special mention is the category of the principal figures in philosophy of science. Here, besides the names of classical scholars like Carnap, Schlick, Popper, Feyerabend, one can find entries for philosophers of science born in the 30s and 40s of the 20<sup>th</sup> century. This is the generation that shaped philosophy of science after the post-positivist turn, which Popper, Lakatos, Kuhn, and Feyerabend were credited with. This generation has not yet been given a common label, partially because it has not yet become a subject of a careful historical reflection. By giving entries to this provisional list of names of people who contributed a lot to the maturation of philosophy of science about the end of 20<sup>th</sup> century, Stathis Psillos has done an important work, which I am sure will be appreciated by the future historians of philosophy of science. One may say that the suggested list of names is too idiosyncratic (in the Introduction Psillos himself admits that the list, which he has produced using "some advice" is "conventional" [Psillos, 2007; xi]) Well, one may easily find reasons to claim that. But it is out of question that Psillos could be eventually blamed only for the omission of one or another name,<sup>7</sup> and not for giving credit to those

whose names he has included in the dictionary. For there is no doubt that people like Peter Achinstein and Larry Laudan, David Lewis and Ian Hacking, Ronald Giere and John Earman, Clark Glymour and Nancy Cartwright have proven their influence on the field in the last quarter of the 20<sup>th</sup> century and, therefore, well deserve their place in the "A-Z" book.

Looking at the different categories of entries in the dictionary, one could ask why there are no entries for the different sub-disciplines of philosophy of science: philosophy of physics, philosophy of biology, philosophy of mathematics etc. The answer is simple: it is difficult, if not impossible to provide a useful general outline of any of these sub-fields in 300-500 words (what is the average length of an entry note in the "A-Z" book). And this answer points to one of the main strengths of the dictionary: the achieved "golden balance" between the number of the entries, the length of the entry papers, and the level of categorization of the entry subjects.

For whom is this book written? In his Series Editor's Preface<sup>8</sup> Oliver Leaman states that it is for those who want "to orient themselves through the subject." I hope this brief review has provided evidence that *Philosophy of Science A-Z* successfully completes the series' aim. But for me it is important to stress that Stathis Psillos' dictionary suggests much more than mere "orientation in the subject." I am not going to convince you in that, I strongly recommend to find out what I mean for yourself. Last but not the least, it should be said that the "A-Z" dictionary is not just another reference book in the large and sophisticated field of philosophy of science. It is a really indispensable guide to the field for both novices and professionals.

### References:

- Fine, K. (2002). *The Limits of Abstraction*. Oxford: Clarendon Press.
- Field, H. (1980). *Science without Numbers*. Oxford: Blackwell.
- Hitchcock, C. (Ed.) (2004). *Contemporary Debates in the Philosophy of Science*. Oxford: Blackwell.
- Kuipers, T. (Ed.) (2007). *General Philosophy of Science: Focal Issues*. Elsevier.
- Ladyman, J. (2002). *Understanding Philosophy of Science*. London: Routledge.

<sup>7</sup> I personally wonder at the lack of entries for Daniel Dennett (born 1942), Michael Ruse (born 1940), or Paul Benacerraf (born 1930).

<sup>8</sup> *Philosophy of Science A-Z* is one of the volumes of the series *Philosophy A-Z*, published by University of Edinburgh Press under the general editorship of Oliver Leaman. By now the series has issued 14 volumes including *Epistemology A-Z*, *Philosophy of Language A-Z*, *Philosophy of Mind A-Z*, *Ethics A-Z*, *Political Philosophy A-Z* etc.



- Losee, J. (2001). *A Historical Introduction to the Philosophy of Science*. Oxford: Oxford University Press.
- Machamer, P., Silberstein, M. (Eds.) (2002). *The Blackwell Guide to the Philosophy of Science*. Oxford: Blackwell.
- Newton-Smith, W. H. (Ed.) (2000). *A Companion to the Philosophy of Science*. Oxford: Blackwell.
- Okasha, S. (2001). *Philosophy of Science: A Very Short Introduction*. Oxford: Oxford University Press.
- Psillos, S., Curd, M. (Eds.) (2008). *The Routledge Companion to Philosophy of Science*. London: Routledge.
- Rosenberg, A. (2000). *Philosophy of Science: A Contemporary Introduction*. London: Routledge.
- Sarkar, S., Pfeifer, J. (Eds.) (2006). *The Philosophy of Science: An Encyclopedia*. London: Routledge.

## INFORMATION FOR CONTRIBUTORS

---

### ***Aims and scope***

The *Balkan Journal of Philosophy* is a peer-reviewed international periodical, academic in spirit and wide-ranging in its potential readership. The aim of the journal is to publish high-quality papers on the most current problems and discussions in philosophy. No philosophical field or expertise is excluded a priori. Special attention will be devoted, however, to the treatment of these problems in the Balkans and South-Eastern Europe, and to their influence on the development of philosophy in this region. There will be two issues per year: one devoted to a specific topic and a second one without thematic limit.

### ***Manuscripts***

The *Balkan Journal of Philosophy* will publish original papers (up to 6000 – 8000 words), short discussion papers (up to 4000 words), and book reviews (up to 3000 words), in English. British English or American English spelling and terminology may be used, but either one should be followed consistently throughout the article.

Manuscripts should be prepared in Word for Windows, formatted in A4 or US Letter, leaving adequate margins on all sides to accommodate reviewer's remarks. Double-space all material, including footnotes and references.

Submitted texts should include short abstracts (up to 150-200 words), 4-6 key words, and should be prepared for a double-blind reviewing procedure. Authors are, therefore, requested to place their name, affiliation, address, and e-mail address on a separate page. Self-identifying citations and references in the article text should either be avoided or left blank when manuscripts are first submitted.

### ***Submission procedure***

Authors should submit their manuscripts by e-mail to: [balkanjournalofphilosophy@gmail.com](mailto:balkanjournalofphilosophy@gmail.com)

### ***Legal requirements***

Submission of a manuscript implies that the work has not been published before; that it is not under consideration for publication anywhere else; and that its publication has been approved by all co-authors, if any. Authors wishing to include figures, tables, or text passages that have already been published elsewhere are required to obtain *permission* from the copyright owner(s). The publisher will not be held legally responsible should there be any claims for compensation.

### ***Annual subscription***

Individuals: 30 EUR

Institutions: 45 EUR

*Payment should be made by bank transfer to:*

UNICREDIT BULBANK, Sofia  
12, Alexander Batenberg Street  
Sofia 1000, Bulgaria  
IBAN – BG92UNCR96603410332105  
BIC UNCRBGSF  
Address: Institute for Philosophical Research  
6, Patriarh Evtimij blvd.  
Sofia 1000, Bulgaria