Contents

ARTICLES

BOOK REVIEWS

REPORTS

The Role of the "Private" in Inter-Gender Misunderstanding[1]

ONDŘEJ BERAN

ABSTRACT: In this paper, I use Wittgenstein's private language argument for reflecting on some folk-linguistic misconceptions. In Section 1, I show that elements of the private language semantics inform common ways of looking at some situations referred to as "misunderstandings". I suggest that it would be appropriate to conceive of the alleged misunderstandings as practical attitudes of mistreatment. This suggestion is explored in Section 2, which is devoted to a commonly assumed prominent example of the problem: the so-called inter-gender misunderstanding. It is believed that men and women use language in systematically different ways, as a result of which they do not understand each other properly, because they miss what their interlocutors "mean". The conceptual apparatus of mentalist semantics presumed here is abused in order to advocate morally reprehensible actions against women. In Section 3, I suggest that the Wittgensteinian accounts of language and mind offer arguments for denying private conceptions of understanding on the grounds of both philosophy of language and ethics.

KEYWORDS: Misunderstanding – private language – gendered languages – sexual violence.

The aim of this paper is to argue against certain folk-linguistic misconceptions. Namely, that what a person's utterance means has to do with how

she means her words, where "meaning something by one's words" refers to an internal, mental procedure. On this view, one misunderstands what another says because one has not established an access into the other's mind. In addition to that, as far as the mental is defined as an essentially hidden or secluded domain, there can be no such access apart from what the speaker manifests, willingly or not, by her words, gestures, actions etc. This kind of access into the other's mind is thus essentially indirect and open to uncertainty and to skeptical doubts. Since the behaviorist turn has begun (and ended), not many a philosopher professes his or her allegation to this theory, at least not in such a straightforward form that I sketched here. Various distinguished names from the *history* of philosophy were, however, its representatives, such as Augustine, Descartes, or Locke.

Not only is this a rather problematic philosophical theory about how language, its meaning and its relationship to what people are thinking (and may not be saying aloud) work. As far as it is inherent to some popular ideas of what *understanding* (or misunderstanding) of another looks like, it is worth challenging with respect to potentially *harmful* practice following or expressing this view. It is in this sense that I focus on it in the text. Specifically, I am interested in popular self-help guidelines as well as more scholarly theories about inter-gender (mis)understanding. If the ways men and women think differ essentially (be it due to different structures of their brains, or due to upbringing), their mutual understanding inevitably faces serious problems: they cannot have access to what the others' words "truly" mean and at best they are only guessing. The assumption that unhindered understanding between speakers of opposite genders is in fact impossible can be easily abused. The best-selling author John Gray can be used as a representative of the theory of separate gendered languages, but I will focus principally on the works by Deborah Tannen that strive at presenting this view in a more scholarly and systematic shape (see Tannen 1996; 2009).

A powerful attack on the above philosophical conception of language has been launched by Wittgenstein in his private language argument. Further elaborations of the later Wittgenstein's philosophy of language can provide certain remedy for the related misrepresentation of inter-gender linguistic relationships as well. I will thus depart from Wittgenstein's argument as my initial point.

Section 1 recapitulates the private language argument and its reach and shows that elements of the private language semantics, criticized by Wittgenstein, inform common ways of looking at situations referred to as "misunderstandings". I suggest that it would be appropriate to conceive of the alleged misunderstandings as practical attitudes of mistreatment. This suggestion is explored in Section 2, which discusses the prominent example of the so-called inter-gender misunderstanding. I argue that the conceptual apparatus of mentalist semantics presumed here is easily abused in order to advocate morally reprehensible actions against women. In Section 3, I suggest that the Wittgensteinian accounts of language and mind offer arguments for denying private conceptions of understanding on the grounds of both philosophy of language and ethics.

1. The private language argument

Wittgenstein's private language argument concerns an imagined language designed to record a person's inner experiences that only show themselves to her (her private sensations). This language thus cannot be understood by anyone else. Its signs cannot be given any public definition. Only the speaker herself can decide whether the experience E happened, hence whether she should use the sign "E" to refer to her experience. Nobody else can distinguish between correct and incorrect utterances "E!" (Wittgenstein 2009, §§ 243 ff).

However, playing a language game is a *rule-governed* activity; the normative distinction between correct and incorrect must hold independently of individual players. If there is no outside corrective authority – requiring a community – then whatever seems right to the lone player *is* right. But such a scenario contradicts the very concepts of "rule", "correct" or "mistake"; one cannot follow a rule privately (Wittgenstein 2009, §§ 201ff). The semantics of our language thus cannot be based on the foundation of naming private (mental) objects. Such is *a* plausible way of summarizing the argument in brief.[2]

---

[2] I stress here Baker & Hacker's (1984) "normativist" reading. Alternative readings are open too: most notably that of Kripke (1982), who links privacy with social deprivation and opposes the idea of private language on similar grounds as Hume's anti-

Wittgenstein's argument is sometimes interpreted in a narrow way (cf. Gert 1986; Sussman 1995): as an argument in favor of the thesis that, literally, private language is impossible. However, the thought experiment itself documents the possibility of diarizing one's inner experiences. Anyone can try it. The problematic point is to classify making such records as a full-fledged *language* activity: as a *game* with properly working *rules*. Private exercises *are* possible but paralinguistic. Language, on the other hand, is an essentially open space every speaker can enter.

Certainly, there are language territories that are more opaque and more treacherous. Talking gently and with insight to someone with profound personal worries is *not* a language territory that the vast majority of speakers can navigate smoothly. It is difficult to grow familiar with it. But this difficulty does not amount to impossibility *by definition* (characteristic of private language attempts). The difficulty accompanying contexts like this is also connected to the utmost importance we assign to them. Other people are often opaque to us and we can cause them great harm if we are not perceptive of what goes on in their lives. They matter to us and their importance has to do with the desirability of a proper appreciation of what goes on in their lives. These dimensions are absent from the diarizing of private experience; according to Wittgenstein's description, it is crystal clear that such a hobby need not be of any interest to anybody else. Access to it by outsiders is also impossible, rather than merely difficult.

Apart from the refusal to label private enterprises as language, Wittgenstein's argument has a more important point: that no relevant philosophical analysis of language can be provided in private terms. The polemics are directed against semantics building upon inner ideas, intentions or intuitions, represented by Descartes or Locke, but also the early Husserl. However, the idea that the meaning of language terms is based on something

skeptical arguments do. More recently, Mulhall (2008) offered an elaboration on Baker and Hacker: the possibility of private language should not be dismissed *a priori*, on the basis of its contradiction of the only reasonable conception of grammar. For the private linguist, this response would be unacceptably blunt, because the pitfalls of her proposal are *latent* to her. Instead, her proposal should be examined patiently "from inside": only by thinking the proposal and the private practice through does one enable the private linguist to see the absurdity. According to Mulhall, this is what Wittgenstein does in Wittgenstein (2009, §§ 258ff).

within speakers' minds and exclusive to them is typical of many *folk-linguistic* intuitions: what the words one utters *mean* is expected to be what one *means* by them. Even though this assumption proves unclear, when thought through, we often do adopt an *attitude* of caution and mistrust towards each other: if there is friction or misunderstanding, there is a strong sense that the speaker's intention should be given *priority* over the listener's interpretation. This priority is reflected in our practice: it is the *speaker* to whom the listeners direct the expectation of clarifying what an unclear utterance is supposed to mean. This assumption has serious implications.

Let us consider a familiar situation: the inability to express oneself in such a way that one is understood by others "as one intended". An artist may feel uncertain about the successful communication of her message (sometimes to the point of obsession, e.g., in Virginia Woolf's case). Or, more commonly, in an online debate one can find herself constantly repeating "That's not what I meant; let me put it in other words", but without reaching the desired end even after multiple attempts. Here, the "what I meant", i.e., the alleged true meaning of the utterance, is dissonant with how the others seem to understand it. We all sometimes reflect on debate situations so that "what I meant" has not been properly received, despite all efforts. In these cases, the misunderstood speaker may be tempted to allude to her neglected private domain: "I had my point in front of my inner eye, but nobody managed to see it, despite all my efforts."

Typically, my interlocutors have *an* understanding of my utterance, only this understanding seems "flawed" to me. As far as what I said offers the possibility of developing it further in the *course of talk* (cf. Rhees 2006), it is *there* that the meaning of my utterance shows itself (its possible, meaningful use is reflected here), not in the failed speech intention. Those who complain about being systematically misunderstood *cannot* distinguish between what they mean and what they only think they mean, because they themselves claim to be the only person capable of understanding their point. For a person who has the feeling of failing repeatedly in communicating something she perceives as highly important, there is no difference whether the others only *happened* to fail to understand her attempts or if her point is essentially impossible to understand. If (as it could have been in Woolf's case) one refers to an inner intention that *eventually* remained blocked from *any* outer interpretation, the

reference makes – from the viewpoint of the private language argument – no sense.[3]

The speaker's claim that this failure has never been overcome amounts to her (implicit) acknowledgement that she cannot be sure whether she does not only think she knows what her hidden intention means, if anything. Just as with private language. At *this* point, there is no difference between *has never been understood* and *cannot be understood*. The only difference lies in others' response: the situation of a speaker who could (and should) have been understood, but never was, calls for pity or sympathy, while someone who wastes her time on what cannot be understood (diarizing one's private experiences) is responded to with indifference or disappointment that she is not interested in something else.

This difference is reflected in Wittgenstein's conception of the foundational attitudes towards other humans (Wittgenstein 2009, II, §§ 19-26): they are "souls" and on the basis of this attitude of ours towards each other we can have conversational exchanges of the kind that it makes sense to have with other beings speaking and understanding the human kind of language. The exchange of factual information – which may sometimes fail – proceeds only on this ground. We do not *infer* that others are beings (capable of) wanting to say something from the scrutiny of their utterances (cf. Winch 1980/1981); we (sometimes) scrutinize their utterances because we *trust* that they are human beings' utterances supposed to convey a sense (see Cockburn 2014).[4]

---

3   Insofar as individuals' inner worlds cannot be spoken of meaningfully, they are insignificant, as good as nothing (see Wittgenstein 2009, § 304). Mulhall's (2008) "resolute interpretation" suggests that such inner objects do not exist at all. This seems stronger than a more literal reading of Wittgenstein: that it makes no sense to postulate such things. But even though a speaker can only think she has access to something that others misunderstand, a certain degree of reality can be ascribed to this "something". It plays a role in motivating the speaker's frustrated and odd behavior that others *can* observe, though they cannot explain it.

4   Cockburn's view of trust as foundational to communication is parallel to the more widely-known principle of charity discussed by Davidson (1973/1974). There are, however, differences. Davidson's principal focus is epistemological: the principle of charity ensures that we understand other's utterances as being coherent overall and expressing what the speaker believes to be *true*. Cockburn agrees, but he adds that, first of all, "relating through speech is central to the human way of *being with* others"

If the intention that failed to communicate was essentially private, then, whether one succeeded in communicating it or not, it would still mean what it was supposed to mean (it would still mean *something*). That is different if the "failure" means a disruption of the foundational interpersonal situation of intelligibility and trust. Then there is nothing beyond the pitiful rudiment: pitiful because of the intensive sense that something meaningful never came to be, though it could have. Failures in understanding must be recognizable as such on the background of the underpinning trust in mutual intelligibility. The presumption of a *preceding* meaning makes understanding and misunderstanding two alternative but *equal* scenarios of what can happen with the meaning. But misunderstandings are *not* routinely present as an equal alternative, depending on the result of the complicated procedure of interpreting signs (cf. Wittgenstein 2009, §§ 503f). To be mistaken is only possible if I am otherwise already in foundational agreement with others about the way we think (cf. Wittgenstein 1969, § 156).

Various elaborations on "what I meant" enter into conversation. For instance, the speaker can state that the public understanding differs significantly or slightly from her intention and she can explain the details of this difference. But she must convey her sense of being misunderstood in a way inviting others to see that there is something about her utterance that is clearly different from how they understood it previously. They are a *part* of the clarification process; a variety of meaningful steps must be open to them, differentiated according to whether they understand the speaker properly or not and how it matters. They must not be excluded from the process as incapable of seeing and appreciating the intention. Otherwise the reference to the intention couldn't enter meaningfully their conversation.

The responses given by others can have different forms; some of them accommodate even the persistent high opacity of the speaker's intention. Even if the speaker's feeling of being misunderstood persists for a long time and her references to her inner intention remain unspecified (open to

---

(Cockburn 2014, 52, my italics). Language is expressive of our *engagement* with each other; some cases reported as failures in understanding are, as I will try to show, more properly understood as failures in engaging with each other as with one's human "neighbors". These failures involve dimensions that need not primarily have to do with what the other believes to be true.

doubt as to their meaningfulness), her worry and distress are real, with full-fledged outward criteria. That is a situation calling for patience, sympathy and attention. These responses require willingness on the recipients' part. But they need not be thus willing, out of idleness, ignorance or ill will. The surmises that communicating essentially involves the possibility of failure (because it starts "inside"), sometimes inevitable as a result of the attempted communication of something incommunicable – assuming that some genuine objects of communication are *essentially* incommunicable – these surmises are synergic with idleness rather than attention.

Language and understanding is, however, a matter of practice. To understand someone in a certain way includes attitudes, actions and interactions (Wittgenstein 2009, § 7). As Davidson (1973) notes in his account of interpretation, in understanding someone I ascribe to them feelings, beliefs, qualities and attitudes and *treat* them as persons who have such feelings, etc. Such *treatment* assigns to the person who is being interpreted a role in the shared practice, implied by the interpreter.

The frustration of the speaker who feels "regularly misunderstood" is often *expressed* in terms of understanding: "The others still don't *understand* what I wanted to say." As I tried to suggest, "what I wanted to say" is – in the cases where the reported misunderstanding persists – a problematic concept to make sense of. It can be fruitfully applied to cases in which the speaker is *subsequently able* to comprehensibly relate her intention, but it fails in cases where she is unable to do so or where nobody listens to her. The common description refers to the private domain: others do not understand me because they are not able to see what is in my mind.

If both success and failure were equal alternatives as results in the enterprise of establishing a connection with the inner, whence the feeling of frustration? The frustration indicates that understanding is, in an important sense, *due*. But understanding in the folk-psychological sense described above *cannot* be due: it is essentially unwarranted. If understanding is an insight into the inner, we could not justly feel frustrated by not being understood. The frustration can only be justified if we do not get what *is* due. We cannot be entitled to understanding in the sense of insight into our minds, but we are entitled to be understood if that means a particular kind of treatment – to be treated in a distinctly *human* manner. Unlike misunderstanding, mis*treatment* is easily shown to be something that makes one

feel *justly* uncomfortable. If the notion of humanity concerns the community of *trust* in mutual intelligibility among beings primitively interested in making themselves intelligible and acknowledging the importance of intelligibility, then the loss of intelligibility – an abandonment of interest, or an indifference to the endeavor – involves something of a loss of humanity. (This is a weak concept of mistreatment: addressing one's interlocutor in a way that neglects or frustrates her trust in being understood. I will endeavor to make it clearer in Sections 2 and 3.)

This may be a reason for seeing mis*treatment* as explanatorily more foundational than misunderstanding. It also helps us to see how serious cases of misunderstanding are ethical issue as well. Clearly, a lack of full attention to each other, accompanying most of our everyday misunderstandings – including innocuous cases – need not be downright immoral or wrong.[5] However, the situation is different if one *constantly* overlooks the other's frustration (because it may suit one's interests) and is in a superior position of power over the person striving to be understood. Here, to explain the feeling of being regularly misunderstood in terms of the opposition between private intentions and public interpretation is not only philosophically misplaced, it overlooks vitally important dimensions of the interplay between the two. Adherence to the idea of understanding as establishing the link with the private, secluded inner only exacerbates the harms related to the inequality of power. The next section will discuss and exemplify potentially serious implications of the interpretation of a particular kind of "persistent misunderstanding" based on the private idea of language.

2. Inter-gender misunderstanding

I have argued that to see misunderstanding as missing what the speaker meant is misplaced and it is better replaced by focusing on mistreatment. The common talk of misunderstanding includes the variety of two mutually misunderstanding partners. A widely suggested case of a mutual persistent

---

[5]   Although Murdoch (1970) argues, the roots of an *unrealistic* vision of the other, failing to appreciate her full humanity as being equal to mine (and thus failing to give her what she is morally due), are located right here.

misunderstanding, intensively discussed in both scholarly and popular literature, is the gender opposition between men and women.

There is a claim of underlying *systematic differences* in their language practices. It is a widely held belief that men and women use language in different ways. This claim has been made by many authors; it has also been variously interpreted.[6] Only in the 'seventies did a more systematic exploration of gender differences in language use begin with the publication of Lakoff's groundbreaking *Language and Woman's Place* (see Lakoff 1975/2004). According to her, the characteristics of "women's language" include indirect techniques, a higher number of polite, correct and excusatory forms, tag questions, diminutives or intensifiers. These speech forms preserve and reproduce the inferior social position of women by articulating and codifying their weakness.

Lakoff uses this linguistic analysis for a feminist critique of the *artifact* which she claims "women's language" to be. And she concludes that if we are able to analyze the forms of language preserving women's subjection to men, it will enable us to change it (see Lakoff 1975/2004, 102). However, these observations – criticized and revised by many scholars, including Lakoff herself – have grown into the widespread and popular picture of the *natural state of affairs* being that there being two distinct gendered languages, which are hardly alterable.

As far as I can see, there is no clear agreement whether the roots of this state are biological or rather cultural.[7] But it is not up to a philosopher to decide the status of a scientific theory. I will confine myself to a discussion of certain elements shared by both views. According to the biological interpretation, the difference in speech styles is clearly considered irreversible. However, culturally acquired grammatical conventions are equally difficult for us to alter as natural and inborn traits, as Wittgenstein (1977, 76) points out. According to the cultural interpretation, children of opposite sexes are nurtured in different ways from the beginning, spending time in different environments and involved in different activities. They are

---

[6]   It was mostly men who were considered to create or preserve language quality and women to neglect or even corrupt it. For a detailed expository survey see Coates (2004, 10ff); and also Cameron (2008, 24ff).

[7]   See the arguments pro in Baron-Cohen (2003, Chapters 7 and 8) and contra in Cameron (2010).

thereby guided to different behavioral standards and they grow into adults behaving like people of different cultures who understand each other only with difficulty or only think they understand each other (cf. Maltz & Borker 1982).

Maltz and Borker's thesis was elaborated by Tannen (2009), with an emphasis on language theorizing. According to her, the languages of men and women represent disconnected systems. Their misunderstanding does not concern the literal meaning of words; they use different interpretive frameworks, ascribe motivations their interlocutors would deny, adopt misguided stances, and infer "meta-messages" incompatible with what the speaker really wanted to express. That results in unexpected, unintelligible and undesired reactions.[8]

Although Tannen does not use the label "private", her thesis of two mutually unintelligible languages is reminiscent of the private accounts of semantics, including their unfortunate consequences described by Wittgenstein. Tannen presupposes that 1) each group speaks differently, because the speakers from one group have only poor (if any) access to the speech intentions of speakers from the other group, which are essentially exclusive to them and accessed by introspection. It means that 2) one is not oriented competently in the discourse of the interlocutor from the other group and its rules, and 3) whenever she attempts to move on such poorly understood grounds, she relies upon what *seems* correct to her, without any instrument for distinguishing between the correct and the seemingly correct. (As Wittgenstein showed, "language" based on foundations of this kind could not work in practice in the way our language does.)

Tannen's claim of mutual misunderstanding is based on her hypothesis that what utterances mean is based on how the speakers "meant" these utterances and that their meaning procedures reflect different *mental conditions* resulting from the different life situations and experiences of the groups, which are incomprehensible to one another. *Within* each group, its members understand each other thanks to the shared mental background of their speech. The agreement is thus not based on an *external* foundation such as following shared rules, but follows from an internal prerequisite, a

---

[8]   Even before Tannen's popular book, similar results were presented by more scholarly authors, e.g. Treichler & Kramarae (1983) or Eckert (1990).

kind of "pre-established harmony". The reason for the supposed communication problems between groups is the same: the clash of mutually inaccessible semantic constitutions in the *absence* of this harmony.

Women's language thus prevents its being understood by male speakers in the same way Wittgenstein's private language prevents its being understood by any outsider. There is more than one speaker of the female language; it is thus not private in the exact Wittgensteinian sense. But since the source of the *intra*-group agreement does not lie outside, in the speakers' interactions, they too may only think they understand each other. A man supposes that another man's words mean what he takes them to mean not because they are both governed by the same independent standard, but because he is counting on the agreement of their inner backgrounds, which he can never guarantee.

However, outsiders cannot even count on this presupposed, if unwarranted, shared background. The mechanism of the constitution of the meaning of utterances by speakers of one gender group leaves, *pace* Tannen, the other group's speakers inevitably in total dependence on uncertain guesswork. A woman's words mean what *she means* by them, and that is determined by her particular female mental patterns of thought. A man cannot have any access to that; whenever he thinks he understands what her words mean (what she means), he cannot be sure whether he does not only think he understands.

Various objections can be raised, both conceptual (meaning conceived as being separated from pragmatics and from how recipients understand the utterances, etc.) and empirical. That does not mean that the situations presented as inter-gender misunderstandings are not real. Only the explanation is debatable. Tannen offers something similar to the "private" view of the regularly misunderstood speaker: her intention is constantly misinterpreted, partly because there is an interpretation offering itself outside, which is easier than fighting one's way through to "what she meant". Similarly, Tannen argues that men have available a substitutive interpretation of what women say and *vice versa*. These interpretations satisfy the needs of each interpreting community, so nothing motivates its members to undertake the labor of insight.

Tannen clearly ignores the real workings of language. Language is a space constituted by confrontations and interactions. It is not just what we *say* but a rule-governed complex of what we *do* (including what we say),

rooted in a richness of practical contexts (Wittgenstein 2009, §§ 241f). To surmise that all speakers' collisions are in essence only dissonant translations of (declarative) sentences is to remain on the surface (Wittgenstein 2009, §§ 65f). An adequate understanding of the alleged male-female miscommunication must reach further than to making "dictionaries" that translate between "male-ish" and "female-ish."

Tannen's strategy is widely anecdotal, as Cameron (2008, 87f) documents in her critique of John Gray and Tannen herself: if a man "understands" a woman's question "Could you empty the trash?" as an information query to which the appropriate answer is "Yes, I am capable of doing that" and not as a request, this is not "misunderstanding", a failure to gain an insight into her speech intention that he cannot understand, until she asks him directly: "Please, empty the trash." (It is disorders on the autistic spectrum that might cause troubles with understanding more than the "literal" content of an utterance.)

If someone reacts thus, it is usually not a case of mistranslation, but of a clash of interests: it suits the man's interest to wear the mask of stupidity so as to avoid the bothersome work. Generally, dissenting "male" or "female" language forms may turn out to be expressions of dissenting positions of interest (Wittgenstein calls the complex standpoints from which language games are played "forms of life").[9] These positions can differ considerably, up to open opposition, but this conflict is *not* (just) a matter of translation.

Though the interpretation of male-female relationships in terms of private "meaning" is problematic, it is pervasive. It is thus not inappropriate to ask about its purpose. The hypothesis of the "misunderstanding" based on the membership of the gender group has a significant impact on interpersonal relationships. If there are only two groups and each human being belongs to one of them, then as a member of the one group to which I inevitably belong, I am confronted with a long-term established set of normative standards, encouraging me in certain expected courses, frameworks and styles of practice (rewarding me for engagement in them) and

---

[9]   These standpoints are established through practice-oriented activity (linguistic activities are mostly practice-oriented) and as such they can also be understood as forms of *power* (Lee-Lampshire 1999).

discouraging me from others (employing various sanctions). This division of labor need not serve my personal interests.

Though "women's language" allegedly displays more finely tuned techniques of maintaining the conversation, it does not thus follow that all female speakers favor this activity more than male speakers, or that they have naturally superior skills in this area. This menial job ("conversational shitwork"; see Fishman 1979 or Coates 2004, 87ff) may simply have been imposed on them as those who were not allowed to decide for themselves. Even if the distribution of skills between the groups is real and described correctly, the inference that "women are better at conversation – women should keep the conversation" incorrectly infers *normative* conclusions from *descriptive* premises. A tacit premise would be needed: e.g., if a task is easier for somebody, that person is obliged to take care of it. But this principle of labor division is far from self-evident; for instance, it would be at odds with learning new skills.

The myths of Mars and Venus prove "remarkably patronizing towards men" (Cameron 2008, 11). On the face of it, they only consider men to be less skilled speakers. But unlike other observations of lower skills, lower skills of this kind are not reflected in a practical disadvantage. The "less skilled" speakers end up with a more advantageous position than their female interlocutors because: being less able, they are *absolved* of the duty to keep the communication going and to take care of understanding each other. A "naïve" view of those who fail to talk to others and claim their right not to bother suggests a *moral* vice: indolence, insensitivity or disregard for others. A "scientific" or "objective" view hastens to correct the "naivety": this is a *natural* consequence of the deep sediment of different skill sets. It may be inferiority, but *intellectual* rather than moral.

Although theorists of gendered languages describe and discuss various structural differences, the example of conversation and communication skills is of special interest to us. It is directly connected to the presumption of the constitution of mental (private) meaning. When the recipients are men, the suggestion of privacy is made: they easily fail to understand women properly, because they do not see what was meant (the implicit assumption being: because it is inside the female speakers' heads). But not even the most enthusiastic advocates of this view can deny that some understanding in practice *has to* be and *is* established. Here the supposed higher communication skills of women enter the stage: the responsibility

lies on them. It would be difficult to explain how – if meaning has to do with private mental meaning intentions – one group can exculpate themselves of blunders, claiming their inability to look inside the others' heads, while the other group is expected to take care of efficient communication, *regardless of that*. But such *is* the expectation: while there is the assumption of a private barrier blocking the from-women-to-men direction, it is less so the other way round.

The asymmetric private language perspective has striking practical consequences, drastically documented in Ehrlich's (2001) extensive survey of sexual assault cases. It is not unusual for accused men to defend themselves with the argument that they didn't understand the woman's refusal to have sex. Although the argument that "No" means "Yes" or "Keep trying" sounds absurd in any other context, here, some defendants have successfully re-interpreted their immoral actions as a case of misunderstanding (*morally neutral*, if regrettable). What is more, the blame falls on the other side: it is the *woman* who has failed because she did not express herself clearly enough (Henley & Kramarae 1991; Cameron 2008, 89ff). The proposed asymmetry in communication skills in favor of women enables a shift in *responsibility* (*women* are primarily responsible for avoiding "misunderstandings") stemming from and again resulting in a real power asymmetry in favor of men.

The problem is not about different *views* of a situation. Each rival position is directly connected to each of the parties' *interests*. The complainant's claim that she has been morally wronged is the more transparent one, for it acknowledges that a certain reading of a situation (stressing that it has a moral bearing) calls for *differentiated* attitudes towards its participants. It recognizes that in some situations one party's interests legitimately deserve to be supported, while the other party should be rejected (or even punished). A part of the sense of this reading is that it is *not* depreciated just because the party claiming it is the one whose interests should be supported.

The alternative tends, on the other hand, to obscure the relationships of interest. It disregards the moral dimension and instead proposes the perspective of morally neutral epistemology (the inquiry into the private) and of the description of the state of affairs (the difference in linguistic skills). This reading is advocated as *unbiased* and in favor of *truth*; that it may also be in favor of one rival party is suppressed, as that could shed doubt on its

claim of impartiality. Despite the professed focus on the factual and epistemic dimension of the case, responsibility *is* assigned, in a way that suits the party advocating this view. Attention towards who is in the position of choosing the interpretive perspective and who profits from the choice is thus appropriate, even when the perspective that is offered claims neutrality.

From the defendants' perspective, understanding is an *unwarranted* result following the attempts to grasp "what the other meant", a result that may, or also may not come, especially in such a "difficult situation".[10] Additionally, whether mistreating the other is morally reprehensible allegedly depends on preceding proper mutual understanding: only then does the assailant have any chance of noticing his mistreatment at all – to see that he is violating the other's wishes.

The chain of logic of this argument is rather complicated, with several weak points: 1) men and women use language in systematically different ways (arguable at best); 2) this difference is inborn, therefore unchangeable (this is not, and perhaps cannot be, proven); 3) this difference is accompanied by different "meaning" procedures of gendered minds, mutually inaccessible as a result of different inborn mental equipment (possibly true, but practically irrelevant); 4) understanding people of the opposite gender is unwarranted, depending on a guess as to "what they meant" (refuted by Wittgenstein's argument); 5) understanding is a matter of decoding the language output (too narrow, at best); 6) understanding the other does not mean acting towards the other (treating her anyhow); 7) (some) cases of mistreating the other can be interpreted thus only if there was preceding

---

[10] In her critique of feminist epistemology, Diamond (1991, 1008) discusses an observation concerning situated knowledge: it cannot be assumed that one's understanding of another is prevented or inhibited by one's *emotions* towards the other. (In her example, it is animal trainers' *love* for their trainees: dogs, horses, etc.) An extrapolation into our example is possible: though the situation is heavily charged with emotions, the assailant's understanding of what his victim thinks or wants is not ruled out by his *emotional* state (be it rage, arousal or hatred). The parallel is not perfect: love may open a space for a deeper understanding that is unavailable for an impartial scientist, but the rage/arousal complex does not seem to mediate this capacity. But it suffices to point out that an emotional state cannot be an automatic excuse for a claimed "misunderstanding".

understanding which was deliberately neglected. Points 6 and 7 deserve some attention.

As I tried to show, understanding another cannot be separated from acting towards her. It can hardly be said that understanding in communication does not affect one's communication partner. First, the view equipping women and men with systematically different language repertoires allocates them to specified positions of practice, implicitly expecting each group to display its own skills and tend to occupations that allow them to make use of their talent. The expectation that men and women have different "beetles in their boxes" (Wittgenstein 2009, § 293) – "what they mean" – on which both how they understand others and how others understand them depend involves adopting different attitudes towards them, treating them differently. But, as Lakoff already pointed out, "women's language" is an artifact. Claiming an artifact to be natural – with the variety of further responses, such as indignant surprise when someone is not behaving "naturally" – produces pressure. It is supposed that women *should speak* a language articulating them as agents of a powerless and dependent nature,[11] whereas the language exclusive to men is, in contrast, apt for solving important problems "in the world". It is not a source of satisfaction that the female role is as impossible for men to occupy as the male role is for women. In effect, the references to the inner and the natural allow some agents to be denied competence.

The second, more specific worry is highlighted by the example I introduced: if morally reprehensible actions can be interpreted as depending on preceding understanding, then one's denial of having understood (with implicit reference to the gender-specific labor division) is an efficient way of exculpating oneself. But the rapist did *not* mis*understand* his victim. He mis*treated* (harmed) her, and by pretending to misunderstand he only aggravates the wrongdoing, because he blames the victim. Public campaigns fighting the culture of victim-blaming (rightly) point out that it makes the victims defend themselves when it is *they* who have been harmed. This is a special case of the same phenomenon: the victim carries an unjustified

---

[11]   That women's language is an artifact is somewhat quaintly documented by the research of fantasy line workers, which showed that they had to learn the ideal "women's language" as a role with which they did not personally identify, but that was demanded by their male customers (Hall 2005).

expressive burden – the duty to establish a link between "what she meant" and how she manifests it in a way transparent enough for the assailant to be able to follow it into her mind.

The picture of the boundary between "meaning" and understanding, partly permeable in one direction only, is useful for the current power asymmetry and is thus sustained by it. The denial of practical competences argued for in the "misunderstanding" line bears marks of ill will. Women urged to defend their behavior in rape trials do not fight philosophers (Locke or Husserl) presenting flawed arguments; they did not get stuck in a scholarly discussion. They face a broad, everyday practice that claims the support of favorable scholarly arguments and ignores adverse ones. The dirty trick lies in presenting situations of inequality in *factual* terms, whereby the ready accusation is avoided that what is going on here is just *morally* wrong.

3. Against the immorality of the "private" claims

I tried to demonstrate, on an example, the possible consequences of linking understanding to the ability to capture what the other "meant". If moral assessment of an action requires proper (preceding) understanding of the other in terms of reaching "what she meant", some actions avoid moral evaluation altogether because acquiring such understanding is by no means guaranteed. It may or may not succeed. If one *can* argue that one has not met this requisite for moral evaluation of an action in question, the results may be horrific.

If men and women do not understand each other because they cannot access what the others "mean", they act as if they were not bound by the same rules, since no rules can be guaranteed as being shared. To follow the same rule governing the meaning of an expression would mean to be able to follow the link between the expression and the mental meaning act. Elaborating further on Wittgenstein's example, since speakers of opposite genders do not participate in the "pre-established harmony" of shared mental backgrounds, they – unlike the private linguist who always thinks she gets it right – must always assume they do *not* get right what the others say. On the grounds of a Mars/Venus theory, they have no tool for distinguishing between grasping the others' utterances incorrectly and only thinking so.

Without access to the semantic standards established inside the others' minds, whatever seems wrong to them is such.

Therefore, wherever understanding the meaning of what the others say plays a role (i.e., practically everywhere), I cannot be properly bound by rules towards my interlocutor of the opposite gender because we only seem to talk to each other. I could only be her genuine dialogue partner if I could with certainty link her private meaning intentions to her language use. In our example, the assailant claims "I didn't understand what she meant, so you cannot blame me for doing anything wrong, such as forcing her into *unwanted* sex". If the moral evaluation of his action depended on proving his ability to grasp what she "meant", he would be quite right.

Certainly, advocates of the misunderstanding view want to maintain the links of mutual moral responsibility. It is said that no "fundamental differ-ence" found by science can be an obstacle to the necessary mutual tolerance (Baron-Cohen 2003, Chap. 1). The slogan "different but equal" is supposed to convey the same message. Unfortunately, only rarely do they seriously attempt to put forward arguments for tolerance that are equally detailed as their arguments for the difference and content themselves with a declara-tion.

An elaboration of these suggestions can draw on Lévinas' (1991) com-pletely independent notion of "ethics". Lévinas endorses that the Other is not fully intelligible to me. To "translate properly" what the Other thinks into my own language means only to reduce the Other to an abstract, sys-tematic idea of her that I am able to create. I would thus treat the Other in a way that assigns her a "suitable" position within *my* worldview. On the other hand, the ethical stance means not to treat the Other in a way that suits my ideas about her, but to let her be in and for herself (see Lévinas 1991, 43). In Lévinasian terms, one's gender counterpart is considered *in-exhaustible* by the interpretive tools one has available: as a private individ-ual over whom one has no power. Consequently, I should respect the Other not *despite* my lack of understanding of her as a person, but *because of* this lack.

Unfortunately, the Lévinasian concept of ethics proves similarly prob-lematic as the rapist's defense was. Indeed, it enables us to grant a person we do not understand the status of a moral subject and to make her thereby protected from *our* abuse. However, such a status only includes *her* right to be treated as a person with certain rights, but I have no leverage that the

other will treat me responsibly in turn – that I have the corresponding status in her eyes. Either way, one side is always left unbound by moral responsibility towards the other.

It seems to me that examples where issues of understanding overlap with issues of morality highlight that there are *multiple* reasons for dismissing private language semantics. The private language argument alone would be enough to show that the semantics of our language cannot be grounded in private intentions. Before I am able to think privacy, I am already in *public*: the source of the conceptual equipment necessary for conceiving skeptical worries (cf. Wittgenstein 2009, §§ 125ff). Wittgenstein's reflections on our foundational attitudes towards one other illuminate the problem from another side. The concept of "soul" – the role fulfilled by those towards whom my *interpersonal* attitudes are directed – implies that a soul is a fellow-being for whom it makes sense to feel sympathy, compassion, solidarity, etc. (cf. Cockburn 1990, Chap. 1)

The original German word for "soul" used by Wittgenstein (*Seele*) is ambiguous: it covers both the English "mind" and "soul". *Seele* is thus both a thinking entity and a *person*: a being of value, towards whom we adopt distinctly personal attitudes we do not adopt towards animals (though there is overlap) or things. The recognition of others as people – with whom I am in a relationship of mutual foundational intelligibility, involving a sense of the other's vulnerability – thus does not have to wait for determination of whether or what they think. On the contrary, if it appears uncertain what the other is thinking, this is parasitic upon the primitive recognition and can only make sense if there is such a primitive recognition (even then only rarely). Otherwise, we would not know what it means that the other hides something from us, etc.; we would not have such concepts as "hiding".

Of course, to recognize another in practice as a human being need not mean acting towards her with particular consideration. Phillips (1992) points out that there are *many* "attitudes towards a soul"; but they all differ in quality from the attitudes of different kinds. To pass un-Samaritanly by an injured man lying by the road means to be callous; there is something wicked about it. With passing by a broken machine, or by a piece of rock, it is not so. Callousness is an attitude towards another soul; we are not callous towards rocks even when we kick them thoughtlessly.

This is where the reason lies for seeing the defendants in rape trials as someone acting in a particularly horrendous way, rather than just seeing

the whole situation as an unfortunate incident. Rhees (2006, 148ff) points out that human communities are distinguished from animals by their members' understanding the lives they lead in terms induced by culture and morality. First of all, in terms of good and evil. Obviously, animals, too, display social complexity, companionship, etc. But it is distinctly human that people *endeavor* and *fail* in the respects of morality, or that they are sometimes altogether "dead" to the dimension of good and evil in their interpersonal relationships. The sense of something horrendous accompanying cases of human cruelty differs from aggression among animals. A crocodile does not kill a deer (or another crocodile) in the spirit of callous negligence of its personal preciousness that it (the crocodile) would or should normally recognize. There is no sense in talking about such recognition here, so there is none in talking about negligence. The rapist's self-defense, on the other hand, bears the meaning of something horrendous just because it is an expression of callous negligence, or, in Rhees' words, of being dead to the sense of good and evil, and to the evil spirit of his actions. He fails to recognize his victim's humanity.

It is a long way from the philosophical accounts of language and meaning in private terms to endings of this kind, but there is a connection. To think of meaning and understanding in private terms is not simply incorrect. This theory can also be exploited in many sorts of idleness, negligence or ill will. The suggestions of a systematic possibility of misunderstanding based on the different mental equipment of interacting agents should be dismissed, rather than seriously discussed in court. The rapist tries to present himself as Descartes' philosopher setting off on a long, difficult journey, and it is only at the end of it that he can say whether the others are not automata. But by shifting the burden of defense and blame onto the victim, he only shows himself as being dead to the sense of good and evil. (Anyway, he could hardly claim to be a philosopher: philosopher would not spare his or her own work, expecting that others should do the job.)

I suggested in the first section that addressing each other as human beings involves engaging with each other in mutual trust in each other's intelligibility, and that detriment to the trusted intelligibility amounts to a certain detriment to humanity. My point was not the Cartesian worry about other minds. It rather concerned the cases of weakened recognition of full humanity in someone who is not quite like me (a natural part of life situations where distinct groups – such as "men" and "women" – get stuck in

the clinch of competing interests). I may then be reluctant to address the other as someone with an equally rich inner life, as someone to whom the difference between the genuine and the seeming matters, as someone with the same depth of understanding and self-understanding as my own, as someone with truly deep or authentic emotions. This can be called a failure of understanding what the other is saying, but deep down, it amounts to detracting from the other's humanity.[12] Sexual violence, too, is an expression of a lack of interest in whether (an implicit distrust that) the victim is a human being with full depths of emotion and understanding.

I have tried to show that Wittgensteinian thinking offers tools for dismantling the potentially abusive attempts to interpret male-female miscommunication in private terms. There is no such thing as "misunderstanding" someone systematically; the arguments offered for the hypothesis of inter-gender misunderstanding would render *intra*-gender understanding equally insecure. Additionally, understanding each other (as persons) is intertwined with a variety of foundational emotional and practical attitudes. It is never only a theoretical task. To postulate the opposition of two mutually private domains and to claim one's natural misunderstanding thus amounts to avoiding one's responsibility to the other as a person. A moral issue is thereby mistaken for an epistemic one. We already understand each other somehow and that understanding involves a sense of personhood in others. The real problem is not how we can understand each other *at all*, but what can be improved between us and how. This is what the true cases of our opacity to each other call for: not for skeptical resignation, but for an effort to increase our mutual understanding. Arguably, this is also one point in feminist politics: *to the extent* that it is a *genuine* lack of understanding what is (partly) responsible for oppression, one should try to make others understand better rather than to sit with one's hands in one's lap, because by claiming the lack of understanding the matter is settled.

The naive notion of language based on private meaning intuition is incorrect not only in the context of semantic analysis, but also as an analytical tool for explaining the relationships between men and women. It facilitates sanctioning the actual inequality, but more so, it can also be abused to excuse blunt or intentionally immoral actions.

---

[12]   The suggestion in this paragraph derives from Gaita's analysis of racism (see Gaita 2002, Chap. 4).

Acknowledgments

References

BAKER, G. P. & HACKER, P. M. S. (1984): *Scepticism, Rules, and Language*. Oxford: Blackwell.

BARON-COHEN, S. (2003): *The Essential Difference*. London: Penguin Books.

CAMERON, D. (2008): *The Myth of Mars and Venus*. Oxford: Oxford University Press.

CAMERON, D. (2010): Sex/Gender, Language and the New Biologism. *Applied Linguistics* 31, No. 2, 173-192.

COATES, J. (2004): *Women, Men and Language*. Harlow: Pearson Education Ltd.

COCKBURN, D. (1990): *Other Human Beings*. Basingstoke: Macmillan.

COCKBURN, D. (2014): Trust in Conversation. *Nordic Wittgenstein Review* 3, No. 1, 47-68.

DAVIDSON, D. (1973): Radical Interpretation. *Dialectica* 27, No. 3-4, 314-328.

DAVIDSON, D. (1973/1974): On the Very Idea of a Conceptual Scheme. *Proceedings and Addresses of the American Philosophical Association* 47, 5-20.

DIAMOND, C. (1991): Knowing Tornadoes and Other Things. *New Literary History* 22, 1001-1015.

ECKERT, P. (1990): Co-operative Competition in Adolescent 'Girl Talk'. *Discourse Processes* 13, 91-122.

EHRLICH, S. (2001): *Representing Rape*. London: Routledge.

FISHMAN, P. (1977): Interactional Shitwork. *Heresies: A Feminist Publication on Art and Politics* 2, 99-101.

GAITA, R. (2002): *A Common Humanity*. London: Routledge.

GERT, B. (1986): Wittgenstein's Private Language Arguments. *Synthese* 68, No. 3, 409-439.

HALL, K. (2005): Lip Service on the Fantasy Lines. In: Cameron, D. (ed.): *The Feminist Critique of Language*. London: Routledge, 321-342.

HENLEY, N. M. & KRAMARAE, C. (1991): Gender, Power and Miscommunication. In: Coupland, N., Giles, H. & Wiemann, J. M. (eds.): *'Miscommunication' and Problematic Talk*. Newbury Park (CA): Sage, 18-43.

KRIPKE, S. (1982): *Wittgenstein on Rules and Private Language*. Cambridge (Mass.): Harvard University Press.

LAKOFF, R. T. (1975/2004): *Language and Woman's Place*. Revised and expanded edition. New York: Oxford University Press.

LEE-LAMPSHIRE, W. (1999): The Sound of Little Hummingbird Wings: A Wittgensteinian Investigation of Forms of Life as Forms of Power. *Feminist Studies* 25, 409-426.

LÉVINAS, E. (1991): *Totality and Infinity*. Lingis, A. (transl). Dordrecht: Kluwer.

MALTZ, D. & BORKER, R. (1982): A Cultural Approach to Male-Female Miscommunication. In: Gumperz, J. J. (ed.): *Language and Social Identity*. Cambridge: Cambridge University Press, 195-316.

MULHALL, S. (2008): *Wittgenstein's Private Language: Grammar, Nonsense, and Imagination in Philosophical Investigations, §§ 243-315*. New York: Oxford University Press.

MURDOCH, I. (1970): *The Sovereignty of Good*. London: Routledge & Kegan Paul.

PHILLIPS. D. Z. (1992): My Neighbour and My Neighbours. In: Phillips, D. Z.: *Interventions in Ethics*. Albany: SUNY Press, 229-250.

RHEES, R. (2006): *Wittgenstein and the Possibility of Discourse*. Oxford: Blackwell.

SUSSMAN, H. (1995): Maxima Moralia: Millennial Fragments on the Public and Private Dimensions of Language. *MLN* 110, 856-87.

TANNEN, D. (1996): *Gender and Discourse*. New York: Oxford University Press.

TANNEN, D. (2009): *You Just Don't Understand*. London: Virago Press.

TREICHLER, P. & KRAMARAE, C. (1983): Women's Talk in the Ivory Tower. *Communication Quarterly* 31, No. 2, 118-132.

WINCH, P. (1980/1981): Eine Einstellung zur Seele. *Proceedings of the Aristotelian Society*, New Series 81, 1-15.

WITTGENSTEIN, L. (1969): *Über Gewißheit*. Oxford: Blackwell.

WITTGENSTEIN, L. (1977): *Vermischte Bemerkungen*. Oxford: Blackwell.

WITTGENSTEIN, L. (2009): *Philosophical Investigations*. 4th revised edition. Oxford: Wiley.

Essentialism and Method[1]

FERNANDO E. VÁSQUEZ BARBA

ABSTRACT: This paper mainly addresses the relation between essentialism and philosophical method. In particular, our analysis centers on the anti-essentialist argument that proposed, given its essentialist bonds, the abandonment of the notion of method. To this end, we make use of the empirical evidence concerning essentialism provided by psychological research, which has shown that our proneness to essentialize is not a by-product of our social and cultural practices as some anti-essentialists have thought. Rather, it is a deeply rooted cognitive tendency that plays a major role in concept formation and so in our understanding of things. Thus, given that such inclination toward essentialism is certain to happen, we argue for a conception of method that, while not overcoming such tendency, avoids the presumed disastrous consequences feared by most anti-essentialists.

KEYWORDS: Cognitive bias – essentialism – method.

0. Introduction

In 2014, the website *edge.org* launched the debate on the following question: What scientific idea is ready for retirement?[2] 178 persons tried

---

      ✉ Fernando E. Vásquez Barba
      Department of Philosophy and Social Sciences, Faculty of Philosophy
      University of Hradec Králové
      Rokitanského 62, 500 03 Hradec Králové, Czech Republic
      e-mail: fernando.barba@uhk.cz

[2]   See https://www.edge.org/response-detail/25366

to give an answer, one of them, Richard Dawkins, said that essentialism is 'scientifically confused and morally pernicious' and so it must be abandoned. One of the reasons for its abandonment, according to Dawkins, is that it 'makes no sense in the light of evolution and other gradualistic phenomena' that is to say, essentialism is unjustifiable (incompatible, incoherent) in the light of other beliefs (e.g. evolution).

The same rejection of essentialism found its way into the realms of Philosophy in the 20th century to the extent of becoming a dirty word. As argued by some anti-essentialists, notably Richard Rorty, notions as essence are part of the representationalist sort of epistemology, which relies on an idea of the mind understood as a mirror that represents an external and independent world populated with 'intrinsic natures' that we must seek to gain true knowledge. Holding such view, Rorty said, lead us to believe that we need a criterion, a method, to determine which representations portray more accurately the world. In this sense, such refusal of essentialism involved the abandonment of other notions such as 'method' allegedly equally connected with the representationalist epistemology.

With this in mind, this paper aims at examining the relation between the notions of 'essentialism' and 'method' and showing that the claimed abandonment of such notions is based on some misconceptions about what essentialism is and how it is related to the notion of method. Beside the introduction and conclusion, the paper is arranged in these sections: 1) Essentialism and method; 2) Anti-essentialism against method; 3) Inevitability of essentialism; and 4) Method without going beyond essentialism. In section 1, we explore the links between essentialism and method. Section 2 critically examines the anti-essentialist argument proposed by Richard Rorty. In sections 3 and 4, we develop our argument. Namely, in the former section, we argue for epistemological essentialism based on the research findings from developmental psychology; and in the latter section, we look at a sample of essentialist theorizing in which the allegedly disastrous bonds between essentialism and method are rendered unproblematic.

1. Essentialism and method

Richard Rorty, in his *The Linguistic Turn*, remarked that with the purpose of bringing to an end the philosophical discussion about methods and

trying to turn philosophy into a science, many philosophers had proposed the adoption of a new method that is purported to limit the task of philosophy (cf. Rorty 1992). In doing so, most of them have claimed to bear no presupposition; however, as Rorty pointed out, no one of those philosophers has succeeded. It is because a philosopher's choice of a particular method is always determined by her metaphysical and epistemological assumptions. It might give us an idea of philosophy according to which philosophy is a matter of opinion, a discipline in which there are no specific criteria for the solution of philosophical problems and, in this sense, no knowledge can be acquired. For this reason, Rorty says, it would be more advantageous, for the sake of finding new topics to discuss, if we focus on unwrapping the presuppositions involved in the utilization of particular methods (cf. Rorty 1992, 2). If we take Rorty's suggestions seriously, we have to agree that every claim about the proper method in philosophy is always idiosyncratic and dependent on one's epistemological and metaphysical commitments.

One of those assumptions is the belief that things or objects have essences. The term 'essence' was introduced into the philosophical vocabulary by St. Augustine who used the term to translate the Greek *ousia*, which was already an important part of the philosophy of Plato and Aristotle. However, despite the many attempts to narrow the scope of the notion of essences, it is still a vague concept; that is to say, it has become an overworked word. With this in mind, without intending to provide a precise definition, the way we understand here 'essence' could be broadly characterized as the existing substance, entity, property or force that causes things to be the way they are, that is, what causes the emergence of other properties; which in turn sustains its identity (unchanging); and, determines its category. It must be mentioned that essences are thought to be intelligible, but sometimes also hidden or unobservable.

Moreover, the first attempt to formulate a complex theory of essences and so to lay the foundations of essentialism as a philosophical view is widely credited to Aristotle. As a result, 'essentialism' has come to designate the metaphysical view, asserting the existence of such entities and pointing to the need to distinguish between essential and accidental properties. Indeed, essentialism is, philosophically speaking, primarily a metaphysical doctrine whose supporters have been trying hard to answer

questions related to the place occupied by essences in the world, its existence, the role they play on the structuring of our world and its classification.

In the same way, there have been relatively recently some attempts to introduce a new particular kind of essentialism, highlighting the epistemological consequences of believing in those entities we call essences, namely epistemological essentialism.[3] The main idea behind epistemological essentialism is the old Aristotelian idea that knowledge presupposes the existence of something that is stable and intelligible, that is, an essence, which turns research or inquiry into an attempt to penetrate true nature, the essence, of things. It is in this way, undeniably metaphorical, that epistemological essentialism has been defined. However, we think that such way of defining epistemological essentialism does not do justice to the complexity of the phenomenon of essentialism as a psychological fact, that is, as intuitions about the constitution of things and their place in a particular group.

For this reason, the way we understand epistemological essentialism here is that it is the guiding *belief* that things have essences or invariable properties that make them what they are, without which those things (or group of things) would cease to exist. Such essences are usually thought to have a causal force that makes the world the way it is, defining membership in a kind and having thus a profound impact on the way we conceptualize, represent, and understand the world, knowledge, thought, language and culture.

It is important to emphasize that we are here concerned with essentialism as a *belief* that has profound consequences on the ways humans perceive and evaluate the world and the way knowledge and inquiry have been construed. Therefore, it means that we are not seeking to make a case for or against the existence of essences, that is, we are not concerned with challenging or supporting any metaphysical claim, but rather our main concern is an epistemological sort of essentialism. Furthermore, we do not think that our arguments for epistemological essentialism support the idea that there are essences somewhere in the world or that they constitute the

---

[3] It also has been called methodological essentialism, but this name never enjoyed enormous popularity (cf. Popper 2002).

objective reality. Certainly, there is no evidence that essences exist, yet there is growing evidence demonstrating that we tend to believe so.

Epistemological essentialism, we argue, provides a comprehensive framework that describes how laypeople, scientists, as well as philosophers, evaluate and understand the process of seeking knowledge. Undeniably, such tendency towards essentialism[4] has been around, either implicitly or explicitly, in the philosophical literature, shaping philosophies for quite some time. It has been giving shape to how philosophers understand inquiry or the search for knowledge. Sure enough, one of the ways in which essentialism affects philosophers' representation of inquiry is that it needs to be done according to a fixed and systematic procedure, that is, it requires a method. A method is always conceived, broadly speaking, as a definable pathway with a constant or definite pattern to be followed when seeking knowledge.

In the light of essentialism, a method appears depicted as an inherent feature of knowledge, without which we would be, paraphrasing Plato, like a blind man. Also, essentialism manifests itself more clearly when philosophers think of true knowledge as the knowledge of such underlying properties that make things what they are, and a method is what makes such knowledge findable. In the same way, the very idea of 'pattern' or 'regular intelligible form' is essentialist in itself. In this sense, we argue that essentialism seems to be an assumption underlying any account of methods in philosophy. In other words, any proposed philosophical method is rendered workable only by adopting a sort of epistemological essentialism and not otherwise.

For example, Plato conceived of his method of dialectics as the only means 'to find its way to the very essence of each thing' so, the one who is skilled in dialectics is "the man who is able to exact an account of the essence of each thing…" (Plato 1969, VII, 532a-534b). Similarly, Aristotle proposed his method of demonstration, which, when put into practice, "…enables us to pursue knowledge of the essence of a thing" (Aristotle 1928, I, 14).

The same nexus between method and the epistemological essentialism can be found in the writings of some medieval theologians. Although

---

4   Hereafter, 'essentialism' and 'epistemological essentialism' are used interchangeably.

sometimes different words (ordo, via) are used, they all have the same connotations clustered around 'method' (procedure, means, way, road and so on). Such is the case of the idea of 'Itinerarium' used by St. Bonaventure of Bagnoregio in his *Itinerarium Mentis in Deum* (cf. Bonaventure 2002).

With this in mind, in his Itinerarium St. Bonaventure describes the various stages which a pious philosopher's mind must traverse through to know the being from which everything emanates, God, that is to say, the way a mind has to go to acquire the knowledge of what is essential to all things. However, for Bonaventure Philosophy alone does not suffice to achieve the knowledge of what is essential; faith must enlighten it. Then, a philosopher who attempts to know the essence, the first principle of everything, must make an effort not

> to read without unction, speculate without devotion, investigate without wonder, examine without exultation, work without piety, know without love, understand without humility, be zealous without divine grace, see without wisdom divinely inspired. (Bonaventure 2002, 39)

God, according to Bonaventure, is the one who enlightens our minds, enabling us to know himself through the vestiges of him that are in the world outside and inside us.

For Descartes, a method is, above all, the right path to follow when attempting to acquire certain and evident knowledge, which is incontrovertible by virtue of its essential nature. Descartes thought that through doubting we could get to know the essences of things and such fundamental knowledge provides a basis for any knowledge. In fact, Descartes claimed that by the use of his method he had come to know his very essence as a human being, which resides in thinking. Descartes made it clear when he said that after doubting the truth of the things he believed that "I thereby concluded that I was a substance whose whole essence resides in thinking" (Descartes 2008, 29). In this way, the phrase 'I am thinking therefore I am' that has become an unmistakable symbol of Descartes' philosophy, delates his essentialism. Likewise, Descartes pointed out a method is a mandatory requirement for the acquisition of such sort of knowledge (cf. Dear 1998).

The essentialism is so pervasive that essentialist characteristics of methods can be also found in recent times. That is the case, for example,

of analytical philosophy movement whose members assumed that philosophical problems could be resolved through analysis of language. So, the analysis was introduced as a method for clarifying the meaning of words and sentences and thereby solving (or, as the case may be, 'dissolving') philosophical problems. Although it has been said that it is hard to find just one model of how such analysis was performed and construed among members of the analytical movement, we can certainly find essentialist assumptions underlying the idea of analysis in the early days of the analytical movement. Significantly, philosophers such as Russell and the early Wittgenstein presupposed that language has a hidden, intrinsic or fundamental logical structure that could be discovered by a process of breaking down language (analyzing) into its constituent elements (words, propositions, etc.). The outcome of such process was purported to be a clarification of the meaning of words and sentences, and a subsequent insight into philosophical problems (see Hylton 1998, 38; Beaney 2000, 97-98).

On the other hand, as G. Hallett says, Kripke's and Putnam's theory of rigid designation are remarkable examples of essentialist theorizing. For the idea of rigid designation advanced by Kripke and Putnam meant that words such as 'water' (a natural kinds term) designate H_2O regardless of whatever 'superficial properties' water may have or may cease to have. In other words, this is also a case of essentialist rigidity (cf. Hallett 1991, 19-22). Although it may be true that Kripke's and Putnam's method may differ from the conception of analysis of the founding fathers of the analytical movement, it is nevertheless true that some of their theories involve essentialism as an inevitable part of it, which makes them typical examples of essentialism within the analytical tradition.

2. Anti-essentialism against methods

Such omnipresence of essentialist theorizing has found many critics among philosophers, resulting in a firm rejection not only of essentialism but also of the idea of method.[5] One of the critics is Richard Rorty who

---

[5]   For further details on Rorty's rejection of the idea of method, see Rorty (1991, 63-77).

claimed that the belief that there are essences or that there is a distinction to be drawn between intrinsic/accidental properties forms a crucial part of the metaphysical dualism, so characteristic of the Western tradition of metaphysics. Likewise, such belief, says Rorty, is complementary to other dualistic distinctions such as mind-body and appearance-reality, which lay at the core of a representationalist way of describing thought, knowledge, and language. Moreover, those dualistic distinctions have come to us through the vocabulary inherited from the Greeks, which in turn has triggered our readiness to believe that there are essences (cf. Rorty 1999, 47-51).

Believing that there is a distinction to be made between what is essential and accidental, according to Rorty, has led philosophers to believe in the existence of some pseudo-problems, namely: "the relation of appearance and reality, of the mind to body, of language to fact" (Rorty 1991, 99). These pseudo-problems suggest that there is something intrinsically privileged inside and outside of us that can be, somehow, described or accurately represented by equally privileged bits of language. An essentialist falls into the delusion that he came to believe in the existence of those intrinsic properties because those objects caused him to believe it and so the essentialist fails to see that those objects are made (rather than given) in the process of beliefs formation.

Moreover, Rorty says that the essentialist is the type who assumes two basic principles, namely Russell's Principle and Parmenides' Principle. The former says, "It is not possible to make a judgment about an object without knowing what object you are making a judgment about," the latter says, "you cannot talk about what does not exist" (Rorty 1991, 105). The adoption of those two principles led philosophers to believe that there exists a special sort of entities that we have to look for and that there is a difference to be made between identifying and describing those entities situated outside us. Rorty thinks that if we accept the idea that there are intrinsic properties, 'things in themselves', then inquiry becomes an activity of discovering, representing or getting hold of those 'objects'.

Instead, Rorty wants to persuade us to think of human mind as a web of beliefs, which is constantly self-reweaving.[6] Such beliefs have to be

---

[6]   It has to be said that the web does not necessarily have any fixed 'core', which would be immune to revision.

understood as 'habits of action' that produce actions, which in turn produce new beliefs that have to be woven into the web of beliefs. We might call this process as 'coming to believe' or 'assimilating a new belief'. It happens at different levels, but sometimes such processes "… provoke the sort of long-scale, conscious, deliberate reweaving which does deserve the name" of inquiry. In that sense, Rorty proposes conceiving of inquiry as 'recontextualizing beliefs'. 'Recontextualizing' here means relocating the new belief in a web of existing beliefs, "for a belief is what it is only by virtue of its position in a web" (Rorty 1991, 94-98).

Given the strong rejection to think of inquiry as having to do with 'essences' or intrinsic properties that we discover or represent, the idea of method seems to become dispensable. In this sense, Rorty says, "by getting rid of the idea of different methods appropriate to the nature of different objects… one switches the attention from the demands of the object to the demands of the purpose which a particular inquiry is supposed to serve" (Rorty 1991, 110). By stressing the idea of 'the demands of the purpose' Rorty seems to be criticizing the idea of method understood as having a filter to percolate our thoughts amongst which there would be some privileged over others because those represent realities or primary qualities more faithfully than others. It would also put, allegedly, some methods in a privileged position in relation to others.

It must be said that we may agree with Rorty's diagnosis in that essentialism has been playing an important role in the way philosophers have construed inquiry and language and that the notion of method is anchored to essentialism. However, the idea that essentialism is embedded in our linguistic practices, which in turn may change contingently in accordance with our social needs and collective purposes, so that we could, as Rorty puts it, 'brush aside' essentialism is highly debatable. Of course, if essentialism is seen as a social construction, it is easy to conclude that essentialism, along with the notions anchored to it, is replaceable. However, cognitive and developmental psychologists have shown, as we shall see, that our tendency to believe that there are essences seems not to be as cultural, historical, socially relative and superfluous as the opponents of essentialism might have thought.

3. Inevitability of essentialism

The reason why essentialism appears to play a central role in our theorizing is that we are cognitively inclined to adopt essentialism. That is the argument of Susan Gelman who asserts that the essentialism is a cognitive bias that appears early in childhood and acts on the way we humans categorize or process information. Therefore, essentialism, Gelman says,

> is the result of several converging psychological capacities, each of which is domain-general yet invoked differently in different domains. (Gelman 2003, 6)

Such psychological capacities or mental abilities are the capacity to distinguish appearances from reality, the capacity to make inductive inference about properties of objects, tracking identity over time, the assumption that properties and events are caused (causal determinism), and deference to experts. These capacities did not appear to help the development of essentialism; rather they collectively have an impact on, or direct implication in, essentialist thinking (cf. Gelman 2003, 316).

According to Gelman, when we attempt to specify the intrinsic property of a thing, we identify it as a member of a category, that is, we identify things falling under certain categories because they possess that intrinsic, non-obvious and unchanging property. Thus, believing that things have intrinsic and non-obvious properties or that animals have innate nature gives shape to our concepts and so, it is

> a powerfully useful psychological disposition that directs adult induction and knowledge acquisition (including the sorts of hypothesis we entertain and attempt to confirm). (Gelman & Wellman 1999, 635)

Gelman suggests neither that there are essences, nor that we can get to know them easily; rather what she is saying is that we believe they exist and that such belief makes easier the development and growth of our conceptual corpus. That is to say, it helps us cope with the world by facilitating the process of categorization, which we, as living systems, must do to organize the amalgam of stimuli we are subjected to.

If this is so, that is, if essentialism is a sort of bias that develops in humans since early childhood without having a significant cultural influence,

then it seems that the idea of essentialism as a contingent bias becomes defective. What is more, as Gelman acknowledges, getting rid of essentialism would demand more than merely running away from it, as Rorty and most anti-essentialists suggest. For, as Gelman puts it,

> even when college students learn about species in a biology course, for example, we suspect it may be difficult to overcome the assumption that species are understood in terms of inherent features that each member possesses (e.g. 'tiger gene'), rather than appreciating that a species is an interbreeding population, characterized by diversity among its members and having no single determinative property. (Gelman & Ware 2012, 471)

It was initially thought that we tended to essentialize most powerfully with respect to natural and social kinds, but not to artifacts made by people. However, Paul Bloom has shown that our tendency to essentialize is not only limited to categories about the natural, but it also applies to categories of objects made by human beings, in which "the essence is seen as the creator's intention" (Bloom 2004, 55). This may be explained, as Gelman and Bloom have suggested, by the fact that essentialism is an implicit assumption, which is manifested in, expands to and adapts to different domains.

In the same way, the cognitive psychologist George E. Newman has found out that human tendency to believe that there are intrinsic properties shared by groups of things is manifested in activities as artwork marketing where the idea of authenticity plays an important role. The way people in those contexts evaluate authenticity presupposes the existence of an intrinsic value, an essence; hence, what is evaluated is whether objects embody or reflect such properties. To illustrate, when people compare an original artwork, let's say of Van Gogh, to its replications, what may give its intrinsic value to the original is that it was actually touched by Van Gogh (cf. Newman & Bloom 2012).

Similarly, it has been shown that the tendency to track essences seems to appear even in the absence of language, so essentialist thinking is prior to the acquisition of language. That is to say, essentialism seems to come into existence earlier than Gelman's studies had shown (cf. Phillips, Shankar & Santos 2010). Given this point, it seems that the argument that

essentialism emerged as a social construction at a certain time in the antiquity, which we have inherited through language, does not really work. Indeed, the fact that essentialism was already present in Plato's or Aristotle's philosophy only means that we have been essentializing for a long time.

The reason why essences seem to be everywhere is that our brains are made to track what is unchanging in the world to make sense of it. That is the argument found on Bruce M. Hood's SuperSense. By appealing to evolution, Hood found that

> [w]hat we do naturally and spontaneously at the most basic level is look constantly for patterns, imagining hidden forces and causes. Even the way we see the world is organized by brain mechanisms looking for patterns. (Hood 2009, 9)

That is to say, the human mind does not work arbitrarily and seems to capture unchanging, intrinsic features in objects, as long as it helps the brain to make sense of the world.

Essentialism gets a bad reputation mostly because of its social consequences. Essentialism, it is said, gives rise to evaluations and social attitudes toward others. So, essentialism: 1. "makes people less responsive to pressure to change their attitudes and judgments"; 2. "Reduces people's motivation to try to eliminate disparities between groups"; 3. "Reduces people's motivation to cross category boundaries", which, in turn, might potentially increase "prejudice toward members of stigmatized groups (Hispanics, homosexuals, etc.), driven by a tendency to essentialize their negatively perceived qualities" (Prentice & Miller 2007, 204). Sure enough, some of the social consequences of essentializing are disastrous. However, we can certainly reject this essentialist understanding of humans without necessarily rejecting essentialism in general.

On the other hand, following the advice of the opponents of essentialism and assuming that it does not play a major role or, even worse, that it is the base of a wrong image of human situation, is of no help in avoiding the negative consequences of essentialist thinking. For the studies, run by Gelman and others, suggest that essentialism is a part of human nature, that essentializing is something we do very well, and that it is mostly a useful tool.

Moreover, as Diana Fuss has shown, the anti-essentialist strategy of claiming that knowledge – including the very notion of essence – is socially constructed and historically determined does not do away with essentialist thinking. For the cases where certain social categories as 'woman' or 'man' are replaced by 'women' or 'men' on behalf of a more pluralistic use of language that stresses the underlying differences between individuals that fall into the same category, "the essentialism at stake is not countered so much as displaced" (Fuss 1989, 4). That is to say, essentialist thinking persists in such claims.

4. Methods without going beyond essentialism

Gelman's research as well as similar studies show that our deeply rooted cognitive tendency to essentialize seems to have a profound impact on the way we categorize and shape our conceptual system. Although Gelman has tried to distinguish between folks' essentialism and philosophers' essentialism, shying away from any possible conclusion about the role played by such cognitive bias on philosophical theories, it would be misleading to claim that there is a difference to be drawn between them. For it would imply that philosophers make use of some conceptual resources different from the rest of mortals, which would be highly debatable. In fact, as pointed out by Lakoff and Johnson, no matter how sophisticated a philosophical theory might be, philosophers employ the very same basic conceptual system shared by ordinary people within a given context when building up their theories (see Lakoff & Johnson 1999, 338).

Certainly, in the case of philosophers' essentialism it is not only an implicit assumption but also an explicit one, forthrightly stated and debugged in philosophers' theories. Nevertheless, philosophers, as well as 'ordinary' people, implicitly assume that there are some essences here or there and it is such belief that may guide the process of drawing inferences about the world, its form, and the way we get to know it.

In this sense, as we have already pointed out, the relationship between essentialism and method is explicitly stated in philosophers' theories. Method, in the light of essentialism, is a defining element of the pursuit of knowledge, that is, the sort of thing without which it would not be reasonable to undertake research about any given matter. Thus, it is method what

enables us to discover the true nature of a given thing. To illustrate this point, let us have a closer look at the Aristotelian and Cartesian conception of philosophical method. As we see them, those conceptions are equally essentialist in that they presuppose that inquiry must be done following a systematic procedure, that is, it requires a method to penetrate the intrinsic nature of things, but it is manifested differently, also having different consequences.

'Method', as it appear in Aristotle's works, sometimes means simply 'inquiry' or 'investigation' (cf. Aristotle 1944; 1983), but in other times, it means – more in line with its modern use – 'mode of pursuing an investigation' (cf. Aristotle 1934; 2002). In brief, for Aristotle 'method' has a general connotation, meaning pursuit of knowledge, inquiry or investigation. Furthermore, inquiry, according to Aristotle, is always goal-oriented, that is, aimed at achieving something. With this in mind, Aristotle made clear that an investigation and the way it should be pursued – that is, a method – is determined by the goal to be attained (Aristotle 1934, 1094a).

Moreover, it is an often-overlooked point that, for Aristotle, pursuing any investigation requires being familiar with the stated goals, that is to say, an investigation is aimed at something, and when we are after something, we do so because we are somehow familiar with that something we are after. Consequently, any given inquiry starts from what is more knowable to us or what is obvious. For, as Aristotle put it himself, "each man judges correctly those matters with which he is acquainted" (Aristotle 1934, 1094b). In that sense, Aristotle's essentialism regarding inquiry is betrayed by the notion of goals or ends, which in turn determine the method to be followed to attain them.

In contrast, in the light of the modern Cartesian conception, 'method' means a handy common set of rules that could be identifiable a priori or drafted beforehand, regardless of its intended use, which can serve as a sort of program of activities (an investigation) that start after the rules of the method have been devised. Equally, a method, as Descartes construed it, is a preset group of rules whose main purpose is the acquisition of a sort of absolute knowledge, which, according to Descartes, 'always remains one and the same' and it should not be wrongly separated and limited by any particular matter. In this sense, Descartes suggested that we would rather give up upon any attempt to undertake research on any matter without

having a predefined method (cf. Descartes 1985b, 360-371). Hence, in the case of Descartes' sort of essentialism regarding inquiry, essentialism is betrayed by his idea of method itself. That is to say, a method, in the light of Cartesian conception, becomes the main or, one may say, an essential feature of knowledge without qualification. Thus, true knowledge is methodical, that is, something acquired according to a given predefined procedure.

As shown above, the relationship between essentialism and method manifests differently in the philosophy of Aristotle and Descartes. It should be noted, likewise, that also the consequences of both sorts of essentialism are different. To explain, Aristotle's essentialism, as we see it, does not necessarily lead to the only-one-method-for-any-subject-matter idea, while Descartes in fact made such idea possible. For a method in the Cartesian sense, as long as it can be devised independently of any subject-matter, becomes something detached from the process of inquiry itself and its purposes, which, in turn, could raise some questions about the usefulness of such device.

Again, it could be argued that it is the same sort of essentialism that led Aristotle to claim the existence of a 'supreme end,' which would make some inquiries more worthy of being pursued than others and for the same reason would lead to the differentiation between disciplines, some being more authoritative than others. However, despite the mentioned concerns, we find Aristotelian essentialism more compelling in that, regarding methods, it does not lead us to think of methods in terms of artificially preset procedures or group of rules that regulate the way we should go about any inquiry, but rather as an activity defined by its goals, some more preferable than others.

In conclusion, we have no doubt that the most critical contribution of Aristotle's conception of method lies in the fact that an inquiry cannot be abstracted from its intended purpose, least of all, to become dependent on a sort of device arbitrarily and capriciously invented and truly different from its applications. Furthermore, what Aristotle's conception would tell us about the modern conception is that even if we keep alive the idea of a group of rules, such group of rules or procedure can only be found in the development of a research activity, of which it is its more or less simplified scheme. That is, a method, keeping its modern sense alive,

would have to be understood as a sort of pattern of a preexistent operation, which is identified *a posteriori*. Of course, it would imply that such device has no practical value for the inquiry process itself. However, as we see it, this look would render the debate about the philosophical method constructive because it would not only force us to focus on identifying the pattern, if there is one, followed by philosophers when pursuing an inquiry, but also discuss what philosophers do when they build their arguments.

5. Conclusion

Essentialism is perhaps one of the notions most vehemently opposed and rejected by philosophers from different strands over the past four decades. However, as shown above, most of the reasons for such a rejection are based on a poor understanding of what it is, how it emerges and on an overestimation of its consequences. Certainly, when social categories are essentialized, it gives us an erroneous account of the nature of humans. Nevertheless, it cannot be denied that the idea of 'penetrating the true nature of things' has been the guiding metaphor behind the success of that body of knowledge that we refer to as science (chemistry, genetics, for example).

In this sense, we argue for an epistemological kind of essentialism, which, in our view, provides an overarching framework that describes and explicates the way we represent inquiry. Such framework is useful to understand why the notion of method is inextricably bound together with our representation of the search for knowledge or inquiry. Indeed, the way we define what a discipline or science is, for example, presupposes the idea of having a method as a fundamental feature.

Likewise, we oppose the anti-essentialist argument that essentialism commits us to a sort of Sisyphean pursuit for knowledge and that it is behind the only-one-method idea. In fact, as has been noted, there are essentialist ways, namely Aristotelian, to construe inquiry such that it does not assume that there is a priviledged single method independently of the purpose of a given inquiry. Moreover, the Aristotelian essentialist conception of method provides a way out of the so-called problem of philosophical

method, for it forces us to focus on what philosophers actually do when they pursue knowledge.

References

Aristotle (1934): Nicomachean Ethics. In: Rackham, H. (ed.): *Aristotle in 23 Volumes. Vol 19.* Cambridge – London: Harvard University Press & William Heinemann Ltd.

Aristotle (1944): Politics. In: Rackham, H. (ed.): *Aristotle in 23 Volumes. Vol. 21.* Cambridge – London: Harvard University Press & William Heinemann Ltd.

Aristotle (1983): Physics: Books III and IV. In: Ackrill, J. L. & Judson, L. (eds.): *Clarendon Aristotle Series.* Oxford: Clarendon Press.

Aristotle (2002): On the Parts of Animals I-IV. In: Ackrill, J. L. & Judson, L. (eds.): *Clarendon Aristotle Series.* Oxford: Clarendon Press.

Aristotle (1928): *Posterior Analytics.* Oxford: Clarendon Press.

Beaney, M. (2000): Conceptions of Analysis in Early Analytic Philosophy. *Acta Analytica* 15, No 25, 97-115.

Bloom, P. (2004): *Descartes' Baby: How the Science of Child Development Explains What Makes Us Human.* New York: Basic Books.

Bonaventure, St. (2002): Itinerarium mentis in Deum. In: *Works of Saint Bonaventure. Vol. II.* Saint Bonaventure, NY: Franciscan Institute, Saint Bonaventure University.

Dear, P. (1998): Method and the Study of Nature. In: Garber, D. & Ayers, M. (eds.): *The Cambridge History of Seventeenth-Century Philosophy. Vol. I.* Cambridge: Cambridge University Press, 147-177.

Descartes, R. (1985a): Principles of Philosophy. In: Cottingham, J. (ed.): *The Philosophical Writings of Descartes.* Cambridge (Mass.): Cambridge University Press.

Descartes, R. (1985b): Rules for the Direction of the Mind. In: Cottingham, J. (ed.): *The Philosophical Writings of Descartes.* Cambridge (Mass.): Cambridge University Press.

Descartes, R. (2008): *A Discourse on the Method.* Oxford: Oxford University Press.

Fuss, D. (1989): *Essentially Speaking: Feminism, Nature & Difference.* New York: Routledge.

Gelman, S. & Wellman, H. (1999): Insides and Essences: Early Understanding of the Non-Obvious. In: Margolis, E. & Laurence, S. (eds.): *Concepts: Core Readings.* Cambridge (Mass.): MIT Press, 613 – 638.

Gelman, S. (2003): *The Essential Child: Origins of Essentialism in Everyday Thought.* Oxford: Oxford University Press.

GELMAN, S. & WARE, E. (2012): Conceptual Development: The Case of Essentialism. In: Margolis, E., Samuels, R. & Stich, S. (eds.): *The Oxford Handbook of Philosophy of Cognitive Science.* Oxford: Oxford University Press, 454-479.

HALLETT, G. (1991): *Essentialism: A Wittgensteinian Critique.* Albany: SUNY Press.

HOOD, B. (2009): *SuperSense: Why We Believe in the Unbelievable*. San Francisco: HarperOne.

HYLTON, P. (1998): Analysis in Analytical Philosophy. In: Biletzki, A. & Matar, A. (eds.): *The Story of Analytic Philosophy: Plot and Heroes.* London: Routledge, 37-55.

LAKOFF, G. & JOHNSON, M. (1999): *Philosophy in the Flesh.* New York: Basic Books.

NEWMAN, G. E., & BLOOM, P. (2012): Art and Authenticity: The Importance of Originals in Judgments of Value. *Journal of Experimental Psychology: General* 141, 558-669.

PHILLIPS, W., SHANKAR, M. & SANTOS, L. R. (2010): Essentialism in the Absence of Language? Evidence from Rhesus Monkeys (Macaca mulatta). *Developmental Science* 13, F1-F7.

PLATO. (1969): Republic. In: *Plato in twelve volumes.* Cambridge (Mass.) – London: Harvard University Press & William Heinemann Ltd.

POPPER, K. (2002): *The Poverty of Historicism.* London: Routledge.

PRENTICE, D. & MILLER, D. (2007): Psychological Essentialism of Human Categories. *Current Directions in Psychological Science* 16, No 4, 202-206.

RORTY, R. (1991): *Objectivity, Relativism, and Truth.* Cambridge: Cambridge University Press.

RORTY, R. (1992): *The Linguistic Turn.* Chicago: The University of Chicago Press.

RORTY, R. (1999): *Philosophy and Social Hope.* London: Penguin Books.

A Few Comments on the Linda Problem[1]

ADAM OLSZEWSKI

ABSTRACT: This paper discusses an experiment in cognitive psychology called the Linda problem. Firstly, some natural conditions for the correctness of the interpretation of psychological experiments (such as the Linda problem) are formulated. The article is essentially a critique of the interpretation of the results of the Linda problem experiment provided by Kahneman and Tversky as well as – indirectly – their concept of heuristics. It is shown that the interpretation provided by Kahneman and Tversky does not meet the aforementioned conditions for correctness. The main argument is justified utilizing such rules of rationality as conditional probability and Grice's conversational maxims. It is also pointed out that this argument can be reformulated in terms of the intuitive system of reasoning.

KEYWORDS: Conjunction fallacy – cognitive psychology – conditional probability – Linda problem.

In this paper, I present a new interpretation of the "Linda problem". The "Linda problem"[2] is the name of a psychological experiment performed by psychologists Daniel Kahneman and Amos Tversky ('KT' for short) in the

---

✉ Adam Olszewski
   Department of Philosophy of Logic, Faculty of Arts
   The Pontifical University of John Paul II in Cracow
   Ul. Kanonicza 25, 31-002 Cracow, Poland
   e-mail: atolszad@cyf-kr.edu.pl

[2] In this paper, I will also use terms such as the "Linda task," often without quotation marks.

1980s. It is described in Kahneman & Tversky (1983); and after a subsequent research and analysis, the summary of the experiment has been presented in Kahneman (2013). The experiment consisted of the description of an imaginary woman named Linda. Her story is as follows:

> Linda is thirty-one years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations. (Kahneman 2013, 211 ff.)

In one version of the study, the researchers then presented eight possible scenarios for Linda's future. They were:

- Linda is a teacher in primary school;
- Linda works in a bookstore and practices yoga;
- Linda supports the feminist movement (abbreviation: (F));
- Linda is a social worker and helps people with mental disorders;
- Linda is a member of the Women's Electoral League;
- Linda is a bank teller (abbreviation: (T));
- Linda is an insurance agent;
- Linda is a bank teller and is active in the feminist movement (abbreviation: $(T \wedge F)$).

The participants of the study were asked which of the following two propositions is more probable. "Linda is a bank teller" (T) or "Linda is a bank teller and is active in the feminist movement" $(T \wedge F)$.

The experiment was repeated many times with different groups and in different versions, as well as with groups whose participants had previously completed courses related to statistics or probability theory. To the surprise of the researchers, most the subjects stated that the conjunction $(T \wedge F)$ is more probable than the proposition (T). This violates the axioms of probability theory, which entail that the probability of the conjunction of two propositions is less than or equal to the probability of the individual members of the conjunction $p(T \wedge F) \leq p(T), p(F)$. In addition, the subjects were unable to explain why they ascribed probabilities to both propositions the way they did.

Relevant here is a well-known quote from Stephen Jay Gould, who, knowing the right answer, wrote that, "[A] little homunculus in my head

continues to jump up and down, shouting at me – 'But she can't just be a bank teller; read the description!'" (Kahneman 2013, 215).[3] Kahneman describes the tremendous impression the results made on him (cf. Kahneman 2013, 213).

The experiment was repeated on a group of PhD candidates from Stanford Graduate School of Business, who had previously participated in probability theory, statistics, and decision theory classes. Ninety-five percent of respondents judged against the laws of probability, and thus, according to the researchers, against the rules of rationality. Moreover, the researchers also carried out tests on a group larger than the initial one and on groups where the participants were considered better (or worse) educated, which substantiated the results.[4]

Kahneman mentions only one group in which 64% of the participants gave the correct answer.[5] They were PhD candidates in sociology from the universities of Stanford and Berkeley.

In their research, Tversky and Kahneman considered, for instance, the transformation of the proposition (T) into the following, "Linda is a bank teller, regardless of her activity in the feminist movement."[6] After conducting an experiment with this new proposition, the number of incorrect answers was reduced to 57%. On the other hand, when conducting the experiments described in Kahneman & Tversky (1983), KT knew that the respondents might treat (T) as $(T \wedge \neg F)$ – see Kahneman & Tversky (1983, 299).[7] At the same time – based on the description – KT knew that the

---

[3]   I prefer to quote this paper here because it is more comprehensive and decisive.

[4]   KT mentioned studying approximately 3,000 subjects; cf. Kahneman & Tversky (1983, 309-310).

[5]   It is assumed by KT that an answer is correct when it is consistent with Probability theory.

[6]   "Linda is a bank teller, whether she is active in the feminist movement or not" (Kahneman & Tversky 1983, 299). KT therefore took into consideration different formulations of (T). I also mention these versions of the Linda experiment, as well as other interpretations made by KT, to emphasize the versatility of their analysis and the fact that the interpretation I will present had already been taken into consideration in a certain form.

[7]   Their reason for rejecting this interpretation is surprising. "To test this interpretation, we asked a new group of subjects (N = 119) to assess the probability of T and of

events (T) and (T ∧ F) are not independent, so there was no reason for them to assume that the subjects would treat (T) as (T ∧ ¬F), since this would imply that (T) and (T ∧ F) are independent, which would be at odds with logic. In this context, another peculiar point is that KT, to exclude an understanding of (T) as (T ∧ ¬F), did not examine the relationship between (F) and (T ∧ F).

What is more, KT knew from previous research that, if Linda's description is limited to the judgment that she is a 31-year-old woman, the results are in accordance with the laws of probability (cf. Kahneman & Tversky 1983, 305). Their conclusion is that C (C being the description of Linda), or the *paradigm,* as they call it, has affected the outcome of the experiment. It is my contention that we are dealing here with ordinary conditional probability, where the condition is C.

Let us ask why this "unexpected" outcome of the experiment had such a strong influence on the authors of the experiment as well as on the scientific community. Since there are, in fact, two questions, I assume that the answer to the first question arises from the answer to the second. Indeed, one of the main arguments of KT is the argument concerning dual process theory.[8] This theory, also known as the dual process account of reasoning, states that there are two systems [minds] in the human brain. System$_1$ is intuitive, and System$_2$ is analytical.[9] I believe that KT acted, as I will attempt to prove later on, within a kind of "paradigm" of this distinction.[10]

T & F on a point scale ranging from 1 (extremely unlikely) to 9 (extremely likely)" (Kahneman & Tversky 1983, 299). KT thought that it seems reasonable to assume that the respondents did not interpret T as T ∧ ¬F, since there is nothing wrong to estimate the probabilities of two events even though one of them is a part of the other.

"The pattern of responses obtained with the new version was the same as before. The mean ratings of probability were 3.5 for T and 5.6 for T & F, and 82% of subjects assigned a higher rating to T & F than they did of T" (Kahneman & Tversky 1983, 299).

[8]   At the time, the view that mind has two systems emerged at least in 1975. In Kahneman & Tversky (1983), this thesis took the form of division into extensional and intuitive reasoning (stated as early as in the title of the article).

[9]   A modern summary of the research and the scientific hypothesis concerning this interesting distinction can be found in Hertwig & Gigerenzer (1999).

[10]   It cannot, however, be claimed that KT were explicitly formulating the dual process theory.

Kahneman expressed this, for example, by commenting on Gould's statement, stating "[T]he homunculus in Gould's head was, *of course*, (italics mine – A. O.) insistent System$_1$" (Kahneman 2013, 213).

Let us now discuss KT's interpretation of the result of the Linda experiment. I use the word "interpretation" with considerable caution, keeping in mind that it is unclear whether psychologists have any unified and standardized theory concerning the interpretation of the results of such experiments as well as the interpretation of *tasks* given to their subjects. In the case of multiple psychological experiments, we are faced with more than one interpretation, two interpretations being dominant.

Experiments similar to the one in question often feature two intersubjective factors (texts) and four subjective factors (interpretations). The first objective factor is a text (t_1), created by the researchers, consisting of the description of the task to be solved by the participants. The second factor is usually also a text (t_2), the aim of which is to present the notions of the participants. The interpretations concern the aforementioned intersubjective texts, and represent the subjects' ways of understanding the texts. These are very often subjective with respect to both, researchers and participants. By using (i_1), we can denote the interpretation of text (t_1) offered by the researchers who performed the experiment. Similarly, by using symbol (i_4) we can denote the interpretation of the same text (t_1) offered by the subjects of the experiment. Both interpretations should be equivalent, as this is usually the aim of the researchers concerning text (t_1). The second aspect is the interpretation (i_2) of the experiment's result which consists of the interpretation of text (t_2) provided by the researchers who created the experiment and the interpretation (i_3) of the experiment's result provided by other researchers interested in the experiment. The interpretations (i_2) and (i_3) should also be equivalent. Therefore, there are two equivalences necessary for the methodological soundness of both the KT experiment as well as any similar experiments. This is, however, insufficient, as no one falsified the KT statements: [11]

$$(i_1) \equiv (i_4)$$

---

[11]   I argue that the interpretations of the Linda problem do not fulfill both of these equivalences. The equivalences are taken here to be stronger then material equivalence, rather as 'to have similar meaning'.

$(i_2) \equiv (i_3)$.

In the case of the KT experiment, neither equivalence holds, as will be explained later in this paper.

Interpretations (i_1) and (i_2) presented by KT were immediately criticized by other researchers in multiple ways. One critical remark was made by German psychologist Gerd Gigerenzer. It referred to the understanding of the term 'probability' by the subjects. When the experiment was repeated, with the question regarding the probability of the proposition replaced with the *frequency* of occurrence (the frequency interpretation), a great majority of the subjects answered correctly.

Some researchers have suggested (e.g., Morier & Borgida 1984, Hilton 1995) that the proposition (T) could have been understood by subjects as $(T \wedge \neg F)$; this case will be described further. Morier and Borgida studied such possible understandings of the proposition T and developed an experiment in which the subjects were asked about the propositions: (T), (F), $(T \wedge F)$ and $(T \wedge \neg F)$. The results of the experiment confirmed the presence of *the conjunction fallacy*. However, after constructing a problem which was logically equivalent to the Linda problem, in which the meaning of the propositions was explained more clearly, to their surprise, the probability of error decreased dramatically (cf. Miyamoto, Gonzalez & Tu 1995, 337-338).[12] However, Hilton claims that the interpretation of (T) as $(T \wedge \neg F)$ is based on rational heuristics (they differ, however, from those desired by KT). The reason for disregarding the rules of probability theory is given by the rules of rational pragmatics, and not by any kind of illogicality (cf. Stanovich 2010, 102).

Returning to the criticism regarding the interpretation of the probability used by KT, it is worth mentioning that there are at least five known interpretations of probability. The participants in the KT experiment were asked to assess the probability of the sentences, which shows that they were essentially asked to make use of Carnap's logical probability. If this were the case, it would provide a basis for criticism of the KT experiment, since if the participants in the study had only basic knowledge of logical probability, this would open the experiment to allegations of circular reasoning.

---

[12]   The authors reflect on conditional probability in the context of the Linda problem.

Let us now turn to the interpretation (i_4) of the result of the Linda problem experiment which refers to the criticism of KT put forward by David Morier and Eugene Borgida (see Morier & Borgida 1984), and later by Denis Hilton (see Hilton 1995). They are based on the analytical System$_2$.[13] Firstly, I will present the interpretation (i_2) as the hypothetical reasoning of a possible participant in the experiment.

 I. Grice's Rules[14] (Maxims of Relevance and Quantity);[15]

 II. Description of Linda C is essential for the task; (from I. and the description)

 III. Based on C, the conditional probability[16] that Linda is a feminist $p_c(F)$ is higher than that she is not, i.e.: $p_c(\neg F) < p_c(F)$.

 IV. $p(T \wedge F) \leq p(K), p(F)$; (from the properties of probability theory; abbreviation: CP)

 V. The sentence K, in the context of Linda's description, may be understood[17] by subjects as an abbreviation for $(T \wedge \neg F)$, which justifies the following inference [VI. – XI.]:

 VI. (F) is true or (F) is false; (from the bivalence of classical logic)

 VII. (F) did not occur, therefore (F) is not true; (from Grice's maxims of relevance and quantity)

 VIII. (F) is false;

 IX. $(\neg F)$ is true;

 X. We have (T);

 XI. Therefore: $(T \wedge \neg F)$; (from classical logic)

 XII. $p_c(T) = p(T)$; (from the independence of (T) and C)

 XIII. $p_c(T \wedge F) = p((T \wedge F) \wedge C) / p(C)$; (from the definition of CP)

 XIV. $p((T \wedge F) \wedge C) / p(C) \equiv p(T \wedge (F \wedge C)) / p(C)$; (from the laws of logic)

---

[13] I am using conditional probability here as an important element.

[14] It means that the subject accepts Grice's conversational rules (maxims).

[15] The Maxim of Quantity requires us to say only what is necessary at a given stage of conversation, and the Maxim of Relevance to say only what is integrally related to the topic of conversation.

[16] The condition of the entire C.

[17] Here, we also require properly designed empirical research.

XV. $p(T \wedge (F \wedge C)) / p(C) = (p(T) \cdot p(F \wedge C)) / p(C)$; (from VI and independence of (T) and (F))

XVI. $p(T) \cdot (p(F \wedge C) / p(C)) > p(T) \cdot (p(\neg F \wedge C) / p(C))$; (from III)

XVII. $(p(T) \cdot p(\neg F \wedge C)) / p(C) = p(T \wedge (\neg F \wedge C)) / p(C)$; (from CP)

XVIII. $p((T \wedge \neg F) \wedge C) / p(C) = p_C(T \wedge \neg F)$; (from CP);

XIX.   Therefore: $p_C(T \wedge F) > p_C(T \wedge \neg F)$; (from VII, XI, XIV and CP).

As we see in the sequence of this reasoning, premises III, V, and VI are critical to the above structure.

For the sake of clarity, I will characterize the steps of this reasoning also in an informal way. I assume that Grice's maxims can be applied, particularly the maxims of relevance and quantity. I think that description C is important. The probability that Linda is a feminist given description C is higher than the probability that she is not a feminist under the same condition, although there are more non-feminists than feminists. The question whether Linda is a feminist is crucial for our discussion. If it is said that Linda is a bank-teller, but feminism is not mentioned, then it should be assumed that Linda is a bank-teller and not a feminist, which is expressed by step V of the reasoning. Next, using the formal properties of the conditional probability we get (XIX), or: under condition C, the probability that Linda is a bank-teller and a feminist is higher than the probability that she is a bank-teller and not a feminist, and we briefly express it in our reasoning by proposition (T).

Concerning the reasoning presented above, we are dealing with two interpretations of the results of the Linda experiment. The first one (i_1), derived from KT, supports System$_1$, while the second interpretation (i_4) is based mainly on System$_2$. As mentioned earlier, KT worked within a certain paradigm that is closely related to the dual process theory, which makes these interpretations mutually inconsistent. Therefore, the second interpretation bears directly on the concept of KT. The following questions arise.

Is anyone wrong here?
If so, who is wrong?
If KT are wrong, why?

A positive answer to the first question stems from the inconsistencies present in both interpretations. This inconsistency arises only if one accepts

the paradigm of KT, particularly the dual process theory.[18] In my opinion, while the estimation of probabilities is made by the intuitive part of the mind, which is similar to what KT desire, the reasons for this estimation come from analyzing the mind, which conflicts with KT's interpretation, i.e. (i_2). For KT, the mechanism on which the subjects based their answers is heuristic, and in this case, the representativeness heuristic.[19] This heuristic, per KT, is being activated cognitively by Linda's description. It should be noted that the Linda task is one of the most important lessons for experimentally confirming the existence of the representativeness heuristic.[20] Regarding the second question, assuming the point of view presented in the second interpretation, KT's interpretation is incorrect. The answer to the third question requires a slightly longer argument.

There are three main arguments against the interpretation of KT. The first one was mentioned earlier and is related to *circular reasoning*. While the concept of logical probability is not well-known, the subjects had no choice but to make use of the intuitive concept of logical probability. For KT, seeking an experimental confirmation for their concept of the role of heuristics and intuition in cognition, in a way, "forced" the participants to use intuition in advance. To phrase this differently, they checked whether the subjects would use intuition while simultaneously provoking them to do so. This is quite surprising, since the title of Kahneman & Tveresky's (1983) refers to both extensional (classical) probability and intuitive probability.

Secondly, as mentioned previously, KT worked within a certain paradigm, and through this, perhaps, made a so-called *systematic error*, where the researchers, "as a matter of principle," tried to interpret the results of any experiment to be in favor of the concept they had initially assumed.

---

[18] If there were only one system of mind, the KT thesis would become somewhat trivial.

[19] This heuristic is an intuitive method of reasoning which allows one to classify a described object in view of its resemblance to a typical representative of the class in question.

[20] This is not the only experiment that supports the concept of KT, as there have been multiple such experiments. Therefore, it cannot be stated, based on my criticism, that the concept of KT has been countered. My comments pertain to the Linda experiment only.

The confirmation of such a suspicion can be found in and reconstructed from what KT offer as reasons for the rejection of the possibility that the subjects had interpreted (T) as (T ∧ ¬F). Let us recall what they say in this context:

> Since assessment of probability makes sense even if one event includes another, subjects had no reason to interpret (T) as (T ∧ ¬F). The response pattern obtained using the new version was the same as before. (Kahneman & Tversky 1983, 299)

This is a peculiar statement, since it seems that KT use the classical concepts of probability (*event*, *inclusion*) in this statement, while during the experiment, they asked about logical probability. Furthermore, they think that the event in the conjunction (T ∧ ¬F) mentioned earlier is contained in the event (T), which, according to KT, is sufficient for rejecting that interpretation of proposition T.[21] Consequently, as KT conclude at one point, comparing the probability of the second event to the probability of the event (T ∧ F) makes more sense than comparing it to the probability of the event (T ∧ ¬F) itself.[22] KT also performed additional (control) experiments designed to exclude the interpretation of (T) as (T ∧ ¬F) by the subjects. However, the subjects continued to attribute higher probability to the conjunction than to its conjuncts (cf. Miyamoto, Gonzalez & Tu 1995, 336-339).[23]

The third objection concerns assuming Linda's task to be the whole task, or a *pars pro toto* error. KT asked the participants to answer the question, (PL) "Which sentence is more probable, (T) or (T ∧ F)?" This problem is, even syntactically, a subproblem of the entire Linda task[24] and the correct answer to this question is, of course, that the probability of the conjunction is equal to or less than the probabilities of its conjuncts. Such an answer, held by KT to be the correct one to (PL), is, according to my

---

[21] In citations (and comments) I preserve the signs used by KT.

[22] Such an argument is even more peculiar. I admit I did not understand it correctly.

[23] This trend in KT's research is not clear to me.

[24] Let us take the Linda problem OP and treat it strictly syntactically as a set of expressions of some language. The subproblem of the problem OP, in a syntactic sense, shall be called problem P, where (P ⊂ OP).

interpretation, incorrect. It is not the answer to the *entire* Linda task, or the *entire* text of the task, since it is merely the answer to the question (PL).

Let us try to summarize our discussion so far. We raised some objections to KT's interpretation (i_1) of the Linda problem experiment interpretation. This interpretation was guided by their assumption that the judgement of the experiment's outcome was understood in terms of activation of the representativeness heuristic that is, using current terminology, the activation of $System_1$. Criticism rested on building a different interpretation (i_4), according to which $System_2$ was used. In my interpretation, there is no need to refer to the heuristics – referring to the analytical $System_2$ is sufficient. Its effect is presented in the form of interpretation (i_2). In general, when it comes to the description of mind, the results of KT's researches in the domain of cognitive psychology and the theory of decision-making are among the most crucial ones. In 2002, Kahneman received the Nobel Prize[25] in economics for his psychological works, which undermined the traditional model of human rationality. To prevent the reader from arriving at any incorrect conclusions regarding this paper, it must be stressed that there are multiple experimental results that confirm KT's conception.[26] Thus, in no case did my paper seek to challenge any general arguments set out by KT. Here I was concerned only with a criticism related to their interpretation of the Linda problem.

It seems that it would be very interesting to perform an experiment in which the reasoning method for a certain problem associated with probability would be compared, on the one hand, in accordance with the representativeness heuristic, and, on the other hand, in accordance with the logic used in mathematics and philosophy, not necessarily classical one. Such an experiment could bring a new look at the Linda problem, although not an attempt at KT's experiment. Taking my considerations into account, it seems very likely that the results of such an experiment would be inconsistent. This would be an interesting conclusion, indicating that the human mind is much more complex, and that it is difficult to predict with certainty the probability of human decisions.

---

[25] A. Tversky died in 1996.

[26] Cf. Kahneman (2013), and the two additions to that work, which are reprints of important KT publications.

References

EVANS, J. & STANOVICH, K. (2013): Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science* 8, 223-241.

HERTWIG, R. & GIGERENZER, G. (1999): The 'Conjunction Fallacy' Revisited: How Intelligent Inferences Look Like Reasoning Errors. *Journal of Behavioral Decision Making* 12, 275-305.

KAHNEMAN, D. (2013): *Pułapki myślenia. O myśleniu szybkim i wolnym* (*Thinking, Fast and Slow*). Poznań: Media Rodzina.

KAHNEMAN, D. & TVERSKY, A. (1983): Extensional Versus Intuitive Reasoning. The Conjunction Fallacy in Probability Judgment. *Psychological Review* 90, 293-315.

MIYAMOTO, J. M., GONZALEZ, R. & TU, S. (1995): Compositional Anomalies in the Semantics of Evidence. In: Busemeyer, J., Hastie, R. & Medin, D. (eds.): *Decision Making from a Cognitive Perspective*. San Diego: Academic Press, 319-384.

STANOVICH, K. (2010): *Decision Making and Rationality in the Modern World*. Oxford: Oxford University Press.

Personal Identity and What Matters[1]

JEREMIAH JOVEN JOAQUIN

ABSTRACT: There are two general views about the nature of what matters, i.e. about the metaphysical ground of prudential concern, the ground of the concern we have for our own future welfare. On the one hand, the identity-is-what-matters view tells us that prudential concern is grounded on one's continuing identity over time; I am concerned with my own future welfare because it is *my own* future welfare. On the other hand, the identity-is-not-what-matters view tells us that prudential concern is not grounded on such continuing identity; rather, it is grounded on some continuity-relation, which only coincides with identity. In this paper, I explore a primary motivation for the latter view—viz., Parfit's fission case—and show that there are interesting ways to resist it.

KEYWORDS: Fission case – identity-is-not-what-matters – identity-is-what-matters – Parfit.

1. Two views about what matters

Suppose a dentist tells you that someone will suffer a terrible toothache tomorrow. For most of us, this prospect would be terrifying, not for anything else, but simply because there is pain involved. Some of us may reason that whenever there is the prospect of a terrible pain (to whomever it may chance upon), it is always rational for us to abhor it. But suppose the

---

✉ Jeremiah Joven Joaquin
Department of Philosophy, De La Salle University
2401 Taft Avenue, 0922 Manila, Philippines
e-mail: jeremiah.joaquin@dlsu.edu.ph

dentist tells you that *you* and not someone else will suffer a fate tomorrow. I suppose that, for many of us, this latter case would be doubly terrifying, not only because of the pain involved, but more importantly, because of the thought that it will be *my* or *your* very own pain. Our commonsense intuitions about these two cases invite an interesting philosophical question about the metaphysical grounding of what matters or of prudential concern, i.e., the concern one has for his or her own future welfare.

There are two general views about the question about what matters: the identity-is-what-matters view and the identity-is-not-what-matters view.[2] The former view is our commonsense intuition about what matters. It tells us that prudential concern is grounded on one's continuing identity over time. Thus, I am hard at work now because I know that I will later enjoy the benefits of these present labors. In the same way that you are preparing tenaciously for an exam tomorrow because you will be the very same person who will take that exam.

On the other hand, the identity-is-not-what-matters view tells us that the prudential concern is not grounded on one's continuing identity over time; rather, it is grounded on some continuity relation that only coincides with identity. On this view, prudential concern is formulated in terms of the continuity of some beliefs, desires, and intentions over time. Thus, my concern for my own future welfare is nothing more than my desire or intention that some of my cherished hopes and dreams will be fulfilled at a later time. My concern is not that *I* experience the fulfillment of my hopes, but that some future person, who is physically and/or psychologically continuous with me, experiences them.

In this paper, I aim to do two main things. In section 2, I explore a primary philosophical motivation for the identity-is-not-what-matters view due to Derek Parfit. This motivation is premised on the possibility of fission. In section 3, I show five ways of resisting the fission case. I argue that at least some of these ways could show that the fission case is not a suitable motivation for the identity-is-not-what-matters view.

---

2   The labels, "identity-is-what-matters view" and "identity-is-not-what-matters view", follow Parfit's nomenclature. See Parfit (1971; 1984).

2. Fission and the identity-is-not-what-matters view

Parfit construes the question about what matters in terms of the importance we attach to our own survival. To motivate this, consider the following case:

Suppose that I will undergo a dangerous surgical procedure. I am surely concerned whether I will survive this. But the question now is: what am I really concerned about when I think about my own survival? Am I concerned whether some future person shares my beliefs, desires, and intentions before the surgery? Or am I concerned whether the person who will survive the surgery is still me? If one goes for the first case, then that leads to the identity-is-not-what-matters view. If one goes for the second, then that leads to the identity-is-what-matters view.

Parfit presents the following argument for the identity-is-not-what-matters view:

1. Identity is a one-one relation.
2. Survival matters.
3. What matters in survival is the obtaining of the relation of psychological connections between psychological states (or what he calls the R-relation).
4. The R-relation need not be a one-one relation.
5. Therefore, identity is not what matters for survival – (cf. Lewis 1976, 19)

As it stands, the argument seems valid. Given that the premises are all true, the conclusion would surely follow. But as we all know the validity of an argument, especially of a philosophical argument, is not sufficient for the acceptance of the truth of its conclusion. More needs to be said about the truth of the premises.

Premise (1) highlights what we already know about the identity relation. As such, it is something that we could take for granted. Premise (2) just reiterates our concern for our own future welfare (or our own future survival). It matters to us whether we will survive some event. And we are concerned for our own future welfare. Again, premise (2) seems a pretty reasonable assumption.

The crucial turn in the argument is found in premises (3) and (4). At the outset, it seems that, given premise (3), the argument already begs the question, since, as it is stated, premise (3) is just the identity-is-not-what-matters view. Premise (4) fares no better than premise (3). To say that the R-relation is not a one-one relation depends on whether the connections between various psychological states do not exhibit a one-one relation. That is, premise (4) claims that the connection between psychological states could be a one-many relation or else a many-one relation. But the truth of this claim relies on the truth of premise (3). But since premise (3) just states the conclusion of the argument, then premise (4) just seems to be an ungrounded assumption (cf. Oaklander 1987).

To answer this circularity objection, Parfit motivates premises (3) and (4) via the fission thought experiment.[3] The fission case proceeds as follows:

Suppose that my whole body develops cancer. The surgeons cannot save my body, but they could save my brain. They remove my brain and transplanted each hemisphere into two brainless bodies, each of which was cloned from my original DNA before the surgery. The operation was a success. After some time, two people wake up – call them Lefty and Righty. Both are psychologically continuous and/or connected with me – they both share my memories, intentions, hopes, desires, and beliefs. Both are also physically continuous with me – they each have half my original brain. Suppose further that the existence of one is unknown to the other. That is, suppose that after the operation Lefty just went on to live his own life independent of Righty. Lefty went to Boracay and opted for a recluse life (which is one of my long-time dreams); while Righty went on to do philosophy all his life (which is also one of my long-time dreams). But now the question is: did I survive? And if so, did I survive as the person in Boracay living a recluse's life, or did I survive as the person doing philosophy?

Parfit claims that if we hold the identity-is-what-matters view, then we have to say that I did not survive the operation; we should regard the prospect of my fission as being nearly bad as death. That is, after the operation

---

[3]   Parfit has used other imaginary cases to motivate these premises, e.g. the Branch-Line Case and the Teletransportation Case. Both make a vivid portrayal of why identity should not what matter to us. See Parfit (1984, 199-201); see also Garrett (1998, 16-17) for other versions of these cases.

no one identical to me exists. Both Lefty and Righty are not identical to me. As such, I did not survive; I have ceased to be. But he thinks that we should not see the matter this way.

Furthermore, Parfit thinks that the prospect of fission is just as good as ordinary survival. Though Lefty and Righty are not identical to me, my relation to each of them contains all that matters to survival. The projects I aimed to pursue, the beliefs I once held, and my desires and intentions prior the operation are all fulfilled and preserved by my two descendants. And since this is so, there is nothing else that needs to be accounted for my survival.

The fission case, thus, motivates Parfit's argument for the identity-is-not-what-matters view. Since all that matters to us, all that we are really concerned about, is the fulfillment and preservation of our psychological states at a future time, it follows then that premise (3) is true. Furthermore, since the connection between psychological states is not necessarily a one-one relation, as evidenced by the case of Lefty and Righty, then premise (4) will be true as well. From this it follows that identity is not what really matters to us.[4]

3. Resisting the fission case

We have seen Parfit's fission case and how it motivates the argument for the identity-is-not-what-matters view. In this section, I show five ways of resisting Parfit's case.[5] I argue that at least some of these ways could show that the identity-is-not-what-matters view is unmotivated.[6]

---

[4] For Parfit, this implies that we should not really attach a special metaphysical status to our own prudential concern. Since prudential concern is not really grounded on our continuing identity, but only in the R-relation which only coincides with identity, it should follow that prudential concern is just the same as the concern we have for other people's welfare. He argues that from this it follows that we have to change our views about rationality and morality. The former implies that the self-interest theory is wrong; while the latter implies that any ethical theory grounded on the self-interest theory is wrong as well.

[5] Some aspects of the subsequent discussions come from Garrett (1998, 59-64).

[6] Contra Parfit, showing that the fission case could be resisted does not necessarily imply the acceptance of the self-interest theory of rationality or even the ethical theories

3.1. I survive as both Lefty and Righty

One way of resisting the fission case is to claim that after fission, I survive as both Lefty and Righty. There are two ways of cashing out this idea. Each way, however, offends commonsense. The first way is to claim that after fission, I survive because I am identical to both Lefty and Righty. The second is to claim that Lefty and Righty are not identical to me, but they are sub-personal constituents of me. That is, I am just the product of both Lefty and Righty.

I find the claim that I am identical to both Lefty and Righty implausible. After fission, Lefty and Righty are two distinct people. Though they are qualitatively similar – i.e., they both share the same physical and psychological characteristics – they are still two numerically distinct people. If Lefty were to get happily married sometime after fission and Righty were to remain a frustrated bachelor, then, according to this view, I would both be a happily married man and a frustrated bachelor all at the same time.

On the other hand, to claim that Lefty and Righty are personal constituents of me is to claim that prior and after my fission, I have two separate consciousnesses. Thus, prior to my fission if Lefty thinks that Trump will be ousted from the presidency and Righty thinks otherwise, then I would have two contradictory beliefs about the matter.

Now, this is not detrimental for the claim that I have two sub-personal constituents, since we could have two conflicting beliefs about some matters. I could believe that I see a dog, but I could surely believe that what I see is not a dog. But the claim pushes us to consider that we could hold contradictory beliefs at the same time. At some particular time, I believe that some dog is there and not there. And this is something that we could not countenance.

What we do accept is that we could change our beliefs after some time. And in such a case we do not have two contradictory beliefs at the same time. It could be that at one time I believe that a dog is there, but upon closer inspection I change my belief since now I believe that it is not a dog. To change a belief involves the passage of time. And this shows that we

---

which are implied by it. These are two distinct worries. The question about whether prudential concern is derivative is a different sort of question from whether I should only be concerned for my own welfare.

could only hold contradictory beliefs because we could change our minds. What we could not accept is that some person holds two contradictory beliefs at the same time. As it stands, the claim that I survive as both Lefty and Righty does not hold water.

3.2. The case has been misdescribed

A second way of resisting the fission case, and a popular one at that, is to claim that the case has been misdescribed. There are many defenders of this view and each defender presents a different version of this (see for example Lewis 1976; Sider 2001; Robinson 1985; and Perry 1972). In the literature, this view is referred to as the multiple occupancy theory.[7]

Defenders of the multiple occupancy theory claim that there is no real tension between the identity-is-what-matters view and the identity-is-not-what-matters view. For example, David Lewis, an ardent defender of the theory, claims that

> [t]he opposition between what matters and identity is false. We can agree with Parfit that what matters in questions of personal identity is mental continuity or connectedness (R-relation), and that this might be one-many or many-one… At the same time we can consistently agree with commonsense that what matters in questions of personal identity – even in problem cases – is identity. (Lewis 1976, 19)

But how does this claim cash out? One way is to show that prior fission Lefty and Righty already exist. Two numerically distinct persons just happen to exist in the same body. But after fission these two persons divide and each lives out his own life.[8]

One motivation for the multiple occupancy theory is to consider that there are two relations involved when we talk about personal identity and what matters: the relation of psychological continuity and/or connectedness (R-relation) and the relation of continuing identity (I-relation). These

---

[7]   The label was coined by Robinson; see Robinson (1985).

[8]   The multiple occupancy theory is different from the view that we have considered above, viz. I survive as both Lefty and Righty. According to the latter view, Lefty and Righty constitute me. But in the former theory, Lefty and Righty just coincide in one particular body. That is, *I* do not exist; only Lefty and Righty do.

two have different relata. When we say that what matters is the R-relation, we are saying that a relation among momentary person-stages is what matters. And when we say that what matters is the I-relation, we are saying that the identity among continuant persons with stages at various times is what matters (cf. Lewis 1976, 20-21). Formally,

(*R-relation*)
For all persons P, P2 matters to P1 iff P1's current stage is R-related to P2's stage.

(*I-relation*)
For all continuant persons C, C2 matters for C1 iff for all person-stages S, S1, S2, S3… Sn are elements of C1 and are also elements of C2.

Let us try to picture this in terms of your life story. This story has a beginning part, a middle part, and an ending part. The beginning part relates to the middle as the middle relates to the end. Thus, the beginning part refers to your birth, the middle refers to your adulthood, and the end refers to your death. For there to be a story, there should be a sequence of events. Your birth came first, then adulthood, and finally your death. But this should not just be a purely ordinal relation. Some events in the beginning part should be intimately related to the middle as middle to the end. This intimate relation is the R-relation. Now your life story is composed of these R-related parts. The aggregate of these parts is your life. Though we could set them apart, these parts essentially compose your story. This aggregate is the I-relation.

The R-relation and the I-relation have different relata. Though this is the case, they are still extensionally consistent with one another. A person is just a maximal set of I-interrelated aggregate of person-stages. Every person-stage is I-related to every other person-stage in the aggregate. That is, each stage in my life story is essentially part of my life story. No other person-stage outside the aggregate could ever be part of it. Since my person-stages are R-related with one another, it follows that the I-relation is just the R-relation when considered extensionally. My person-stages are my stages because they are R-related with one another and since they are R-related with one another and they belong to me, it follows that I am the same continuant person who has those stages.

So how does this bears to what matters? The original Parfit argument gives us a tension between the identity relation and the R-relation. Defenders of the multiple occupancy theory arrest this tension by claiming that identity is nothing really but an aggregate relation between R-related person-stages. As such, we could consistently hold both the identity-is-what-matters view and the identity-is-not-what-matters view.

Furthermore, this implies that the fission case has been misdescribed. To say that Lefty and Righty are not identical to me is true because I am not even present in the case. Only two continuant persons are present, Lefty and Righty. Both the person-stages of Lefty and Righty overlap to one single body before fission. But after fission they continue on to live each of their lives.

Moreover, Lefty and Righty are concerned with each of their own future welfares. This means that the continuant, Lefty, is concerned whether some future person-stage is still psychologically continuous with his present person-stage in the same way that Righty is also concerned about his future person-stage. Defenders of the theory see that there is no need to account for what matters to me because I am not even part of the case; i.e., I do not even exist in the fission case.

There are several objections which could be raised against this view. One has something to do with the assumed ontology of the multiple occupancy theory; another is that the view seems self-defeating.[9]

One objection against the multiple occupancy theory is that it presupposes a four-dimensionalist ontology. Four-dimensionalism is the view that aside from spatial parts, ordinary objects (people included) also have temporal parts.[10] Persons are extended in time as well in space. The Me five minutes ago is a part of Me just like my forefinger is also a part of me. The talk of person-stages and continuant persons presupposes this

---

[9]   There are other objections to the multiple occupancy theory. One is about the semantics of first-person judgments: if the theory is correct, then prior fission, my first-person judgments will be ambiguous between the judgments of Lefty and Righty; see Garrett (1998). Another objection has something to do with counting: if the theory is correct, then prior fission, we should count two people in one body; see Lewis (1976) and Sider (2001).

[10]   For a clear discussion of this view see Sider (2001).

kind of ontology. A person-stage is a temporal part of a particular continuant person. Thus, the aggregate of person-stages just is a continuant person.

There are ways to resist this kind of ontology. One argument for four-dimensionalism hinges on the analogy between space and time. It is claimed that whatever may be said about space could also be said about time. Thus, since there are objects in space, it follows that there are objects in time as well. Since objects in space have spatial parts, it follows that objects in time have temporal parts as well. But there is a certain disanalogy between space and time. We could say that one and the same thing cannot be in two different places at one and the same time. But we cannot say that one and the same thing can be at two different times in one and the same place (cf. Chisholm 1976, 140). As such, it would seem that not everything we could say of space could also be said of time.

Defenders of the multiple occupancy theory could reply that we could have the same theory without presupposing the four-dimensionalist ontology (cf. Markosian 2010). Such a theory could be put in terms of a three-dimensionalist ontology where persons are not composed of person-stages, but are wholly present at different times. On this view, Lefty and Righty are both wholly present in the same space at the same time prior fission.

But this again is hard to understand. How can two numerically different persons exist in the same place at the same time? A three-dimensionalist version of the multiple occupancy theory is just as unintelligible as the idea that two qualitatively similar tables occupy the same place at the same time.

Another objection against the multiple occupancy theory is that its main claim seems self-defeating. The theory cannot consistently hold the common sense view that identity is what matters without distorting our ordinary conception of personal identity. Recall that the main aim of the theory is to reconcile the identity-is-what-matters view and the identity-is-not-what-matters view. Defenders of the theory claim that my identity is just the aggregate of my person-stages and what matters to me is just the continuity of this aggregate. But surely this is not how we understand what matters in one's continuing identity. Common sense tells us that what matters is whether I will be the very same person who will experience some future event, and not whether some future person-stage is related to one of my

person-stages (cf. Parfit 1976 and Sider 2001). As such, the main aim of the multiple occupancy theory is left unsatisfied.

3.3. I am neither Lefty nor Righty

A third way of resisting the fission case is to claim that I would not survive it. None of the fission products will be me. Parfit welcomes this consequence. Neither Lefty nor Righty is identical to me. But since what matters is already contained in this description, he claims that personal identity is not what matters. Our continuing identity does not ground our prudential concern. But there are other ways of interpreting the claim that I am neither Lefty nor Righty without implying the identity-is-not-what-matters view.

One way of interpreting the "I am neither Lefty nor Righty" response is to say that when I divide, there are two equally good candidates for my successor. But since Lefty and Righty are both equally good candidates, it follows that I am neither of them. This is what is known as the best-candidate theory or the closest continuer theory.[11] The motivation behind this theory is that the fission case is not really an argument for the identity-is-not-what-matters view; rather, it only shows that an analysis of personal identity is *extrinsically* grounded.

The best candidate theory tells us that the question whether identity is what matters should be taken independently of the question whether personal identity admits analysis. Unlike Parfit who claims that we should have a negative answer to the former question if we have a positive answer to the latter, we could have positive answers to both questions.

The main claim of the theory is that personal identity admits analysis. But this analysis is extrinsically grounded. Whether I continue to exist depends on whether I have one or two close continuers. Defenders of this view take the fission case as an exemplification of this main claim. After fission, two equally good continuer candidates are available. Since Lefty and Righty are equally good continuer candidates, I have some reason to think that I will no longer survive. My existence, then, depends on whether just one continuer exists.

---

[11]   This view is originally developed and defended by Nozick (1981). A later version of this could be seen in Noonan (2003).

One objection that can be raised against the best candidate theory is as follows: how can my existence depend on the existence or non-existence of some other future person? Suppose we have a possible world where during a fission transplant a nurse dropped half of my brain which is supposed to occupy Righty's body. In this world, I survive as Lefty. Suppose that in another world, the nurse did not drop half of my brain, and thus Righty and Lefty both exist. According to the best candidate theory, in this second world I did not survive. The theory, therefore, gives us two verdicts in these two worlds. In the first world I did survive and in the second world I did not. Thus, my survival depends on the existence or non-existence of some future person who is not even causally related to me. But how can someone who does not exert any causal influence on me cause my death?

3.4. Fission is not even possible

A fourth way of resisting the fission case is to claim that it is not even possible. There are two ways of cashing this out. One way is to claim that though the fission case is *conceptually* possible, it is still *nomologically* or even *metaphysically* impossible.[12] The other way is to claim that we are emotionally ill-equipped to handle the Fission Case; as such, we cannot imagine how this scenario would spell out. Animalists, like Eric Olson, may be said to hold the first view; while Bernard Williams and Richard Swinburne the latter view.[13]

Animalists hold that because we are necessarily biological entities of a certain sort, some governing natural laws necessarily apply to us. And though it is true that we could imagine cases where people divide, this remains to be nomologically impossible. Certain evolutionary laws prohibit the fission of people. Other biological entities, like an amoeba and certain type of cells, do divide. They divide because this is their only way to reproduce. But people are not a biological entity of this sort. As such, though the fission case is conceptually possible, it offends nomological possibility (see Olson 1997, 46-57).

---

[12] This general outlook is also shared by Gendler (2002); Gunnarsson (2008); and Wilkes (1988).

[13] For a recent version of the latter view, see Kind (2004).

There are ways to address this resistance to the fission case. First, the claim that we are necessarily biological entities of a certain sort needs to be argued for. Indeed, we could say of ourselves that we have biological characteristics, and these characteristics are governed by natural laws. But it does not follow from this that we are necessarily biological entities. If all that we are is the preservation of our psychological states, then it is possible to for us to leave the biological body that we have now and be put into an artificial body, but still manage to exist. The burden of proof now lies on the animalists to prove their claim.

Second, even if we are governed by nomological laws, it still seems possible for parts of our brains to be transplanted to another body. Medical science made it possible for us to transplant different organs of our bodies to some other body. Heart, liver, and kidney transplants are now a commonplace. If the brain holds our mental life, then if it were to be transplanted to another body, it seems plausible that the notion of who we are would likely be transferred to that body as well. So, though we are governed by nomological laws, no natural law is broken in imagining the fission case. As such, this resistance to the fission case is not promising.

Another way of cashing out the "fission is not possible" response is to say that we cannot imagine what it would be like for us to undergo fission. There is just no fitting emotional response to this case. When we are confronted with the possibility of fission, there is no correct emotional response if we consider the case from our own point of view.

Suppose that a mad surgeon captures you and announces that he is going to transplant your left cerebral hemisphere into Lefty's body, and your right hemisphere to Righty's. He is going to torture one of the resulting persons and will give the other a million dollars. You can choose which of the resulting persons is going to be tortured and which will receive the sum of money. The mad surgeon is true to his word; he would do whatever it is that you will choose. But the question is how would you choose? Would you choose to give Lefty the money and let Righty suffer the torture? Or would you choose to do otherwise – give Righty the money and let Lefty be tortured?[14]

---

[14] A version of this thought experiment is due to Williams (1973, 46-64); cf. Swinburne (1984, 18).

The fission case tells us that Lefty and Righty are both physically and/or psychologically continuous to you, but they are not exactly you. And there is no marked difference if we think of them as such. We should care for their welfares. It is does not matter whether either one of the two is identical to you or not. What matters is that they are both continuous to you. But if this is the emotion that the fission case wants to elicit from us, then it would have to wait until we develop a different set of attitudes from what we naturally have.

All that we could have in the case where we have to choose between a future torture and a future pleasure is a mixed emotional response. We cannot decide who of Lefty or Righty is to receive torture because there is a risk in the choice. The risk could be spelled out in terms of the notion of identity. I cannot decide which of Lefty or Righty should be tortured after the surgery because I do not, and cannot, know who of the two will be me. I care for my own future welfare. And if it turns out that I had made a wrong choice – i.e. I choose to torture Lefty and not Righty, and it turns out that I am Lefty, then I am doomed to suffer the torture.

If the fission case, and hence Parfit's argument itself, wants us to say that prudential concern is not grounded on our continuing identity, then we should feel that there is no risk in choosing either of the two future outcomes. But since there is a risk involved, it follows that my prudential concern is really grounded on my continuing identity.

A defender of the identity-is-not-what-matters view could reply that the risk that one feels when confronted by this scenario only occurs because it is assumed that identity is what matters. If we abandon this assumption, then no risk would occur. But as this reply runs, it seems to commit to a vicious circle. Identity is not what matters because what matters is the R-relation. The R-relation matters because it does not involve a risk. It does not involve a risk because identity is not what matters.

So, as far as the argument goes, the idea that fission is not possible when we think of it from our point of view holds true. As such, it would seem that fission could be resisted. And thus we have no motivation to accept Parfit's argument for the identity-is-not-what-matters view.

3.5. I survive as either Lefty or Righty

Another plausible way of resisting the fission case is to say that I survive as either of the two resulting persons. Immediately after fission, though both Lefty and Righty are physically and psychologically continuous to me, I would have to be one of them. As such, even if Righty believes that he is me, but actually I am Lefty, then Righty's belief is wrong.

What motivates this response is that if we grant the possibility of the fission case, then either of two things will happen. Either my first-person point of view is preserved in Lefty or it is preserved in Righty. That is, if I can still refer to myself as myself in one of the two resulting persons, then my identity is preserved.

There are problems with this way of resisting the fission case. One is that since both Lefty and Righty are symmetrically related to you, then the only way for you to claim that you are either of the two is to conjure up a Cartesian Ego which grounds the first-person point of view. But since we can never really know whether there is such a thing as a Cartesian Ego, then we could surely deny that the first-person point of view needs to be accounted for.

Furthermore, from the third-person point of view, there is no real difference between the consciousnesses of Lefty and Righty. Both of them claim that they are continuous with me. Lefty says that he remembers all my memories; while Righty claims to have the same memory. As such, there is no real difference between the two claims.

There are responses to these objections. First is that the first-person point of view does not need a Cartesian Ego to ground its existence. We obviously cannot know whether there are Cartesian Egos. But my first-person point of view need not be grounded on a Cartesian Ego or even on anything to know that I have a first-person point of view. My experiences have a unique phenomenology, a feeling of what it is like to experience such and such. And this is what matters to me. It simply is not the case that I sustain having two or more sets of symmetrical phenomenological experiences. That is, I cannot see the world from two points of view at the same time. I cannot have an experience of the world from both Lefty's eyes and Righty's eyes.

Furthermore, since I have a unique set of phenomenological experiences, if I am Lefty, then Lefty also has this unique set. Righty could share

my memories and experiences, but he has a different point of view from me. That is, we could both share the same type of experiences, but since I am Lefty, Lefty's point of view is different from that of Righty's. Righty sees the world from his own first-person perspective. Thus, this shows that Lefty and Righty are not really symmetrically related to me. Since one could preserve my point of view, while the other does not.

The "I survive as either Lefty or Righty" response offers a unique way to preserve the identity-is-what-matters view. Since it matters to me that some future person is identical to me, i.e. that I can still see the world from my own personal perspective, it follows that what matters in my survival is that my first-person point of view is preserved. The concern I have for my own future welfare therefore is grounded on the continuity of my first-person point of view.[15]

This does not imply, however, that the continuity of the first-person point of view is reductive in nature. I am not reducing my existence to this point of view. My personal point of view is in-itself a person-involving concept which could not be reduced to other non-person-involving concept. Nor does it imply that my continuing identity is analyzable in terms of the continuity of my personal point of view. Therefore, like the previous response discussed in the last section, this resistance to the fission case is also plausible.

4. Conclusion

The question about what matters asks whether our prudential concern is grounded on our continuing identity or whether it is grounded on some relation, which only coincides with identity. In section 2, I discussed Parfit's fission case as a motivation for the identity-is-not-what-matters view. In section 3, I have examined five ways of resisting the fission case, and have found that there are plausible ways of showing that the fission case is not a good motivation for the identity-is-not-what-matters view.

---

[15]   This is what Dainton refers to as phenomenological continuity. See Dainton (2008).

Acknowledgments

References

CHISHOLM, R. M. (1976): *Person and Object: A Metaphysical Study.* London: George Allen & Unwin Ltd.

DAINTON, B. (2008): *The Phenomenal Self.* Oxford: Oxford University Press.

GARRETT, B. (1998): *Personal Identity and Self-Consciousness.* London: Routledge.

GENDLER, T. (2002): Personal Identity and Thought-Experiments. *Philosophical Quarterly* 52, No. 206, 34-54.

GUNNARSSON, L. (2010): *Philosophy of Personal Identity and Multiple Personality.* London: Routledge.

KIND, A. (2004): The Metaphysics of Personal Identity and Our Special Concern for the Future. *Metaphilosophy* 35, No. 4, 536-553.

LEWIS, D. (1976): Survival and Identity. In: Rorty, A. O. (ed.): *The Identities of Persons*. Berkeley, (CA): University of California Press, 55-77.

MARKOSIAN, N. (2010): Identifying the Problem of Personal Identity. In: Campbell, J. K, O'Rourke, M. & Silverstein, H. S. (eds.): *Time and Identity.* Boston: MIT Press, 129-148.

NOONAN, H. (2003). *Personal Identity.* 2nd ed. London: Routledge.

NOZICK, R. (1981): *Philosophical Explanations.* Oxford: Clarendon Press.

OAKLANDER, L. N. (1987): Parfit, Circularity, and the Unity of Consciousness. *Mind* 96, No. 384, 525-529.

OLSON, E. T. (1997): *The Human Animal: Personal Identity Without Psychology.* Oxford: Oxford University Press.

PARFIT, D. (1976): Lewis, Perry, and What Matters. In: Rorty, A. (ed.): *The Identities of Persons.* Los Angeles: University of California Press, 91-107.

PARFIT, D. (1971): Personal Identity. *The Philosophical Review* 80, No. 1, 3-27.

PARFIT, D. (1984): *Reasons and Persons.* Oxford: Oxford University Press.

PERRY, J. (1972): Can the Self Divide? *The Journal of Philosophy* 69, No. 16, 463-488.

ROBINSON, D. (1985): Can Amoebae Divide without Multiplying? *Australasian Journal of Philosophy* 63, No. 3, 299-319.

SIDER, T. (2001): *Four-Dimensionalism: An Ontology of Persistence and Time.* Oxford: Oxford University Press.

SWINBURNE, R. (1984): Personal Identity: The Dualist Theory. In: Swinburne, R. & Shoemaker, S.: *Personal Identity.* Oxford: Wiley-Blackwell, 3-66.

WILKES, K. (1988): *Real People: Personal Identity without Thought Experiments*. Oxford: Oxford University Press.

WILLIAMS, B. (1973): The Self and the Future. In: Williams, B.: *Problems of the Self: Philosophical Papers 1956-1972*. Cambridge: Cambridge University Press, 46-63.

Inferential Erotetic Logic in Modelling of Cooperative Problem Solving Involving Questions in the QuestGen Game[1]

PAWEŁ ŁUPKOWSKI – OLIWIA IGNASZAK

ABSTRACT: In the paper problem solving processes that involve reasoning with question are analysed. These reasonings with questions are compared to normative solution scenarios based on A. Wiśniewski's Inferential Erotetic Logic. An on-line game with a purpose QuestGen has been used to gather data for the analysis.

KEYWORDS: Erotetic implication – erotetic search scenario – games with a purpose (GWAP) – inferential erotetic logic – questions – scientific discovery games.

0. Introduction

The main aim of this paper is to present our analysis of solutions of tasks retrieved form the on-line game QuestGen. The game has been designed for collecting the data for research focused on problem solving with

---

✉ Paweł Łupkowski
   Department of Logic and Cognitive Science, Reasoning Research Group
   Institute of Psychology, Adam Mickiewicz University
   Szamarzewskiego 89a, 60-586 Poznań, Poland
   e-mail: Pawel.Lupkowski@amu.edu.pl

✉ Oliwia Ignaszak
   Reasoning Research Group, Institute of Psychology, Adam Mickiewicz University
   Szamarzewskiego 89a, 60-586 Poznań, Poland
   e-mail: oignasz@gmail.com

questioning involved.[2] QuestGen consists of detective-like stories, where players have to solve a presented puzzle – initial problem. They collaborate to achieve the solution playing against the game rules within a time limit. What is crucial from our perspective is that for each story in the game there exists a pre-defined normative solution of a given puzzle which is based on certain logical concepts.

The underpinning of QuestGen stories is Inferential Erotetic Logic (hereafter IEL; see Wiśniewski 1995; 2013b). IEL is a logic which focuses on inferences whose premises and/or conclusions are questions (so called *erotetic inferences*), and which gives criteria of validity of such inferences. Thus it offers a very useful and natural framework for analyses of the questioning process. We can point here to IEL's applications in modelling cognitive goal-directed processes (see Wiśniewski 2003; 2001; 2012; and Urbański & Łupkowski 2010). As a consequence of this line of research, IEL is also used as a theoretical background in the context of empirical research. Moradlou & Ginzburg (2014) present a corpus study aimed at characterising the learning process by means of which children learn to understand questions. The authors assume that for some stages of this process children are attuned to a very simple erotetic logic. Urbański et al. (2014) present research on correlations between the level of fluid intelligence and fluencies in two kinds of deductions: simple (syllogistic reasoning) and difficult ones (erotetic reasoning). The tool used to investigate erotetic reasoning is the Erotetic Reasoning Test which exploits IEL concepts (such as erotetic implication) – see Urbański et al. (2016).[3] Our research reported in this paper is in line with these studies. We explore how the normative yardstick, established with the use of IEL, relates to the real solutions of certain (well defined) problems. The novelty of our approach is the use of an on-line game to collect the necessary language data. Such a solution is

---

[2] The idea of QuestGen is presented in Łupkowski (2011b). Details of the implementation may be found in Łupkowski & Wietrzycka (2015); see also Łupkowski et al. (2015).

[3] It is worth mentioning that IEL-based concepts have proven useful for many other domains, including the Turing test's adequacy (cf. Łupkowski 2011a; Łupkowski & Wiśniewski 2011); abductive reasoning (see Komosinski et al. 2014); or proof-theory (see Wiśniewski 2004b; Leszczyńska 2004; 2007; Wiśniewski & Shangin 2006; Urbański 2001a; 2001b; 2002).

inspired by a successful use of game-like elements in the scientific domain (see e.g. *Foldit* (Cooper et al. 2010), *Galaxy Zoo* (Darg et al. 2010), or *Wordrobe* (Venhuizen et al. 2013)). What these games have in common is that when playing the game (and having fun) players solve a serious scientific problem (or rather well defined parts of it). Games of this type are referred to as *games with a purpose* (GWAP – Von Ahn 2006) or s*cientific discovery games* (see Cooper et al. 2010).[4]

The outline of the paper is the following. In the first section we introduce basic notions and concepts from IEL, which are used for the Quest-Gen design and in the following modelling of solutions. Second section covers the game and the tasks used as well as overview of the collected data. In the third section we model and discuss selected solutions of Quest-Gen tasks. In the last section we address possibilities of future developments and improvements of our approach.

1. Erotetic inferences and their modelling in Inferential Erotetic Logic

1.1. Language $L^?_{cpl}$

In what follows, we will use propositional language with questions. The reason for this is that the expressive power of such a language is just-sufficient for the analysis presented.[5]

We will use $Q, Q^*, Q_1, \ldots$ as metalinguistic variables for questions and A, B, C, D, possibly with subscripts, as metalinguistic variables for declarative well-formed formulas, X, Y, … represent sets of declarative well-formed formulas. We will use **d**Q for the set of direct answers to a question Q.

---

[4]   See also an overview of such games presented in Kleka & Łupkowski (2014); Łupkowski & Dziedzic (2016); and Dziedzic (2016).

[5]   IEL introduces a series of semantic concepts about questions. Semantics for questions are provided by the means of the so called Minimal Erotetic Semantics (MiES for short) – for more details see Wiśniewski (2013b, Chap. 4). It is worth stressing that MiES allows for enriching any formal language with questions, provided that this language allows for partitioning declarative formulas into true and untrue ones (cf. Wiśniewski 1996; 2001; 2013b).

Following Wiśniewski (2013b, Chap. 2) we present language $L^?_{cpl}$. Let us start with L_{cpl} which is the language of Classical Propositional Logic (CPL, for short). Language L_{cpl} contains the following primitive connectives: ¬ (negation), → (implication), ∨ (disjunction), ⊥ (exclusive disjunction), ∧ (conjunction), ↔ (equivalence). The concept of a *well-formed formula* (wff for short) is defined in a traditional manner.

We use p, q, r, s, p_1, … for propositional variables. CPL-valuation (v) is understood in a standard way.

At this point, we introduce another object-level language – $L^?_{cpl}$. The vocabulary of the new language is the vocabulary of L_{cpl} extended with the following signs: ?, {, }, and the comma. This allows us to represent the *erotetic formulas* (e-formulas) of the language. Consequently we say that $L^?_{cpl}$ has two categories of well-formed expressions: *declarative well-formed formulas* (hereafter d-wffs) and *erotetic well-formed formulas* (i.e. questions, hereafter e-wffs). The categories of d-wffs and e-wffs are disjoint. D-wffs of $L^?_{cpl}$ are simply well-formed formulas of L_{cpl}, and e-wffs of $L^?_{cpl}$ are expressions of the form:

(1) $?\{A_1, …, A_n\}$

where $n > 1$ and A_1, …, A_n are nonequiform (i.e. pairwise syntactically distinct) d-wffs of $L^?_{cpl}$ (i.e. CPL-wffs). If $?\{A_1, …, A_n\}$ is a question, then each of the d-wffs A_1, …, A_n is called a direct answer to the question. As we can see, each question of $L^?_{cpl}$ has a finite set of direct answers and each question has at least two direct answers.[6]

Any question of the form (1) may be read (informally):

Is it the case that A_1, or …, or is it the case that A_n?

In what follows, for the sake of simplicity, we will adopt some notational conventions.

A simple yes-no question (i.e. a question whose set of direct answers consists of a sentence and its classical negation) of the form:

---

[6] It is worth mentioning that in IEL also other types of questions (including the ones with infinite sets of possible answers) are considered – see Wiśniewski (1995, Chap. 3).

(2)   ?{$A, \neg A$}

are simply presented as:

(3)   ?A

Questions of the form (3) can be read (informally):

Is it the case that A?

1.2. Erotetic implication

In IEL erotetic inferences of two kinds are analysed:

1. *Erotetic inferences of the first kind*, where a set of premises consists of declarative sentence(s) only, and an agent passes from it to a question – grasped under the notion of *question evocation* (see Wiśniewski 2013b, Chap. 6); and
2. *Erotetic inferences of the second kind*, where a set of premises consists of a question and possibly some declarative sentence(s) and an agent passes from it to another question – grasped under the notion of *erotetic implication* (e-implication for short).

In this paper we will be interested only in the erotetic inferences of the second kind. E-implication is a semantic relation between a question Q, a (possibly empty) set of declarative well-formed formulas X, and a question Q_1. It is an ordered triple $\langle Q;X;Q_1 \rangle$, where Q is called an interrogative premise or simply initial question, the elements of X are declarative premises and the question Q_1 is the conclusion or the implied question – see Wiśniewski (2013b, 51-52).

The intuition behind e-implication might be expressed as follows. Let us imagine an agent who is trying to solve a certain (possibly) complex problem. The problem is expressed by her initial question (Q). We assume that the agent does not have resources to answer the initial question on her own. Thus the initial question has to be processed/decomposed. This decomposition is aimed at replacing the initial question with an auxiliary question – Q_1. The auxiliary question obtained as a result of the decomposition process should have certain characteristics. First of all, it should stay on the main topic. In other words, no random questions should appear here.

However, the main characteristic that we are aiming at here is that the answer provided to the auxiliary question should be at least a partial answer to the initial question (i.e. it should narrow down the set of direct answers to the initial question, see Wiśniewski 2013b, 43). It should bring our agent closer to solving the initial problem. Summing up, we can perceive the discussed process of replacing one question with an auxiliary one as a well-motivated step from the problem-solving perspective. Before we provide a formal definition of e-implication we will introduce the necessary concepts of MiES. The basic semantic notion to be used here is that of a partition (see Wiśniewski 2013b, 25-30).

Definition 1 (Partition of the set of d-wffs)

Let $D_L{}^?{}_{cpl}$ designate the set of d-wffs of $L^?{}_{cpl}$. A partition of $D_L{}^?{}_{cpl}$ is an ordered pair:

$$\mathsf{P} = \langle \mathsf{T_P}, \mathsf{U_P} \rangle$$

where $\mathsf{T_P} \cap \mathsf{U_P} = \varnothing$ and $\mathsf{T_P} \cup \mathsf{U_P} = D_L{}^?{}_{cpl}.$

Intuitively, $\mathsf{T_P}$ consists of all d-wffs which are true in P, and $\mathsf{U_P}$ is made up of all the d-wffs which are untrue in P (see Wiśniewski 2013b, 25).

Definition 2 (Partition of the language $L^?{}_{cpl}$)

By a partition of the language $L^?{}_{cpl}$ we mean a partition of $D_L{}^?{}_{cpl}$.

The concept of the partition is very general, thus Wiśniewski (2013b, 26, 30) introduces the class of admissible partitions being a non-empty subclass of all partitions of the language. This step allows for defining useful semantic concepts.

Definition 3 (Admissible partition of $L^?{}_{cpl}$)

A partition $\mathsf{P} = \langle \mathsf{T_P}, \mathsf{U_P} \rangle$ of $L^?{}_{cpl}$ is admissible iff for some CPL-valuation v:

(i) $\mathsf{T_P} = \{A \in D_L{}^?{}_{cpl}: v(A) = \mathbf{1}\}$, and
(ii) $\mathsf{U_P} = \{B \in D_L{}^?{}_{cpl}: v(B) = \mathbf{0}\}.$

The set of truths of an admissible partition of $L^?{}_{cpl}$ equals the set of d-wffs which are true under the corresponding CPL-valuation.

Partitioning of the language concerns only declarative formulas. A question is neither in T_P nor in U_P, for any partition P – MiES does not presuppose that questions are true or false (cf. Wiśniewski 2013b, 26). As a counterpart of truth for declarative formulas, for questions we introduce the notion of soundness (see Wiśniewski 2013b, 37).

Definition 4 (Soundness)

A question Q is sound in a partition P iff $\mathbf{d}Q \cap T_P \neq \varnothing$.

A question is sound (in a partition) iff at least one direct answer to this question is true in the partition.

Now we need to introduce the definition of multiple-conclusion entailment (mc-entailment) – see Shoesmith & Smiley (1978) and Wiśniewski (2013b, 33).

Definition 5 (Multiple-conclusion entailment)

Let X and Y be sets of d-wffs of language $L^?_{cpl}$. We say that X mc-entails Y in $L^?_{cpl}$ (in symbols $X \vDash_L{}^?_{cpl} Y$) iff there is no admissible partition $P = \langle T_P, U_P \rangle$ of $L^?_{cpl}$ such that $X \subseteq T_P$ and $Y \subseteq U_P$.

The intuition behind mc-entailment is that it holds between the sets of d-wffs X and Y iff the truth of all d-wffs in X warrants the presence of at least one true d-wff in Y.

Now we may introduce the definition of erotetic implication (see Wiśniewski 2013b, 68).

Definition 6 (Erotetic implication)

A question Q implies a question Q_1 on the basis of a set of d-wffs X (in symbols, $\mathsf{Im}(Q,X,Q_1)$) iff:

    (1) for each $A \in \mathbf{d}Q$: $X \cup \{A\} \vDash_L{}^?_{cpl} \mathbf{d}Q_1$, and
    (2) for each $B \in \mathbf{d}Q_1$ there exists a non-empty proper subset Y of $\mathbf{d}Q$ such that $X \cup \{B\} \vDash_L{}^?_{cpl} Y$.

The first clause of the above definition warrants the *transmission of soundness* (of the implying question Q) and truth (of the declarative premises in X) into soundness (of the implied question Q_1). The second clause expresses the property of *open-minded cognitive usefulness* of e-implication,

that is, the fact that each answer to the implied question Q_1 narrows down the set of direct answers to the implying question Q.

If a set X of declarative formulas is empty, an e-implication of this sort is called a *pure e-implication* (see Wiśniewski 2013b, 76).

Let us now consider simple examples of e-implication, starting with a pure one.

(4) $\mathsf{Im}(?\{A, B \vee C\}, \varnothing, ?\{A, B, C\})$

In (4) Q is $?\{A, B \vee C\}$, Q_1 is $?\{A, B, C\}$ and set X is empty. The first condition for a pure e-implication is met. The same applies to the second condition. One may observe that the proper subset Y of the set of direct answers to the question Q is the following: (*i*) for the direct answer A to question Q_1 it is $\{A\}$, (*ii*) when it comes to the answer B it is $\{B \vee C\}$, and (*iii*) for the answer C it is also $\{B \vee C\}$.

Let us consider another example:

(5) $\mathsf{Im}(?A, A \leftrightarrow B; ?B)$

In (5) we may also notice that two conditions of e-implication are met. $?A$ is a simple yes-no question, thus the set of direct answers to this question is $\{A, \neg A\}$. The set of direct answers to the implied question $?B$ is $\{B, \neg B\}$. For each direct answer to $?A$, if it is true and the premise is true, then at least one direct answer to $?B$ is true (it is B for A and $\neg B$ for $\neg A$). As for the second condition of the e-implication, it is also met. The required proper subset Y of the set of direct answers to the implying question $?A$ is the following: (*i*) for the direct answer B to the question $?B$ it is $\{A\}$, and (*ii*) for the direct answer $\neg B$ to the question $?B$ it is $\{\neg A\}$.

1.3. Erotetic search scenarios

When we think about e-implication used for decomposing questions as described above it is easy to imagine that it might be repetitively applied while solving a particular complex problem. The intuition behind such a process is perfectly grasped under Wiśniewski (2013b, 103):

> **EDP** (*Erotetic Decomposition Principle*) Transform a principal question into auxiliary questions in such a way that: (a) consecutive auxiliary questions are dependent upon the previous questions and, possibly,

answers to previous auxiliary questions, and (b) once auxiliary questions are resolved, the principal question is resolved as well.

This leads us to the notion of an erotetic search scenario (e-scenario in short). As the name suggests it is a scenario for solving a problem expressed in the form of a question. The pragmatic intuition behind the e-scenario is that it

(…) provides information about possible ways of solving the problem expressed by its principal question: it shows what additional data should be collected if needed and when they should be collected. What is important, an e-scenario provides the appropriate instruction for every possible and just-sufficient, i.e. direct answer to a query: there are no "dead ends". (Wiśniewski 2013a, 110)

In this paper – following Wiśniewski (2013b) – we will present the e-scenario as a family of interconnected sequences of the so-called erotetic derivations.[7] It is worth mentioning that e-scenarios can also be viewed as labelled trees (see Leszczyńska-Jasion 2013).

Erotetic derivation is defined as follows (cf. Wiśniewski 2013b, 110–111):

Definition 7 (Erotetic derivation)

A finite sequence $\mathbf{s} = s_1, \ldots, s_n$ of wffs is an erotetic derivation (e-derivation for short) of a direct answer A to question Q on the basis of a set of d-wffs X iff $s_1 = Q$, $s_n = A$, and the following conditions hold:

(1) for each question s_k of $\mathbf{s}$ such that $k > 1$:
   (a) $\mathbf{d}s_k \neq \mathbf{d}Q$,
   (b) s_k is implied by a certain question s_j which precedes s_k in $\mathbf{s}$ on the basis of the empty set, or on the basis of a non-empty set of d-wffs such that each element of this set precedes s_k in $\mathbf{s}$, and
   (c) s_{k+1} is either a direct answer to s_k or a question;

---

[7]   See also Wiśniewski (2001; and 2003) where the idea of e-scenarios has been presented for the first time.

(2) for each d-wff s_i of **s**:
   (a) $s_i \in X$, or
   (b) s_i is a direct answer to s_{i-1}, where $s_{i-1} \neq Q$, or
   (c) s_i is entailed by a certain non-empty set of d-wffs such that each element of this set precedes s_i in **s**.

The e-derivation is *goal-directed*: it leads from an initial question Q to a direct answer to this question. Clause (1a) of the above definition requires that an auxiliary question (i.e. a question of an e-derivation different from Q) appearing in an e-derivation should have different direct answers than the initial question Q. Clause (1b) amounts to the requirement that each question of the e-derivation which is different from the initial question Q must be e-implied by some earlier item(s) of the e-derivation. Clause (1c) requires that an immediate successor of an auxiliary question in the e-derivation must be a direct answer to that question or a further auxiliary question. Clause (2) enumerates reasons for which a d-wff may enter an e-derivation. Such a d-wff may be: (2a) an element of a set of d-wffs X; (2b) a direct answer to an auxiliary question; (2c) a consequence of earlier d-wffs.

Definition 8 (Erotetic search scenario)

A finite family Σ of sequences of wffs is an erotetic search scenario (e-scenario for short) for a question Q relative to a set of d-wffs X iff each element of Σ is an e-derivation of a direct answer to Q on the basis of X and the following conditions hold:

(1) $\mathbf{d}Q \cap X = \varnothing$;
(2) contains at least two elements;
(3) for each element $\mathbf{s} = s_1, \ldots, s_n$ of Σ, for each index k, where $1 \leq k < n$:
   (a) if s_k is a question and s_{k+1} is a direct answer to s_k, then for each direct answer B to s_k: the family contains certain e-derivation $\mathbf{s^*} = s^*_1, s^*_2, \ldots s^*_m$ such that $s_j = s^*_j$ for $j = 1, \ldots, k$, and $s^*_{k+1} = B$;
   (b) if s_k is a d-wff, or s_k is a question and s_{k+1} is not a direct answer to s_k, then for each e-derivation $\mathbf{s^*} = s^*_1, s^*_2, \ldots s^*_m$ in Σ such that $s_j = s^*_j$ for $j = 1, \ldots, k$ we have $s_{k+1} = s^*_{k+1}$.

E-derivations being elements of an e-scenario will be called *paths* of this e-scenario.

For our purposes notions of query of an e-derivation (cf. Wiśniewski 2013b, 112) and query of an e-scenario (cf. Wiśniewski 2013b, 113) will also be needed.

Definition 9 (Query of an e-derivation)

An element s_k (where $1 < k < n$) of an e-derivation $\mathbf{s} = s_1, \ldots, s_n$ is a query of $\mathbf{s}$ if s_k is a question and s_{k+1} is a direct answer to s_k.

Definition 10 (Query of an e-scenario)

A query of an e-scenario is a query of a path of the e-scenario.

As an illustration of the above concepts, let us consider a simple example – see Figure 1. The initial question of our exemplary e-scenario is $?p$. Only one declarative premise is employed here, namely $p \leftrightarrow q$. This e-scenario contains two paths (i.e. two e-derivations):

(6a) $?p, p \leftrightarrow q, ?q, q, p$
(6b) $?p, p \leftrightarrow q, ?q, \neg q, \neg p$

$$?p$$
$$p \leftrightarrow q$$
$$?q$$

$$q \qquad\qquad \neg q$$
$$p \qquad\qquad \neg p$$
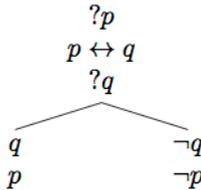
Figure 1: Example of an e-scenario for the question $?p$
relative to the premise $p \leftrightarrow q$

The e-scenario has only one query, i.e. $?q$. The query is e-implied by the initial question and the declarative premise (see e-implication scheme (5)).

An e-scenario might be viewed as providing a search plan for an answer to the initial question. This plan is relative to the premises a questioner has, and leads through auxiliary questions (and the answers to them) to the answer to the initial question. Each path of an e-scenario leading from the

root to one of the leaves represents one of the ways in which the process of solving the initial problem might go. This allows us to consider issues referred to as *distributed internal question processing* (see Wiśniewski 2013b, 105).

The key feature of e-scenarios is that auxiliary questions appear in them on the condition that they are e-implied. Thus we may use e-scenarios to provide some insights into questioning strategies. This approach is efficient for contexts where a questioner wants to obtain an answer to the initial question, which should not be asked directly (as e.g. in the Turing test situation, where asking a direct question 'Are you a human or a machine?' would be fruitless as a satisfactory way of obtaining a solution to the problem of agent identification).[8] To obtain an answer to the initial question, a questioner usually asks a series of auxiliary questions in these situations. Answers to these auxiliary questions build up to be an answer to the initial one. It is easy to imagine a context such as this in real life situations, as for example while teaching, when we want to check if our student really understands a given problem. Figure 2 presents a natural language example of a questioning plan which has the structure of an e-scenario.

Is it the case that there is a zero voltage reading?
When there is a non-zero voltage reading, there is a difference in charge.
If there is a zero voltage reading, there is no difference in charge.
There is a difference in charge if and only if the voltage is "created" or "used up".
In a switch, is there any voltage being "created" or "used up"?

Yes	**No**
There is a difference in charge.	There is no difference in charge
There is a non-zero voltage reading.	There is a zero voltage reading.

Figure 2: Example of a questioning plan with an e-scenario structure. The example is based on a tutor-student dialogue from The Basic Electricity and Electronics Corpus (see Rosé et al. 1999), file BEE(F), stud37

---

8 See Łupkowski (2011a), Urbański & Łupkowski (2010); and also Genot (2009) and Genot & Jacot (2012) for the discussion of these issues in the framework of Interrogative Games.

2. QuestGen – the game with a purpose used
for collecting the data

The idea of the game was presented in Łupkowski (2011b, 89-91). The aim of the QuestGen game is to engage players in generating a large collection of questions for a certain piece of story written in natural language. What is crucial from our point of view is that each story used in QuestGen is based on an e-scenario, which serves as a normative yardstick for the pre-established solution of this story. This allows us to compare and discuss the normative view on a given solution and real solutions retrieved via the game.

An important disclaimer is needed at this point. Our claim here is not that logical concepts are the ultimate explanation of the gathered linguistic data (i.e. that people are/or should process questions according to IEL) – see the detailed discussion in Łupkowski (2016). Our approach here is rather that logic provides a very useful normative yardstick to study, describe and analyse these phenomena (see Stenning & Van Lambalgen 2008, 130). Logic may be, and is, successfully applied within research concerning actual human reasoning as reviewed and discussed by Urbański (2011) or Sedlár & Šefránek (2014). For this context, IEL offers convenient tools for modelling natural language phenomena and for their better understanding. Using these tools we will consider the issue of motivation for certain moves in the game (or in a broader context, in a dialogue). On the other hand, empirical data (like that retrieved from language corpora) allows for *better tailored logical concepts* – see e.g. concepts of weak erotetic implication introduced by Urbański et al. (2016) as a consequence of analysis of solutions to Erotetic Reasoning Test tasks.

In the QuestGen game, two randomly chosen players are engaged in solving a detective puzzle. One of them plays as the Detective, while the other is called the Informer. The aim for the Detective is to solve the presented puzzle by questioning the Informer. Each story in the game has two formulations (one for the Detective and one for the Informer), containing all the additional data necessary to solve the puzzle. Each story should be solved within a given time limit.

The basic rules for the game are the following:

1. The Detective is allowed only simple yes/no questions.

2. The Detective is not allowed to ask directly for the solution to a puzzle.
3. The Detective may ask as many questions as she/he wants (within the time limit).
4. The Informer should provide information accordingly to her/his version of the story.

QuestGen, in the version described in this paper, consists of six stories, entitled: *Hrabina* (Countess), *Teleturniej* (Quiz), *Zaginiony chłopak* (Lost boyfriend), *Tablet*, *Arsen L.*, *Bomba* (Bomb). The stories were written (or adapted) by the second author of the paper.

The process of preparing all these stories started with an appropriate e-scenario. In what followed the plot was built on the basis of the e-scenario structure. The *Arsen L.* story is a slightly modified example of the e-scenario in action taken from Wiśniewski (2003, 392). The *Bomb* is adapted from one of the aforementioned Erotetic Reasoning Test tasks presented in Urbański et al. (2016, 4-6). Story *Countess* is based on the e-scenario from Łupkowski (2010, 78); *Quiz* uses the e-scenario from Wiśniewski (2013b, 110). Stories *Lost boyfriend* and *Tablet* are both based on the same e-scenario taken from Wiśniewski (2004a, 16). The idea behind this decision, was to check whether two stories with the same logical structure will be solved in a similar manner. Table 1 presents the basic characteristics of stories used in QuestGen.

Table 1: Stories used in QuestGen: Informer (I), Detective (D)

Title	Premises	Facts (I)	Words (D)	Words (I)	Time limit
Countess	3	3	150	146	3 min
Quiz	3	3	113	110	3 min
Lost boyfriend	3	4	146	118	3 min
Tablet	3	4	111	132	3 min
Arsen L.	4	3	155	113	4 min
Bomb	6	6	169	169	6 min

Let us now take a closer look on one of the stories. Due to the space restrictions we will not be able to present all six stories, they can be found at the project's webpage.[9] It is also worth mentioning that all the data collected with the use of QuestGen are now included in the Erotetic Reasoning Corpus (ERC)[10]. ERC constitutes a data set for research on natural question processing – see Łupkowski et al. (2017). All the data is in Polish, however the tag-set used for the annotation allows for the data analysis for English-speaking researchers. For this paper we have decided to present the story *Bomb*. The reasons for this are twofold. Firstly, it is the most complex story in QuestGen (with the highest number of premises for Detective and facts for Informer – see Table 1). Secondly, the story has been adapted from the Erotetic Reasoning Test task (see Urbański et al. 2016, 4-6), and thus it is possible to compare and discuss the results from the test and from Quest-Gen. Changes to the original version cover different names used in the story and different wires colours (changed from green, red and orange to purple, orange and pink, in order to avoid popular references to e.g. films). Let us remind that each QuestGen story has its two formulations, one for Detective and one for Informer. Below we present a translation of the story from the game.

Bomb: Detective version

There was a bomb planted in the main train station of Nibyjunkcja. You coordinate actions of the sapper unit. The chief of the local police managed to establish the following evidence, which he is sharing with you now:

1. There are three wires in the bomb: purple, orange and pink.
2. To disarm the bomb either the purple or the orange wire must be cut. Cutting the wrong wire will cause an explosion.
3. If the bomb has been planted by Anthony, cutting the purple wire will disarm it.
4. If the bomb has been planted by Roger, cutting the orange wire will disarm it. Moreover, no one but Roger would have used the orange wire.

---

[9]   https://plupkowski.wordpress.com/projects/questgen-game/.

[10] https://ercorpus.wordpress.com/.

5. If the bomb has not been planted on an even day of the month, the culprit is Anthony.
6. The bomb has been planted by Anthony, or by Roger, or by someone else.

Which wire should be cut to disarm the bomb?

Before you will make a decision you can ask questions to the chief of security, who is responsible for the place where the bomb is planted. Remember the time is limited. You can ask only yes/no questions. There is no sense to ask directly which wire should be cut to disarm the bomb, because the chief of security does not know this.

Bomb: Informer version

You are the chief of security at the train station where the bomb was planted. The coordinator of the sapper unit is trying to establish which wire to cut in order to disarm the bomb. He has the following facts at his disposal:

1. There are three wires in the bomb: purple, orange and pink.
2. To disarm the bomb either the purple or the orange wire must be cut. Cutting the wrong wire will cause an explosion.
3. If the bomb has been planted by Anthony, cutting the purple wire will disarm it.
4. If the bomb has been planted by Roger, cutting the orange wire will disarm it. Moreover, no one but Roger would have used the orange wire.
5. If the bomb has not been planted on an even day of the month, the culprit is Anthony.
6. The bomb has been planted by Anthony, or by Roger, or by someone else.

After checking security cameras you know the following:

1. The bomb was planted by Roger YES
2. The bomb was planted by Anthony NO
3. The bomb was planted on an odd day of the month. NO
4. The bomb was planted on an even day of the month. YES
5. The bomb was planted by someone else. NO
6. To disarm the bomb the orange wire should be cut. YES

Questgen

Gra rozpoczęta

Koniec gry

Jesteś detektywem

pierwsze podejrzenie padło na wnuka-urwisa, który często, pomimo zakazów hrabiny, jeździ na deskorolce wewnątrz posiadłości. Ponieważ nie byłby to pierwszy wybryk wnuka, hrabina może już na podstawie wcześniejszych doświadczeń wywnioskować, że:

1. Jeśli to wnuk rozbił wazę, to ze strachu przed konsekwencjami nie pojawił się tego dnia na podwieczorku.
2. Jeżeli wnuk był sprawcą, to po kilku godzinach spędzonych w ukryciu zaproponował gosposi pomoc w pracach domowych, ponieważ dręczyły go wyrzuty sumienia.
3. Jeżeli gosposia potwierdzi, że wnuk nie zjadł podwieczorku, a po jakimś czasie zaoferował jej pomoc, to hrabina będzie pewna, że to on stłukł jej ulubioną wazę.

Zanim jednak wyda wyrok, hrabina musi skonsultować się z gosposią, która pomoże jej w rozwiązaniu zagadki i znalezieniu odpowiedzi na pytanie "czy to wnuk stłukł wazę?".

Przebieg gry

Pytanie: Czy wnuk zjadł podwieczorek?
Odpowiedz: **Nie**
Pytanie: Czy wnuk zaproponował pomoc w pracach domowych?

Zadaj pytanie:

wpisz swoje pytanie　　wyślij pytanie

Gra rozpoczęta

Koniec gry

Jesteś sędzią

Od wielu lat jesteś gosposią w posiadłości hrabiny. Częstym gościem jest w niej wnuk właścicielki, któremu zdarza się niechcący powodować straty w trakcie zabawy. Zaaferowana hrabina zwraca się do Ciebie, ponieważ podejrzewa, że wnuk rozbił jej ulubioną wazę z dynastii Qing. Hrabina uważa, że:

1. Jeśli to wnuk rozbił wazę, to ze strachu przed konsekwencjami nie pojawił się tego dnia na podwieczorku.
2. Jeżeli wnuk był sprawcą, to po kilku godzinach spędzonych w ukryciu zaproponował gosposi pomoc w pracach domowych, ponieważ dręczyły go wyrzuty sumienia.
3. Jeżeli gosposia potwierdzi, że wnuk nie zjadł podwieczorku, a po jakimś czasie zaoferował jej pomoc, to hrabina będzie pewna, że to on stłukł jej ulubioną wazę.

Ty możesz przekazać hrabinie następujące informacje:
1. Wnuk nie pojawił się tego dnia na podwieczorku. TAK
2. Wnuk zaproponował ci pomoc w pracach domowych. NIE
3. Wnuk rozbił ulubioną wazę hrabiny. NIE

Przebieg gry

Pytanie: Czy wnuk zjadł podwieczorek?
Odpowiedz: **Nie**
Pytanie: Czy wnuk zaproponował pomoc w pracach domowych?

Odpowiedz:

◉ TAK
○ NIE
○ NIE WIEM

wyślij odpowiedź

Figure 3: A screenshot of QuestGen game. The Detective's screen is visible on the top, while the Informer's screen is presented below. For the Detective's part the story is presented in the left column. Below the story there is a field for typing question. The right column presents the game as it is progressing. For the Informer's screen we also have the story in the left column, but below it there are pre-established answers to Detective's question to be used by the Informer.

As we have stressed above the structure of the story is based on the e-scenario, which is presented in Figure 4 (see Urbański et al. 2016, 38). Propositional variables represent the following sentences:

p – Cutting the purple wire disarms the bomb.
q – Cutting the orange wire disarms the bomb.
v – Cutting the pink wire disarms the bomb.
s – The bomb has been planted by Anthony.
r – The bomb has been planted by Roger.
t – The bomb has been planted on an even day of the month.
u – The bomb has been planted by someone else.

$$?\{p, q, v\}$$
$$p \vee q$$
$$s \rightarrow p$$
$$r \leftrightarrow q$$
$$\neg t \rightarrow s$$
$$r \bot s \bot u$$
$$?\{r, \neg r\}$$

$$r \qquad\qquad \neg r$$
$$q \qquad\qquad \neg q$$
$$\qquad\qquad\quad p$$
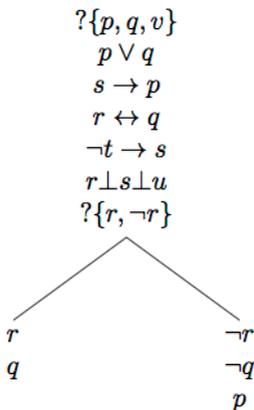
Figure 4: E-scenario for the *Bomb* story

We treat the e-scenario as presenting the normative solution for the given puzzle. For our Detective it will be enough to ask only one question, namely *Was the bomb planted by Roger?* (?$\{r, \neg r\}$). After obtaining the affirmative

answer from the Informer (r), the Detective would reach the answer to the initial question – the orange wire should be cut in order to disarm the bomb (q). Such a procedure is optimal in the sense that no spare, non-necessary auxiliary questions are asked. Auxiliary question ?$\{r, \neg r\}$ is e-implied by the initial question ?$\{p,q,v\}$ on the basis of premises $p \vee q$ and $r \leftrightarrow q$.

In what follows, we will say that the solution to a given puzzle in Quest-Gen is *correct and normative*, when the Detective will provide pre-established solution and, what is more, she/he will reach this solution by asking auxiliary questions accordingly to e-scenario used as an underpinning for the story. (It is worth mentioning that there are puzzles in QuestGen that require asking more than one auxiliary question, in these cases the order of asking auxiliary questions is not important.) We would say that the solution to a given QuestGen puzzle is *correct but non-normative* in cases, when Detective will reach the pre-established solution, however the process of reaching this solution does not involve asking auxiliary questions from the appropriate e-scenario. In cases where more than one auxiliary question is required also solutions where Detective does not ask all the required auxiliary questions are counted as non-normative. For cases where the answer to the puzzle is different than the pre-established one, we say that the solution is *not correct*.

To improve readability in the following analysis we propose to present solutions in form of schemata based on an e-scenario. Such a schema for the normative solution of the *Bomb* puzzle is presented in Figure 5. In the root we have auxiliary question that should be asked by the Detective. Below the root we have answers that might be provided by the Informer (Y – 'yes', N – 'no'). Information that should be provided accordingly to the Informer's version of the story is circled. At the leaf a solution to the initial question (which is delivered by the Detective) is presented.
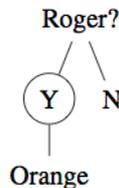
Figure 5: The schema of the normative solution for the Bomb puzzle.
See e-scenario in Figure 4

3. An analysis of solutions to QuestGen tasks

The data analysed in this paper has been recorded during February 2015. QuestGen was published on-line. Anyone, who completed the registration could play the game. Each randomly chosen pair of players went through all six stories of QuestGen. For each story the players switched roles, from the Detective to the Informer and *vice versa*. Players were not supervised in any way, they were just playing the game. (Although an information about the scientific aim of the game was provided in the *Contact* section of the game web-site). Overall we have collected 116 game transcripts from 40 players. The general solution statistics for the study sample (all six stories) is the following: 91 solutions are *correct*, out of which 44 are *normative*, i.e. solved exactly accordingly to the e-scenario underpinning a given story. In 18 cases Detectives provided *incorrect* solutions and in 7 they did not provided *any solution* (mostly due to the time constraints).

There are several regularities which may be observed across the collected data. First of all, for all six stories we observe solutions which may be classified as correct and normative. What is more for each story there are *more* correct than incorrect solutions, but the majority of correct solutions are not normative (let us remind here that this means, that Detectives gave the correct answer to the initial question but they reached it not in the way which is predicted by the underlying e-scenario). The summary of this data is presented in Table 2.

Table 2: Number of correct and correct and normative solutions
to the QuestGen stories

Title	Correct solutions	Correct and normative solutions
Countess	8	4
Quiz	14	6
Lost boyfriend	17	7
Tablet	18	10
Arsen L.	15	11
Bomb	18	6

Disproportion observed between correct and correct and normative solutions provides a good basis for studying different strategies of solving QuestGen stories. We provide such a detailed analysis for the *Bomb* story below.

What is also visible is the tendency to learn how to solve QuestGen stories. Tables 1 and 2 present this stories in the order in which they were presented to our players. It may be noticed that with each story, players were getting better and more accurate in solving them.

On the basis of the all gathered data two other general tendencies may be derived. It seems that our players tend to process one premise after another in the order in which they appear in a given story. We often observe that the questions asked and their order reflect the order of premises. What is more, it is often the case that premises are paraphrased by players, mostly by reformulating negative sentences into affirmative ones.

Last but not least, it is also worth to mention that QuestGen players comprised themselves to the game rules (which is important, as QestGen is simply an on-line game and the process of data collection is not supervised). All questions asked in the game were (exactly as required) yes/no questions, and forbidden questions (i.e. asking directly for the solution) were rare. The low number of games without a solution being provided by the Detective suggests also that the difficulty level of stories and time constraints were chosen adequately.

Let us now focus on the story presented in details in the previous section, i.e. the *Bomb*. For this story 20 solutions were gathered. Out of these 19 ended with the solution given by the Detective, 18 solutions were correct and 1 was not correct (the Detective pointed out to the pink cable – see scheme D1 below). 1 game ended without a solution being delivered by the Detective. In this case the Detective asked a forbidden question ("Purple one?") and then the game ended. Let us now take a closer look on correct solutions. From the normative point of view only one auxiliary question was needed to reach the solution – see Figures 4 and 5. Only six out of eighteen correct solutions were reached exactly in the way predicted by the normative model. However more complex solutions were also recorded. Players asked more than one question usually referring to premises about

the day the bomb has been planted. Below we present them in a schematic form.

Planted on an even day?

Y N

Roger?

Y N

Orange

(B1)

Roger?

Y N

Planted on an odd day?

Y N

Orange

(B2)

Not Roger?

Y N

Planted on an even day?

Y N

Orange

(B3)
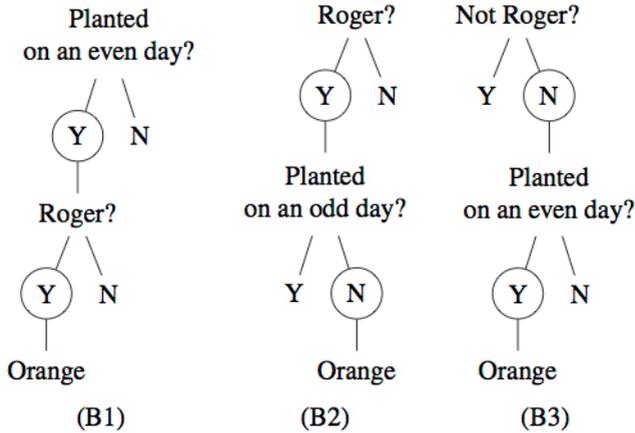
Solution (B1) starts by asking an auxiliary question concerning the day the bomb was planted. Information retrieved by the Detective is not enough (from the normative point of view) to reach the answer to the initial question, thus the second auxiliary question is asked. This time the answer might be reached. Solutions (B2) and (B3) are even more interesting. Form the normative point of view, the Detective is able to solve the initial problem after obtaining the answer to first auxiliary question asked. Why players decided to ask yet another auxiliary question in these cases is an open question. At this point we should mention one of the main drawbacks of our method of using the game to collect data as compared with the Erotetic Reasoning Test presented in Urbański et al. (2016). The Test is designed in such a way, that it requires an answer associated with its explanations. This allows for better understanding of certain choices made by subjects. QuestGen provides a flexible environment collecting data, but the cost is that gathered solutions are not enriched with additional explanations.

It is worth stressing that QuestGen players often reformulate premises (as we have mentioned, such a behaviour is observed for all stories in the

game). This is visible in the discussed schemata (B1), (B2) and (B3). Detectives sometimes ask about whether the bomb has been planted on odd day, and sometimes they decide to ask whether it was an even day (compare with the fifth premise "If the bomb *has not been planted* on an even day of the month, the culprit is Anthony"). We can even observe rather unexpected questions with negation as in (B3). Despite this variety of formulation of auxiliary questions by Detectives it may be observed that Informers (in vast majority of cases) are able to provide the correct answers with respect of their version of a given story. This may be interpreted in favour of the cooperative game design, where Detective and Informer play together against the game rules and the time limit.[11]

We also observed two solutions of the type presented in scheme (C). Here we may hypothesize that Detectives use a kind of heuristics addressing culprits mentioned in the premises. Observe that the order in which their names appear as questions is the same as the order in which they appear in premises (first Anthony, then Roger). This suggests a simple strategy of testing one option after another (we may hypothesize that it is done without a deeper analysis of available premises).
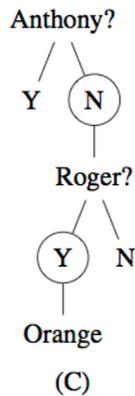
(C)

Let us now take a look at solutions presented in (D1) and (D2). In (D1) the Detective arrives to a wrong solution to the initial question. In his case auxiliary question required by the normative solution does not appear during the process. It seems that the player in this case is not able (or willing) to use given premises to decompose the initial question. In favour of this interpretation is that after obtaining clear information that it was not Anthony who planted the bomb, the player uses the premise related to Anthony as the culprit. What is more the second premise stating clearly that only purple or orange wires should be taken into account for disarming the bomb is ignored here – Detecitve points at pink wire as the one to be cut. This solution is far from the normative one, and may be a result of pure guessing (possibly enforced by the time constraints of the game). As such, solution (D2) is also interesting. The Detective asks here whether the bomb has been planted on an odd day. The information given by the Informer is correct, however (given the premises) it is certainly not enough to reach the solution to the initial question.

Anthony?

Y N

Planted on an odd day? Planted on an odd day?

Y N T N

Pink Orange

(D1) (D2)
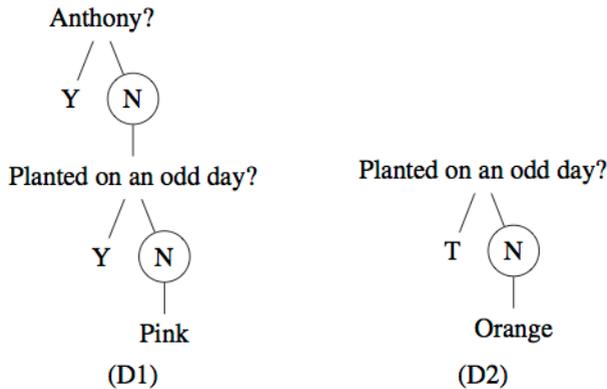
What is also interesting, there were cases when QuestGen design allowed to cope with certain more complex Detective moves. These are solutions depicted as (E1) and (E2).

(E1)   (E2)
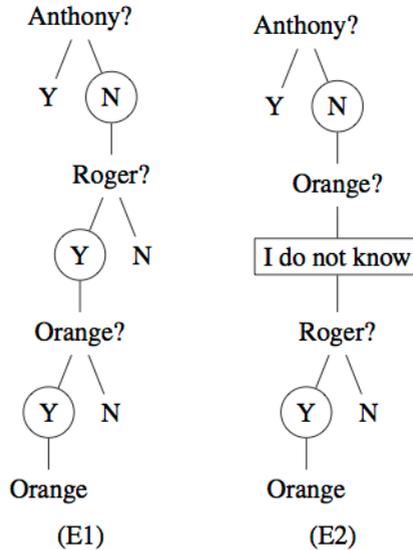
Let us remind that in QuestGen a direct question for a solution is for-bidden. Thus we allow the Informer to react with "I do not know" in sit-uations when the Detective asks for such an information – which is in line with the version of a story for the Detective. (The Informer may also use this response in a situation when she/he cannot resolve a question form Detective, i.e. there is not enough information in her/his version of the story). What we find puzzling with (E1) and (E2) is that they start with question about one of the culprits. It is not simply the case that the first question appearing is the forbidden one. What is more puzzling is that (E1) and (E2) differ with respect to Informer's reaction to a forbidden question. In (E1) we see "Yes" answer, while in (E2) we observe behav-iour with accordance to the game rules – i.e. "I do not know" response. A closer look on the way questions are formulated here sheds some light on these cases.

In (E1) we are dealing with well-formed questions. (Original spelling is preserved in following dialogues.)

DETECTIVE: Czy Arkadiusz ma coś wspólnego z bombą? [*Is it the case that Anthony has something to do with de bomb?*]

INFORMER: No.

DETECTIVE: A więc to Roman jest winny?! [*So it is the case that Roger is guilty?!*]

INFORMER: Yes.

DETECTIVE: **Czyżby** pomarańczowy? [*Orange,* **isn't it***?*]

INFORMER: Yes.

DETECTIVE: Pomarańczowy. [*Orange.*]

In our opinion the question about the orange wire should be interpreted here as a tag question in this context. Its formulation suggests that Detective already knows the answer. Informer seems to correctly interpret this move and thus does not use "I do not know" response, and simply confirms the answer given by Detective. In the case of (E2) we observe – more typical for QuestGen – extremely simply formulated questions.

DETECTIVE: arek? [*Anthony?*]

INFORMER: No.

DETECTIVE: fioletowy? [*purple?*]

INFORMER: I do not know.

DETECTIVE: roman?

INFORMER: Yes.

DETECTIVE: Pomarańczowy. [*Orange.*]

Here the Informer's response is well justified. For such a formulation of a question about the wire, there is no way (without actually hearing the question) to decide whether it is a proper question or whether the Detective knows the answer and just wants to make sure. What is interesting, after "I do not know" response, the Detective seems to use the heuristics observed in (C) and reaches the correct answer to the initial question.

4. Summary and discussion

In this paper we present and discuss data gathered with the use of the on-line game QuestGen. The data consists of solutions of detective-like stories, which are formulated accordingly to erotetic search scenarios. This allows us to compare normative point of view on these solutions with the solutions delivered by players. There are at least two conclusions from the presented analysis which point out the future research areas. First of all, we may conclude that QuestGen offers a convenient platform for gathering the valuable language data. Of course there is still a room for the improvements. As we have mentioned above, there are no additional explanations collected in QuestGen. This somehow restricts interpretation of certain solutions. However, we may reach for the results presented in Urbański et al. (2016). The *Bomb* puzzle in QuestGen is analogous to the Bomb task used in Erotetic Reasoning Test. The key difference is in structure of the task. In both cases we have detective-like story with initial problem and gathered evidence presented, but in the Erotetic Reasoning Test the task of a subject is to pick a question (one out of four listed below the story), each answer to which will lead to some solution to the initial problem. The subjects are also asked to justify their choices. On the basis of the analysis of these justifications Urbański et al. (2016) propose the notion of a weak erotetic implication in order to tackle the rationality behind using questions about the day of the month or about Anthony. These are still useful for the solution, however they do not meet requirements of e-implication (they are not cognitively useful). For the weak erotetic implication the second condition is restricted for some (not all – see Definition 6) direct answers to the implied question (cf. Urbański et al. 2016, 42). This illustrates how empirical perspective concerning normative models may enrich the formal tools used as a point of departure. As we may read in (Urbański et al. 2016, 45), "modelling the solutions by means of weak e-implication introduces an important descriptive factor into the formal framework of IEL". In our future research we plan to analyse the gathered solutions using this notion. What is especially interesting here are these solutions that are correct but cannot be counted as normative, when we think about regular e-implication. As for the second conclusion, the analysis performed for the needs of this paper show

how complicated and demanding is the task we are dealing with here. We are convinced that including the collected data into the Erotetic Reasoning Corpus project will simplify the future processing and analysis of the discussed data type.

Acknowledgments

References

COOPER, S., KHATIB, F., TREUILLE, A., BARBERO, J., LEE, J., BEENEN, M., LEAVER-FAY, A., BAKER, D., POPOVIĘC, Z. et al. (2010): Predicting Protein Structures with a Multiplayer Online Game. *Nature* 466, No. 7307, 756-760.

DARG, D. W., KAVIRAJ, S., LINTOTT, C. J., SCHAWINSKI, K., SARZI, M., BAMFORD, S., SILK, J., PROCTOR, R., ANDREESCU, D., MURRAY, P. et al. (2010): Galaxy Zoo: the Fraction of Merging Galaxies in the SDSS and Their Morphologies. *Monthly Notices of the Royal Astronomical Society* 401, No. 2, 1043-1056.

DZIEDZIC, D. (2016): Use of the Free to Play Model in Games with a Purpose: the RoboCorp Game Case Study. *Bio-Algorithms and Med-Systems* 12, No. 4, 187-197.

GENOT, E. J. (2009): The Game of Inquiry: the Interrogative Approach to Inquiry and Belief Revision Theory. *Synthese* 171, No. 2, 271-289.

GENOT, E. J. & JACOT, J. (2012): How Can Questions Be Informative before They Are Answered? Strategic Information in Interrogative Games. *Episteme* 9, No. 2, 189-204.

KLEKA, P. & ŁUPKOWSKI, P. (2014): Gamifying Science – the Issue of Data Validation. *Homo Ludens* 6, No. 1, 45-54.

KOMOSINSKI, M., KUPS, A., LESZCZYŃSKA-JASION, D. & URBAŃSKI, M. (2014): Identifying Efficient Abductive Hypotheses Using Multicriteria Dominance Relation. *ACM Trans. Comput. Logic* 15, No. 4, 28:1-28:20.

LESZCZYŃSKA, D. (2004): Socratic Proofs for Some Normal Modal Propositional Logics. *Logique et Analyse* 47, No. 185-188, 259-285.

LESZCZYŃSKA, D. (2007): *The Method of Socratic Proofs for Normal Modal Propositional Logics*. Adam Mickiewicz University Press.

Leszczyńska-Jasion, D. (2013): Erotetic Search Scenarios as Families of Sequences and Erotetic Search Scenarios as Trees: Two Different, yet Equal Accounts. *Research report* no 1(1)/2013. Technical report, Adam Mickiewicz University, Poznań. Available at: https://intquestpro.wordpress.com/resources/reports/.

Łupkowski, P. (2010): Cooperative Answering and Inferential Erotetic Logic. In: Łupkowski, P. & Purver, M. (eds.): *Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue*. Poznań: Polish Society for Cognitive Science, 75-82.

Łupkowski, P. (2011a): A Formal Approach to Exploring the Interrogator's Perspective in the Turing Test. *Logic and Logical Philosophy* 20 No. 1/2., 139-158.

Łupkowski, P. (2011b): Human Computation – How People Solve Difficult AI Problems (Having Fun Doing It). *Homo Ludens* 3, No. 1, 81-94.

Łupkowski, P. (2016): *Logic of Questions in the Wild. Inferential Erotetic Logic in Information Seeking Dialogue Modelling*. London: College Publications.

Łupkowski, P. & Dziedzic, D. (2016): Building Players' Engagement – a Case Study of Games with a Purpose in Science. *Homo Ludens* 9, No. 1, 127-145.

Łupkowski, P., Ignaszak, O. & Wietrzycka, P. (2015): Modelowanie interrogacyjnego rozwiązywania problemów w środowisku gry QuestGen. *Studia metodologiczne* 34, 239-254.

Łupkowski, P., Urbański, M., Wiśniewski, A., Paluszkiewicz, K., Ignaszak, O., Żyluk, N., Urbańska, J., Gajda, A., Marciniak, B., Bładek, W., Juska, A., Kostrzewa, A. & Pankow, D. (2017): *Erotetic Reasoning Corpus Architecture: Components, Tags, Annotation*. Technical report, Poznań: Adam Mickiewicz University, available at: https://ercorpus.wordpress.com/home/.

Łupkowski, P. & Wietrzycka, P. (2015): Gamification for Question Processing Research – the Questgen Game. *Homo Ludens* 7, No. 1, 161-171.

Łupkowski, P. & Wiśniewski, A. (2011): Turing Interrogative Games. *Minds and Machines* 21, No. 3, 435-448.

Moradlou, S. & Ginzburg, J. (2014): Learning to Understand Questions. In: Muller, P. & Rieser, V. (eds.): *Proceedings of SemDial 2014 (DialWatt)*. Edinburgh, 116-124.

Rosé, C. P., DiEugenio, B. & Moore, J. (1999): A Dialogue-based Tutoring System for Basic Electricity and Electronics. In: Lajoie, S. P. & Vivet, M. (eds.): *Artificial intelligence in education*. Amsterdam: IOS, 759-761.

Sedlár, I. & Šefránek, J. (2014): Logic and Cognitive Science. In: Kvasnička, V., Pospíchal, J., Návrat, P., Chalupa, D. & Clementis, L. (eds.): *Artificial Intelligence and Cognitive Science IV*. Bratislava: Slovak University of Technology Press, 219-236.

Shoesmith, D. J. & Smiley, T. J. (1978): *Multiple-conclusion Logic*. Cambridge: Cambridge University Press.

STENNING, K. & VAN LAMBALGEN, M. (2008): *Human Reasoning and Cognitive Science*. MIT Press.

URBAŃSKI, M. (2001a): Remarks on Synthetic Tableaux for Classical Propositional Calculus. *Bulletin of the Section of Logic* 30, No. 4, 194-204.

URBAŃSKI, M. (2001b): Synthetic Tableaux and Erotetic Search Scenarios: Extension and Extraction. *Logique et Analyse* 44, No. 173-175, 69-91.

URBAŃSKI, M. (2002): Synthetic Tableaux for Łukasiewicz's Calculus Ł3. *Logique et Analyse* 45, No. 177-178, 155-173.

URBAŃSKI, M. (2011): Logic and Cognition: Two Faces of Psychologism. *Logic and Logical Philosophy* 20, No. 1-2, 175-185.

URBAŃSKI, M. & ŁUPKOWSKI, P. (2010): Erotetic Search Scenarios: Revealing Interrogator's Hidden Agenda. In: Łupkowski, P. & Purver, M. (eds.): *Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue*. Poznań: Polish Society for Cognitive Science, 67-74.

URBAŃSKI, M., PALUSZKIEWICZ, K. & URBAŃSKA, J. (2016): Erotetic Problem Solving: From Real Data to Formal Models. An Analysis of Solutions to Erotetic Reasoning Test Task. In: Paglieri, F. (ed.): *The Psychology of Argument: Cognitive Approaches to Argumentation and Persuasion*. London: College Publications.

URBAŃSKI, M., PALUSZKIEWICZ, K. & URBAŃSKA, J. (2014): Deductive Reasoning and Learning: A Cross Curricular Study. Research report no 2(4)2014. Technical report, Poznań: Adam Mickiewicz University, available at: https://intquestpro.wordpress.com/resources/reports/.

VENHUIZEN, N., BASILE, V., EVANG, K. & BOS, J. (2013): Gamification for Word Sense Labeling. In: *Proceedings of the 10th International Conference on Computational Semantics* (IWCS-2013). Postdam, 397-403.

VON AHN, L. (2006): Games with a Purpose. *Computer* 39, No. 6, 92-94.

WIŚNIEWSKI, A. (1995): *The Posing of Questions: Logical Foundations of Erotetic Inferences*. Boston – London: Kluwer AP, Dordrecht.

WIŚNIEWSKI, A. (1996): The Logic of Questions as a Theory of Erotetic Arguments. *Synthese* 109, No. 1, 1-25.

WIŚNIEWSKI, A. (2001): Questions and Inferences. *Logique et Analyse* 44, No. 173-175, 5-43.

WIŚNIEWSKI, A. (2003): Erotetic Search Scenarios. *Synthese* 134, No. 3, 389-427.

WIŚNIEWSKI, A. (2004a): Erotetic Search Scenarios, Problem-Solving, and Deduction. *Logique et Analyse* 47, No. 185-188, 139-166.

WIŚNIEWSKI, A. (2004b): Socratic Proofs. *Journal of Philosophical Logic* 33, No. 3, 299-326.

WIŚNIEWSKI, A. (2012): Answering by Means of Questions in View of Inferential Erotetic Logic. In: Meheus, J., Weber, E. & Wouters, D. (eds.): *Logic, Reasoning and Rationality*. Springer.

WIŚNIEWSKI, A. (2013a): *Essays in Logical Philosophy*. Berlin – Münster – Wien –
   Zürich – London: LIT Verlag.
WIŚNIEWSKI, A. (2013b): *Questions, Inferences, and Scenarios*. London: College Pub-
   lications.
WIŚNIEWSKI, A. & SHANGIN, V. (2006): Socratic Proofs for Quantifiers. *Journal of
   Philosophical Logic* 35, No. 2, 147-178.

Jan-Werner Müller: *What Is Populism?*
University of Pennsylvania Press, Philadelphia, 2016, 123 pages[1]

Populism is democracy's evil twin brother. It presents itself as democracy, as making good on democracy's highest ideals, "Let the people rule!" (p. 6). But behind the mask it is demagoguery (p. 11), a degraded form of democracy (p. 6). The author pits representative democracy, with its mass suffrage, against populism, and the latter depends conceptually on the former: "Populism arises with the introduction of representative democracy; it is its shadow" (p. 20). A claim attributed to the Nazi ideologue Carl Schmitt and his Fascist counterpart Giovanni Gentile (p. 28) stayed with me throughout the book; the claim, namely, that Fascism is more democratic than democracy is capable of, or that only Fascism can realize the full potential of democracy. Whether Fascist or just populist, this claim makes a mockery of democracy.

The author is a political scientist of German origin who works in the United States. Müller casts his net wide, drawing extensively on recent examples from Europe, North and South America, with Asia and Africa taking more of a backseat. He presents his book as a conceptual analysis of populism, which qualifies it as a treatise on political philosophy, but he also offers hands-on advice on how to counter populist politicians and voters in the day-to-day political fray. His advice is congruent with his analysis and thus elucidates the latter from an applied angle. The advice is helpful, since populist politicians and their supporters play by other rules than their competitors in the politicial marketplace. The author is openly in favour of democracy and openly against populism, but what I found appealing about the book is that it resists the temptation to speak condescendingly of populist voters, picturing them as unlettered, unwashed masses stoked up on hatred, disdain and aggression directed at everyone who is 'them' and not 'us'. It would be contrary to the democratic mindset not to engage with them as equals, without necessarily buying into their framing, and it would also be counterproductive, because it plays

---

[1]   ✉ Bjørn Jespersen
       Department of Computer Science, VSB-Technical University of Ostrava
       17. listopadu 15/2172, 708 33 Ostrava – Poruba, Czech Republic
       e-mail: bjornjespersen@gmail.com

right into the populist narrative of being outsiders whose rightful place is in society's mainstream.[2]

Though billed as an analysis of populism, I came away with the impression that the book is at heart an examination of democracy. The author's methodology, accordingly, is to study the healthy body of representative democracy by studying one of its typical maladies, namely populism.[3] The diagnosis is that the 'shadow' of democracy is a fully functional method of government, which when fully implemented is not democracy in full bloom, but straight-up dictatorship. Not surprisingly, the final of the 'Seven Theses on Populism' that form the Conclusion of the book, states, "Populism, then, should force defenders of liberal democracy to think harder about what current failures of representation might be" (p. 103).[4]

<p style="text-align:center">*   *   *</p>

---

[2]   Müller tracks some of the different meanings that the terms 'populism' and 'populist' have had since the 19th century, applying to different political situations and landscapes. Müller's book is specifically about the most recent trends, and his use of the terms is tailored to those. Portions of the book are devoted to critiqueing rival interpretations of populism and strategies for addressing populists. I am leaving those portions out of consideration here. I am also disregarding attempts to explain why populism has become such a political force. Briefly, though, Frum (2017) makes this interesting, if somewhat sweeping claim: "Outside the Islamic world, the 21st century is not an era of ideology. The grand utopian visions of the 19th century have passed out of fashion. The nightmare totalitarian projects of the 20th have been overthrown or have disintegrated, leaving behind only outdated remnants: North Korea, Cuba. What is spreading today is repressive kleptocracy, led by rulers motivated by greed rather than by the deranged idealism of Hitler or Stalin or Mao. Such rulers rely less on terror and more on rule-twisting, the manipulation of information, and the co-optation of elites." Populist machinations would, accordingly, facilitate the power grab of 'repressive kleptocrats' and their clients. But I am not entirely convinced, nor would Müller be, as he emphasizes that the mass appeal of populism is driven by identity politics rather than by primarily economic concerns.

[3]   Rosanvallon (2011) says something along the same lines: "Si nous voulons mieux comprendre la démocratie, il nous faut donc aussi mieux saisir ce qu'est le populisme. Car l'intelligence de la démocratie est inséparable d'une intelligence de ses perversions."

[4]   See, e.g., p. 60: "… the defect that weaker socioeconomic groups do not participate in the political process and do not have their interests represented effectively".

The author does not spell out in any detail what he understands by democracy, but he lists (p. 55) the following political rights as being among those constitutive of democracy:

- freedom of speech and assembly
- media pluralism
- the protection of minorities

together with this requirement (p. 55):

- citizens must be well-informed about politics

Moreover, he ascribes (p. 40) the following tenets to democratically-minded politicians:

- representation is temporary and fallible
- contrary opinions are legitimate
- society cannot be represented without remainder
- it is impossible for one party or politicians permanently to represent an authentic people apart from democratic procedures and forms

What permeates these tenets is a fallibilist, bottom-up, empirical approach to politics and of how to structure society. Populism turns this upside-down. Müller describes populism in terms of the following package of tenets:

- polarization between the people versus the elite(s)
- monolithic culture, antipluralism, delegitimization of opponents
- populist governments characterized by open displays of:
    - suspension of the separation of powers
    - state jobs being handed out to loyalists
    - corruption, cronyism
    - mass clientelism
    - discriminatory legalism[5]
    - suppression of civil society
    - proneness to conspiracy theories.

---

[5] "For my friends, everything; for my enemies, the law" (p. 46). For instance, a populist government can decide to single out particular individuals, companies and organizations for nitpicking tax audits while leaving the rest alone.

Rosanvallon (2011), which Müller references (p. 109, fn. 43), also lists the people/elites dichotomy as the first of three 'simplifications' that characterize populism. The second simplification is a variation on the theme of suspension of the sepation of power, namely the elimination of 'intermediaries', such as nonelected judges, between the people and the ruler(s). The third simplification is that social cohesion becomes a matter exclusively of cultural identity, rather than the social rapport between individuals and groups of individuals that is found in day-to-day civic society. This, again, is symptomatic of the monolithic vision of cultural identity and the adjacent lack of tolerance. It is important to note, as Müller does, that his description of populism is of a form rather than of a content. For instance, it is not necessary that xenophobia must be part of any populist agenda. Xenophobia, say, will be part of a given agenda only if a particular populist definition of a particular people states or entails that foreigners have no place in the country because they are foreigners.

A noteworthy feature of the list of tenets above is that populism will intersect with totalitarianism, while it is possible to have one without the other. A totalitarian regime that imposes itself from above even without pretending to represent the people will not be populist, and a populist regime that, for instance, does not persecute dissidents (though it will harass them) will not qualify as totalitarian. However, there is not much conceptual wiggle room, so for this reason I would have appreciated a closer comparison between totalitarianism and populism. What the author does offer is a remark like this (p. 93):

> One implication of the analysis presented in this book is that National Socialism and Italian Fascism need to be understood as populist movements – even though, I hasten to add, they were not just populist movements but also exhibited traits that are not inevitable elements of populism as such: racism, a glorification of violence, and a radical 'leadership principle'.

In other words, Nazism and Fascism were (are) populism-plus.[6] But it seems to me that Nazism and Fascism (including the avowedly Fascist regimes that held sway in southern Europe well into the 1970s) did (do) little more than put the icing

---

[6] I am not sure Italian Fascism was distinctly racist. Once the movement had shed its progressive ambitions and become reactionary, Fascism strikes me as run-of-the-mill imperialism. To be sure, Fascism definitely had a strong fondness for violence, dominance, cult of personality, hierarchy, and militarism. (I suppose that 'leadership principle' is a nod to the *Führerprinzip* practiced in the Third Reich.)

on the populist cake by concentrating all institutional power in one point while soaking the oxygen out of civil society. It can be no accident that a string of populist government or party leaders have become household names, unlike the leaders of perhaps most democratic countries.

*   *   *

The populist makes a number of assumptions that strike the philosophical mind as being exceptionally strong. First and foremost, it is assumed that there is such a thing as *the people*, e.g. the Turkish people, the Korean people, the Danish people, above and beyond the individuals of flesh and blood who at any given time self-identify as Turks, Koreans, Danes, etc. Next, it is assumed that the people has a clearly circumscribed set of properties defining its culture. Third, it is assumed that it can be known what these properties are. Populism is infallibilist, because it takes for granted that it knows what the people is like and what it wants, that there is exactly one view of how to run the country that is true and legitimate, and that that view is that of the populist. This point of departure explains why the populist thinks that any alternative view is a deviation from the truth, and also why the populist is adamant about eliminating alternative news sources and alternative communities within civil society. Instead of the citizens being exposed, or at least having access, to a wealth of different sources and views, they are locked inside an echo chamber with a clear framing of questions and answers – "and, amazingly, it always happens to be the [answers] we were expecting" (p. 36).[7]

Starting from these assumptions, the first move the populist makes is to instill a dichotomy between *the people* (*populus*, hence the term) and *the elite*. It bears repeating that the people thus understood is not the totality of a number of individuals sharing the same passport or living on the same territory or self-identifying as belonging to the people in question at a given moment. This conception of the people Müller calls the empirical one. The mythical or moral conception, as he sometimes calls it, is to deem only some, if most, of the people part of the people. They, and only they, are the real people. The term 'populism' stands for an exclusionary, rather than inclusionary, conception of what it means to belong to a given people. The populist, self-declared or not, speaks on behalf of the 'morally real people',

---

[7]   1930s Germany offers the clear-cut example of *Lügenpresse* (newspapers not toeing, or even defying, the Nazi line being called a lying press) and *Gleichschaltung* (the 'equalization' of all mass media, meaning that they would all stick to the official line).

excluding the 'illegitimate' ones that, though empirically belonging to the people (by possessing Turkish or Danish citizenship, say), fail to belong to the people proper. Thus, a member of the Turkish elite, say, does not belong to the Turkish people.[8] The populist decides what is required to belong to the people, making the notion a thoroughly politicized one, which is, furthermore, plastic enough so as to be remoulded to fit an altered political landscape despite its pretence to be atemporal.

The polar opposite of the people (the various elites) remains a fluffy notion, and I would have liked to hear a bit more about how populists picture these elites. There is a sketch (p. 57) of elites standing for "economic liberalism, a pluralistic and tolerant 'open society', and the protection of fundamental rights", but the reader is expected to have an intuitive grasp of this key notion. Müller ought perhaps to have spelt out in more detail the fact that the populist is not opposed to elites per se. First, the populists when politically successful become (part of) the elite themselves (though they would not label themselves as such, of course). Second, even when in opposition, the populists will look kindly upon those portions of the existing elites that serve the populist cause. But then, one wonders, do those portions of the elite also belong to the people as well? It might be tempting to say they do, but then the supposedly sharp dichotomy between people and elite becomes conceptually muddled. Third, Müller notes that the supporters of those populist politicians who already belong to the established elite know that they do: "what matters [to the supporters is the populists'] promise that as a proper elite, they will not betray the people's trust and will in fact faithfully execute the people's unambiguously articulated political agenda" (p. 30). The contrast between 'proper elite' and (I presume) 'improper elite' is intuitively clear, but since *elite* is a key notion this contrast requires more elucidation.

Müller contrasts populism with democracy, while also bringing up technocratic government briefly, but he might have wanted to elaborate on a hint he drops to the effect that the technocrat and the populist are both monolithic, unlike the democrat, who is pluralistic (cf. p. 97). The technocrat believes that there is exactly one correct, rational policy, while the populist believes that there is exactly one

---

8   This exclusionary move reminds me of Stalin's term 'rootless cosmopolitan', which was code for Jewish intellectuals in the Soviet Union, but which feeds on the idea that some citizens do not belong to the people and, by implication, have no place amongst them, and, by implication, should be done away with.

correct, authentic policy, and in both cases there is no need for democratic discourse and inclusion of alternatives.[9]

<div align="center">*   *   *</div>

Polarization is the hallmark of the populist modus operandi. There is no attempt to unify conflicting views, seek compromise or strike a balance. The critical platform is provided by the monolithic notion of the people as 'the silent majority' that has, finally, acquired a voice. It is obvious enough how polarization works as part of a campaign strategy, but Müller emphasizes that the populist continues to polarize also when in power. This suggests to me that populism, as Müller understands it, is conceptually incapable of full implementation of its alleged program of the people being the sole political and cultural force of the country; for there is always going to be an elitary residue. This is one place where the conspiracy theories come in handy. Even when real opposition is feeble, there are mistakes that need to be explained away, and 'foreign agents', a 'fifth colonne' of unreformed members of past elites, etc., conspiring against the people and its leadership by causing those mistakes are easy scapegoats. Just as a hammer needs nails to pound, a populist needs enemies to hammer away on.

An interesting observation Müller makes is that populists in power can quite openly get away with, not least, abolishing the separation of power, not only the separation between the legislative, the executive, and the judiciary branches, but also the additional checks and balances provided by, inter alia, the press, the military, the police, and the intelligence agencies. They can do so in perfect keeping with the official programme of serving the people, because the people, as understood by the populist, wants all power to be centralized and put at their service.

The people is a unit that appears to play much the same role as, say, a deity when making a political argument: the people/the deity wants this or does not want that, and its wishes and demands surpass the law of the land. Or, the presumed will of the people/the deity is turned into official law. Whether the populist regime happens to be theocratic or secular, what happens then is that "the [new] constitution sets a number of highly specific policy preferences in stone, when debate about

---

9   The interview found at *https://www.youtube.com/watch?v=9SH5Jy8xxsY* elaborates on this point, among many other. I seem to remember a recent Danish government that was fond of speaking of 'the policy of necessity' ('nødvendighedens politik').

such preferences would have been the stuff of day-to-day political struggle in non-populist democracies" (p. 65). Those who find that the appeal to the people or a deity has no purchase on political discourse will be dismissed as not belonging to the people and, therefore, having no legitimacy.

The picture that begins to emerge of populism is that it has a hollow core "because their claim [to represent the people] is of a moral and symbolic – not an empirical – nature, it cannot be disproven" (p. 39).[10] The populist modus operandi also looks like a Macchiavellian strategy for seizing and retaining power in the sense that it maintains a façade of being a benign policy (one serving almost everybody) while in fact it is much more self-serving (one serving a few select individuals and very specific segments of society). The masses, to put it bluntly, are being played (again). Müller (p. 49) speaks of a 'final great irony' of populism:

> Populism in power brings about, reinforces, or offers another variety of the very exclusion and the usurpation of the state it most opposes in the reigning establishment it seeks to replace. What the 'old establishment' or 'corrupt, immoral elites' supposedly have always done, the populists will also end up doing – only, one would have thought, without guilt and with a supposedly democratic justification.

Populism presupposes democracy conceptually but not chronologically. A country that has not known democracy can be a breeding ground for populism, provided its citizens value the idea(l) of democracy.

<p align="center">*   *   *</p>

The author is keen to give populism a fair hearing. His discussion of, e.g., East Germany and 19th-century USA leads me to think that one can actually make a favourable case for populism, provided the country in question is far from being a democracy. In East Germany people (not the people, but people, lots of people) would take to the streets in 1989 and chant "*Wir* sind das Volk", meaning that the Party was wrong to claim that it represented the people and its interests. And the USA was founded on the idea that the country should not be governed by clergy-

---

[10] As Eco (1995) says about Mussolini, "Mussolini did not have any philosophy: he had only rhetoric."

men, monarchs or aristocrats, as in Europe, but by the people in the form of representative democracy. Hence the disappointment when a new elite class of capitalists came into being that had little time for the vast majority. However, making a favourable case for populism in modern-day democracies is hard. The most Müller is prepared to concede can, I guess, be summarized like this: populists highlight real problems that most others in power would rather not address, so that speaks in their favour, but they prescribe a cure that will kill the patient.

One last thing. Müller distinguishes between right-wing and left-wing populism, without making too much of the distinction. While it is a distinction with a difference, the two 'wings' seem to me to be very close indeed, rhetoric aside. Perhaps the right/left distinction, which has been with us since the French Revolution, is not always clear-cut enough to be of much use. Perhaps a more profound distinction nowadays would be between populist authoritarianism and democratic pluralism with thorough separation of powers. Just a thought.

$$* \quad * \quad *$$

This 100-page book is an easy read, lively and brisk-paced, occasionally even funny, for instance, when quoting politicians from the Polish PiS party. But also unfailingly scholarly, in part thanks to its extensive Notes. And rich in content, as one discovers when pausing to mull over a side remark or a tangential line of thought. The prose is refreshingly free of politological clichées, though a 'neoliberal hegemony' does crop up on p. 98, and the author is eager to reach out to his mixed readership. (I do miss an index of subjects, though.) This engaging monograph is a fine example of applied political philosophy. It is also exceptionally timely.

Bjørn Jespersen

Acknowledgments

References

ECO, U. (1995): *Ur-Fascism*. Available at: http://www.nybooks.com/articles/1995/06/22/ur-fascism/.

FRUM, D. (2017): How to Build an Autocracy. *The Atlantic*, March 2017 issue, available at: https://www.theatlantic.com/magazine/archive/2017/03/how-to-build-an-autocracy/513872/.

ROSANVALLON, P. (2011): *Penser le populisme*. Available at:
http://www.laviedesidees.fr/Penser-le-populisme.html.

Olivier Roy, Allard Tamminga and Malte Willer (eds.):
Deontic Logic and Normative Systems: 13th International Conference,
DEON 2016, Bayreuth, Germany, July 18-21, 2016
College Publications, Milton Keynes, 2016, 302 pages[1]

Deontic logic enjoys increasing popularity. First and foremost, there are the biennial DEON conferences dedicated to deontic logic and related topics (since 1991). Moreover, *Handbook of Deontic Logic and Normative Systems* came into existence in 2013. The crucial importance of this publication for further rise of deontic logic is beyond question. But let us move three years forward. The 13th DEON conference took place in Bayreuth (Germany) on July 18-21, 2016. The reviewed book contains the proceedings of this conference. Interestingly enough, the special focus was "Reasons, Argumentation and Justification". The clever choice of special focus has led to an interesting cooperation between argumentation theory and deontic logic. The conference had four keynote speakers, namely John Broome, Janice Dowell, Xavier Parent, and Gabriella Pigozzi.

The book contains eighteen interesting and original papers that are usually structured as follows: first, the authors introduce their topic, provide us with some background and some motivations for developing a new logical system, or a couple of them. Second, syntax, semantics and some inferential machinery are introduced. The effectivity and the problem-solving potential of the systems are usually demonstrated in passing. Next, the formal properties of the systems are proved, or at least mentioned. Finally, the papers conclude the achieved results, providing us

[1]   ✉ Daniela Glavaničová
         Department of Logic and Methodology of Sciences, Faculty of Arts
         Comenius University in Bratislava
         Gondova 2, 814 99 Bratislava, Slovakia
         e-mail: dada.baudelaire@gmail.com

also with some directions for further research. This is how the nicely structured face of modern logic looks!

The first paper *Cumulative Aggregation* comes from Ambrossio, Parent and van der Torre. The paper employs the framework of *Abstract Normative Systems*. It is concerned with conditional obligations, more specifically, with two principles of aggregation (simple and cumulative). Two systems are introduced, the system FA for simple aggregation, and the system FC for cumulative aggregation. As the authors acknowledge, their contribution is mostly technical. From the philosophical point of view the paper lacks extensive discussion of the relevance of these results to normative reasoning. There is a typo at the end of p. 4. The authors write that "Let FA = {FD, AND}". But FA is a certain triple, whilst {FD, AND} is a set of rules, so there should be "R" instead of "FA" (R is a set of rules).

Anglberger, Faroldi and Korbmacher are the authors of the second paper, *An Exact Truthmaker Semantics for Permission and Obligation*. The paper proposes semantics for permissions and obligations. This account is hyperintensional, so deontic operators are not closed under logical equivalence. Hyperintensionality tends to be an efficient weapon against paradoxes, and deontic logic is no exception. The authors ascribe truth-values to obligations and permissions (p. 23), though intuitively these are not truth-apt. It would be better to speak about the truth-values of deontic propositions and about the satisfaction (fulfilment, validity) of obligations. As regards hyperintensionality, though the author of this review herself holds a hyperintensional stance on deontic logic, there are some dangers that come with this feature. For instance, the formula $P(q \lor \neg q)$ does not hold in the proposed system (see p. 24), but the formula $P(p \land q) \to P(q \land p)$ is valid (since $p \land q$ has the same truthmakers as $q \land p$, as is clear from p. 29). Regarding the former, the authors list the intuitive counterexample: it may not be permitted to kill the cat or not kill the cat (p. 24). However, if "or" stands for disjunction, the formula $P(q \lor \neg q)$ does not allow for more than a tautology. But if we are allowed to list such sentences as counterexamples, one can list a similar sentence against the formula $P(p \land q) \to P(q \land p)$. For instance, it may be permitted to close the door and open the window, but not permitted to open the window and close the door (since the order matters). Yet of course, if the order matters, we are no longer dealing with conjunction – but why should the former sentence count as a good counterexample, but the latter as a bad one? Finally, there is a missing "O" at the p. 30, derivation (ii), line (e).

The third comes the paper *A Structured Argumentation Framework for Detaching Conditional Obligations* written by Beirlaen and Straßer. As the first

paper, this paper is too devoted to conditional obligations. The paper starts with abstract argumentation framework, subsequently instantiating it with deontic arguments, thus generating a structured deontic argumentation framework. The authors claim that obligations that are violated should not be detached (p. 42). Yet intuitively, the fact that an obligation is violated does not imply that the obligation no longer holds.

The fourth paper *Argumentation Frameworks with Justified Attacks* is written by Dyrkolbotn and Pedersen. As in the previous paper, argumentation frameworks are used here. However, and interestingly enough, argumentation frameworks are used to analyse the argumentation itself. A system for metalogical reasoning is thus developed. A well-known troublesome example from default logic is analysed within the proposed system, aiming to resolve a meta-level disagreement pertaining to examples of this sort.

The fifth is the paper *Arguments, Responsibilities and Moral Dilemmas in Abductive Default Logic* written by Dyrkolbotn, Pedersen and Broersen. This paper employs two frameworks, default logic and again, argumentation frameworks. The paper presupposes that agents are responsible for something only when they have had a choice. The agent should not be blamed for something that is a designer's fault. However, there seem to be two weak points. First, the epistemic aspect is neglected. It is stated that "we do not assume that the agent knows (or does not know) the (implicit) consequences of applying certain rules" (p. 66). However, in reality, agents have some epistemic capacities. For instance, the epistemic ability of an artificial intelligence is given in advance: we know what the agent knows, and to what extent it can carry out reasoning (though it is not at all trivial to speak about the responsibility of a machine). On the other hand, the epistemic competence of human agent is not given in advance. Despite that, we can presuppose something like Jago's *bounded rationality*: an agent is neither a deductive machine, nor incapable of trivial inferences (see Jago 2014a; 2014b). The second weak point is the very presupposition that agents are never responsible when the element of choice is missing. Consider the following example: "imagine a young woman wondering whether to enlist in the army. (…) [I]f she chooses to enlist and then decides to kill someone – intentionally – on the orders of a superior officer, we would hesitate to say she is morally responsible" (p. 73). But imagine two similar scenarios: (1) a wealthy agent A, a racist, as well as a psychopath, freely decides to join army with an intention of killing people of other races and (2) a poor agent B, an unemployed father of three children and a pacifist decides to join army with an intention

of earning money for his family. Intuitively, we would say the agent A is responsible for the subsequent killing (no obligation forced A to join the army) but the agent B is not responsible for the killing (B was forced to join the army by obligation of earning money for his family).

The sixth paper *Basic Action Deontic Logic* is written by Giordani and Canavotto. This paper develops a system of dynamic deontic action logic that consists of ontic part (logic of states and actions) together with deontic part (abstract and actual deontic ideal).

Governatori, Olivieri, Calardo and Rotolo wrote the seventh paper, *Sequence Semantics for Norms and Obligations*. This paper proposes semantics for sequences of (compensatory) obligations and for (ordered) sequences of permissions. The suggested sequence semantics is an extension of neighbourhood semantics. Thanks to this, the authors provide us with a nice adaptation of a standard completeness proof for neighbourhood semantics. There should be $\langle \|a_1\|_V, \dots, \|a_n\|_V \rangle$ and $\langle \|a_1\|_V, \dots, \|a_{n-1}\|_V \rangle$ instead of $\langle \|a\|_1, \dots, \|a_n\| \rangle$ and $\langle \|a\|_1, \dots, \|a_{n-1}\| \rangle$ in the proof of the theorem 5.9 (the completeness of the system $D^{\otimes}$), p. 104; and there are seven, not six detachment schemata (p. 105).

Ju and van Eijck are the authors of the eighth paper entitled *To Do Something Else*. The paper proposes two dynamic action logics, stemming from an idea that normative notions can be defined in terms of consequences of actions carried out (note that this background idea may be criticised by the proponents of deontological ethics). The initial system of dynamic deontic logic provided by Meyer led to incorrect reading of refraining from doing something. The present paper offers a new reading of refraining in terms of *doing something else*.

Multivalued Logics for Conflicting Norms, the ninth paper, is written by Kulicki and Trypuz. The authors develop three systems of multi-valued deontic action logics, whilst the main focus is on normative conflicts and on merging norms. The background idea is that one can compute deontic values of actions just as one can compute truth values in propositional logic. The first proposed system offers a pessimistic view on normative conflicts, the second system an optimistic view, and the third a neutral view. The authors seem to be sympathetic to the second and the third system that have the optimistic flavour and liberate the agent from the burden of guilt (p. 133). To motivate the optimistic view, they claim that in cases of normative conflicts, if we follow one obligation, it is enough to make the decision good (p. 132). Yet one may object that following one obligation is not enough if we have more obligations. For instance, if one has two kids in the kindergarten, it is not enough to pick just one of them up. Finally, the axiom (30)

$N^b(\alpha \sqcap \beta) \to N^b(\alpha) \vee N^b(\beta) \vee (O(\alpha) \wedge F(\beta))$ (p. 134) should contain one more disjunct: $F(\alpha) \wedge O(\beta)$. The system will not be sound otherwise, since this axiom is not a tautology: if the value of α is f and the value of β is o, then the value of $\alpha \sqcap \beta$ is $\top$, so the value of the antecedent formula $N^b(\alpha \sqcap \beta)$ is 1, but neither of the three disjuncts holds in this case, so the value of the consequent will be 0, and the resulting value will be 0 too.

Liao, Oren, van der Torre and Villata are the authors of the tenth paper, *Prioritized Norms and Defaults in Formal Argumentation*. The paper introduces a prioritized abstract normative system and analyses three different approaches to nonmonotonic reasoning in terms of it. It is claimed that "If priorities are disregarded, then this logic program has two answer sets: {a, p, x} and {a, p, ¬x}. Thus, considering priorities, the former is the unique preferred answered set, as pointed out in Example 2.6" (p. 144), but the authors obviously meant the latter, not the former set.

The eleventh position in the book belongs to the paper *Reasons to Believe in a Social Environment* written by Liu and Lorini. The paper devises a new system of Dynamic Epistemic Logic of Evidence Sources, DEL-ES. There should be "justification logic" instead of "justication logic" (p. 156 and p. 169). And the approach in this paper is quantitative, not qualitative, contrary to what the authors suggest at the page 157.

Marra is the author of the twelfth paper, *Objective Oughts and a Puzzle about Futurity*. The paper is concerned with future-dependent objective oughts. The author attempts to briefly defend the usefulness of the objective oughts, though not persuasively enough. What is right or best seems to be relative at least to some package of norms, or values, and there may be considerable differences between the best action for one agent and the best action for another agent. Moreover, the article tries to avoid determinism, but comes with the commitment to indeterminism. Both of them are strong metaphysical commitments – they should be either avoided or it should be argued for the chosen option.

The thirteenth position in the book is occupied by the *Rights in Default Logic* written by Mullins. The author argues that default logic is an appropriate framework for reasoning about rights and consequently develops his account in terms of Horty's default logic. The example author uses as an illustration of the role of strong and weak permission (p. 193) does not seem apt. Strong permissions are understood as positive (explicitly given) and weak permissions as negative (no norm requires us to act in a certain way). The example in question concerns the prohibition of insulting speech overridden by the right to freedom of

political communication. However, what's going on in this example is some priority ordering of norms, not the distinction between strong and weak permissions.

Pavese is the author of the fourteenth paper, *Logical Inference and Its Dynamics*. The author provides us with an argument from dynamic conception of inference to a dynamic conception of inference rules, which motivates her subsequent proposal. However, the account seems to be vulnerable to the paradox of inference. For instance, the author holds that a context supports some sentence just in case the result of updating context with this sentence is the context itself (p. 206) and that sentence with "therefore" is informationally empty (p. 208). Intuitively, an inference brings some new (analytic) information (see Duží 2008 for this line of thought).

Peterson and Kulicki wrote the fifteenth article *Conditional Normative Reasoning with Substructural Logics*. The starting point is Peterson's system CNR (Conditional Normative Reasoning) that aims to be paradox-free deontic logic but does not have De Morgan validities and the Law of Excluded Middle. Because of this, the authors propose an "intermediate" logic that is stronger than CNR but still avoids undesirable paradoxes.

Silk is the author of the sixteenth paper, *Update Semantics for Weak Necessity Modals*. The paper is concerned with formal analysis of weak and strong necessity modals (should and must), mostly in their deontic reading. The author claims that the weakness of should consists in a failure to presuppose that the relevant worlds in which the prejacent is necessary are candidates for actuality. The author analyses the sentence "Alice must be generous" (p. 244), but it is not clear what is his view on ambiguity of this (and similar) sentences. Obviously, the sentence has deontic, as well as epistemic reading. The author claims that according to the proposed semantics, when we say something like the above sentence, we don't update information, just "place a necessity claim on the conversational table" (p. 246). Yet this does not seem correct. When we say that Alice should be generous, we update information that it is *desirable* that Alice is generous. Default reading may be useful here: when we say that Alice should be generous (in deontic sense), we are simply saying that Alice ought to be generous unless some more preferred ought overrides this obligation. Finally, there should be "the first update of the fourth line" instead of "the first update of the third line" (p. 248), since the author is obviously discussing the difference between $[\mathbf{w} \mid \mathbf{w} = \bot \omega]$ and $[\mathbf{w} \mid \mathbf{w} = \top \omega]$.

The seventeenth paper *Coarse Deontic Logic (short version)* is written by Van De Putte. The author devises a group of multi-modal logics based on Cariani's semantics. Importantly, these logics invalidate Inheritance (i.e. the inference from OA and A entails B to OB) and allow for coarseness (OA can be true even if there are intuitively impermissible ways of making A true). Resulting logics are compared to some existing deontic logics. The author claims that (C+) implies that (A_B) and (A_I) are equivalent – there should be (A_P) instead of (A_I) (p. 268).

Žarnić is the author of the last, eighteenth paper: *Deontic Logic as a Study of Conditions of Rationality in Norm-related Activities*. Later Von Wright suggested a reinterpretation of deontic logic as the study of rationality conditions of the norm-giving activity. The paper formalizes Von Wright's suggestion within the set-theoretic approach, thus providing us with certain logical pragmatics. Yet it seems to be questionable to what extent is the alleged reinterpretation a genuine reinterpretation, since the theorems are in either case the same. The author writes: "What has been previously understood as a conceptual relation, later becomes a normative relation; a norm for the norm-giving activity, and not the logic of the norms being given" (p. 279-280). However, one can argue that there is no such disanalogy between *proper* logic and deontic logic: *any* logic is primarily concerned with the *right* usage of language, and with *correct* inferences, not with the actual (often flawed) inferences carried out by real humans.

Finally, the end of the book. I have a confession to make: I really enjoyed the reading! The book is definitely a must-read for anyone who is curious about the state of art in deontic logic. I heartily recommend to buy a copy (it is cheap, thanks to College Publications!). However, target readers are certainly not exhausted by the circle of deontic logicians. Since the special focus was argumentation, and the spectrum of used frameworks was incredibly broad, the book might be interesting for any logician or analytic philosopher.

Daniela Glavaničová

Acknowledgments

References

DUŽÍ, M. (2008): Semantic Information and the Paradox of Inference. In: Zouhar, M. (ed.): *Jednotliviny, všeobecniny a významy*. [*Particulars, Universals and Meanings.*] Bratislava: Filozofický ústav SAV, 34-59.

GABBAY, D. M. et al. (eds.) (2013): *Handbook of Deontic Logic and Normative Systems*. College Publications.

JAGO, M. (2014a): *The Impossible: An Essay on Hyperintensionality*. New York: Oxford University Press.

JAGO, M. (2014b): The Problem of Rational Knowledge. *Erkenntnis* 79, No. 6, 1151-1168.

Jan Dejnožka: *Bertrand Russell on Modality and Logical Relevance*
CreateSpace Independent Publishing Platform, 2016, 647 pages[1]

As the title indicates, *Bertrand Russell on Modality and Logical Relevance* investigates two main topics: modality and logical relevance in the work of Bertrand Russell. It claims to be the only study of Russell's views about modality and logical relevance ever written (p. xi) and as such deserves attention of anyone interested in the magnum opus of the philosopher. In the scope of more than six hundred pages, Dejnožka brought to light many aspects of Russell's philosophy which, implicitly or explicitly, record Russell's interest in modal matters. Dejnožka's strategy is quite straightforward: to gather together relevant quotations including modal notions and, consequently, interpret them in a systematic and 'Russell friendly' way. True, such a comprehensive overview is unique and of interest of a wider group of philosophers. Projects of this character though often face a threat of misrepresentation, overestimation of one's position, or simply a danger of going (far) beyond what the particular papers and books bear. Although I am not claiming this is Dejnožka's case, I will try to show some potential risks of the project.

Dejnožka's excursion into the philosophy of Bertrand Russell comes in ten chapters. After an extensive introduction, Dejnožka presents his main objective: to resist a view dubbed as 'V': the view that 'not only did Russell not offer a

[1]   ✉ Martin Vacek
       Institute of Philosophy, Slovak Academy of Sciences
       Klemensova 19, 811 09 Bratislava, Slovak Republic
       e-mail: martinvacekphilosophy@gmail.com

modal logic (he did not), but also that he ignored modality or was against mo-
dality' (p. 35). Namely, the chapter 'Propositional Functions and Possible
Worlds' goes through several Russell's arguments for a theory according to
which logical modalities are certain specific properties of propositional functions
(MDL). Having distinguished between logic and ontology, Dejnožka proceeds
to say a bit more about the latter. In chapters 'Russell's Three-Level Theory of
Modality' (Chapter 3) and 'The Ontological Foundation' (Chapter 4) two main
issues are discussed: three (not rival) senses of 'exist' and three senses of 'possi-
ble' as their correlates; and the most fundamental level of Russell's ontology and
theory of modality (I will return to these chapters in due course). Chapter 5
moves the reader's attention to several Russell's critics or, in other words, pro-
ponents of 'V'. In particular, Dejnožka critically examines Rescher's reasons for
'unwillingness to recognise the merely possible (the *contingently* possible) as a
distinct category' (p. 77, emphasis in original), yet concludes that 'he [Rescher]
agreed…that he had not read much Russell' (p.86).

    The more logic burden part of the book starts with Chapter 6 entitled 'Russell's
Eight Implicit Modal Logics'. The chapter is a test-based since, as Dejnožka sug-
gests, his goal is to describe 'two tests for imputing an implicit modal logic to Rus-
sell' (p. 87). In practice, he aims to show that some bits of Russell's writings can
be paraphrased into MDL and it variants: implicit alethic logic (pp. 92-105); im-
plicit causal logic (pp. 105-110); implicit epistemic logic (pp. 110-116); and im-
plicit deontic logic (pp. 116-119). Chapter 7, 'Russell's Implicit Possible Worlds
Semantics', focuses on Russell's use of possible worlds talk and philosophical is-
sues such talk raises: negative facts (p. 168), rigid designation and trans-world
identity (p. 171), essential properties (p. 173) and the problem of alien individuals
and alien properties (p. 181). Again, through the numerous references, Dejnožka
illustrates Russell's inclination toward possible worlds talk although, as he points
out, in some passages Russell regards such talk as mere "phraseology" (p. 184). In
Chapter 8 the author goes back to roots of Russell's modal considerations. In 'The
Motives and Origins of Russell's Theory of Modality' Dejnožka reintroduces Rus-
sell's three-level theory of modality, summarises his account, boldly concludes that
Russell's critics 'do not know him [Russell] very well' (p. 211) and offers ten crit-
icisms on his own.

    The last two, and the most extensive, chapters move from modality to rele-
vance. The heart of Chapter 9 named 'Russell's Implicit Relevance Logic' is found
in Dejnožka's claim that Russell has a largely explicit theory of both relevance and
an implicit relevance logic. Chronologically, relevance is detected in Russell's

Principles (p. 244), 'Necessity and Possibility' paper (p. 245), *Principia* (p. 247), 'Our Knowledge of External World' (p. 251), *The Philosophy of Logical Atomism* (p. 251), *Introduction to Mathematical Philosophy* (p. 252), *An Outline of Philosophy* (p. 272), *The Analysis of Matter* (p. 252), *Inquiry* (p. 275), *Human Knowledge* (p. 276) and *My Philosophical Development* (p. 276). 'Probability as Degree of Logical Relevance', the final chapter, mostly concerns probability, its relation to induction and the causal relation, and the way Russell treats them. Historically oriented reader will definitely find interesting Dejnožka's 'History Chart of Relevance Rules' (p. 480), 'History Chart of Common Terms for Relevance' (p. 481) and the 'Relevantist Members of the Inner Temple (p. 481).

So much about the structure of the book. In the second part I would like to discuss some (partly interconnected) issues Dejnožka raises in the context of 'Russell's writings + Dejnožka's comment' package. Namely, I will look at the issues of modality as bearing three different, yet interwoven, problems: the problem of semantics, the problem of epistemology and the problem of metaphysics.

On several places in the book, Dejnožka relates the notions of existence and possibility. The core of the debate, Chapter 3, is the distinction between three senses of 'exist'. The first, primary, sense reflects, according to Dejnožka, Russell's robust sense of reality summed in a motto 'to be is not to be nothing'. The second sense of 'exist' is Berkeleyan and Humean conceived as 'to be correlated with other particulars (sense data) in appropriate ways'. It seems that this sense has an epistemological reading since only things we have an appropriate relation to, an acquaintance with, exist. Finally, the third 'exist' concerns the logical structure of existence assertions. Such Fregean definition of existence is to be understood as a property of a propositional function (see also Chapter 2 for more about propositional functions).

So far so good. Now, following the above mentioned distinction, Dejnožka goes on and maps it into an analogous distinction within the scope of 'possible'. According to a primary sense of 'possible' all and only existents are possible. According to the second sense only groups of correlated particulars are possible. Finally, a tertiary sense concerns the logical structure of possibility assertions.

This distinction, however, does not exhaust the structure of Dejnožka's interpretation of Russell. Beside the primary, secondary and tertiary levels, he imposes 'modal features of interest' on every level. To start with the primary level, we get the following: (i) the primary existence of a sense-particular is logically contingent and can be known only through empirical acquaintance (p. 56); (ii) there is no such

thing as a merely possible particular; and (iii) the existence of a particular is transcendentally necessary with respect to thought and language. Descending to the secondary level, another three modal features are identified: (i*) the secondary existence of a group of correlated particulars, qua secondary existence, is not absolutely, but plainly relatively contingent; (ii*) there is a clear sense of relative structural possibility of secondary existence given a primary existence of some particular; (iii*) the secondary existence or ordinary things is not transcendentally necessary. Finally, the tertiary level is further analysed via the following modal aspects: (i**) a propositional function is possible if and only if it is sometimes true; (ii**) a propositional function is possible if and only if it describes something which has secondary reality; (iii**) a tertiary existence assertion may be said to have transcendental necessity in a derivative sense if and only if it is logically deducible by existential quantification over logically proper names. The so-called MDL {1, 2, 3} articulates Russell's full theory of modality.

Dejnožka suggests that MDL {1, 2, 3} is then an articulation of Russell's full theory of modality. To do so however, at least two things should be shown: one, all of (i) – (iii**) are mutually consistent; two, they provide an exhaustive analysis of modal discourse. Dejnožka argues on behalf of the former by resisting the view that, despite its logical consistency, 'x is unicorn' turns out to be impossible on Russell's view. Namely, anything that has secondary existence not only has features (i*) – (iii*) relative to its secondary existence, but also features (i) – (iii) relative to its primary existence. Importantly, 'with respect to modal feature (ii*) on the secondary level of existence unicorns are *combinatorially* both possible and contingent' what is 'perfectly consistent with their being impossible according to modal feature (ii) on the primary level of existence' (p. 60, my emphasis).

Two caveats. First, Dejnožka seems to appeal to a combinatorial theory of modality or, more precisely, a combinatorial theory of possible worlds. Such theory has roots in the Wittgenstein's *Tractatus* and appeared (in some form) in Quine's 'Propositional Objects' (Quine 1968), Cresswell's 'The World is Everything that is the Case' (Cresswell 1972), Skyrms's 'Tractarian Nominalism' (Skyrms 1981) or Armstrong's *A Combinatorial Theory of Possibility* (Armstrong 1990). The core of these theories is a construction of some distribution of matter throughout a spacetime region, be it a Newtonian spacetime or some non-classical spacetime. This however, raises at least two worries: the worry from circularity, and the worry from incompleteness.

The worry from circularity is the following: the set-theoretic constructions, or simply the recombinations, determine the position of mereological atoms (with

'atom' being rather a neutral notion) and their sums as well as the situation on the world at the micro-level. Yet, if it is so (and we have few reasons to think it is not) there should be a tight connection between micro-level and macro-level. Put even stronger, in order to metaphysically explain the going-on in the actual world (explanandum) by means of recombinations (explanans) one has to posit a necessitation relation between the two. Since the relation is modal in nature, we deal with a circular analysis (what can be a reason for Russell's scepticism about modality as a fundamental or irreducible concept).

The worry from incompleteness arises as far as we recombine actual atoms only and omit possibilities of the being merely possible atoms. Although I am not sure how strong the intuition 'there could be worlds with more matter' is, one can still back it up with a simple (transcendental) consideration: a world to which no individuals, worlds, or properties are alien would be an especially rich world. There is no reason to think we are privileged to inhabit such a world. Therefore any acceptable account of possibility must make provision for alien possibilities (cf. Lewis 1986, 93). Dejnožka discusses alien individuals and alien properties in several places (pp. 52, 81, 166, 182) yet he, in my opinion, does not square MDL {1, 2, 3} with this (again, maybe disputable) possibility properly.

Bertrand Russell on Modality and Logical Relevance is literally a full-length study of Russell's views on modality. It does both, highlight the 'modality bearing' passages in which Russell implicitly or explicitly comments on the problems of modality, and interprets them in a spirit of the overall unity, systematicity and Russell's ingenuity. To repeat, it is always a hard and risky enterprise to find an important, although to the date ignored, features in the life works of the most influential philosophers of 20th century. But Dejnožka's book does present one such enterprise and as such is a stimulative and worthy contribution to (the history) of philosophy.

Martin Vacek

References

ARMSTRONG, D. M. (1990): *A Combinatorial Theory of Possibility*. Cambridge University Press.

CRESSWELL, M. J. (1972): The World Is Everything That Is the Case. *Australasian Journal of Philosophy* 50, No. 1, 1-13.

LEWIS, D. (1986): *On the Plurality of Worlds*. Oxford: Blackwell.

QUINE, W. O. (1968): Propositional Objects. *Crítica: Revista Hispanoamericana de Filosofía* 2, No. 5, 3-29.

SKYRMS, B. (1981): Tractarian Nominalism. *Philosophical Studies* 40, No. 2, 199-206.

Petr Glombíček: *The Philosophy of Young Ludwig Wittgenstein* [*Filosofie mladého Ludwiga Wittgensteina*] Nakladatelství Pavel Mervart, Červený Kostelec, 2016, 216 pages[1]

There are numerous monographs about Ludwig Wittgenstein, but only a few of them were published in the territory of former Czechoslovakia. Most of them are translations of books authored by foreign authors, while those by Czech or Slovak authors are rare. Most notably, they include two books by Ondřej Beran, namely *"Střední" Wittgenstein: cesta k fenomenologii a zase spátky* (*The "Middle" Wittgenstein: His Journey to Phenomenology and Back Again)* and *Soukromé jazyky* (*Private Languages*) – see Beran (2013a; 2013b). A collection of papers *Studie k filosofii L. Wittgensteina* (*Studies on the Philosophy of L. Wittgenstein*) published by the Institute of Philosophy of the Czech Academy of Sciences is also worth mentioning (see Dostálová & Schuster 2011). Those who are interested in philosophy are certainly pleased by the fact that a new book by Petr Glombíček *Filosofie mladého Ludwiga Wittgensteina* (*The Philosophy of Young Ludwig Wittgenstein*) has appeared.

Capturing the gist of young Wittgenstein's philosophy is by no means an easy goal. Analysing selected topics *cum grano salis* of "a Schopenhauerian interested in formal logic" with the aim to map and outline the influence of other thinkers on his development is far from a routine task. This was Petr Glombíček's aim, though he admitted that this aim has changed in the course of writing the book. I think one should appreciate the change of focus because the result of Glombíček's effort is a book that is unique, at least in our geographical area.

[1]   ✉   Marián Ambrozy
       Department of Social Sciences, College of International Business ISM Slovakia in Prešov
       Duchnovičovo námestie 1, 080 01 Prešov, Slovak Republic
       e-mail: ambrozy.marian@gmail.com

The book is not a standard introduction to the philosophy of "early" Wittgenstein. The readers who expect this kind of content should certainly choose a different monograph to read. On the contrary, the book presumes that its readers have read Wittgenstein's texts and know the basic facts about his life. It is thus a book for an advanced student of Wittgenstein's philosophy. At the same time, the monograph does not present Wittgenstein's issues in a typical analytic way. This does not mean that reading it cannot be beneficial for philosophers with analytic focus. One should appreciate that knowledge of mere fundamentals of logic, as taught in courses for undergraduate students of philosophy, is sufficient to understand the book.

One cannot deny that the author is very well acquainted with both Wittgenstein's texts and biographical facts. Petr Glombíček tries to reveal early Wittgenstein's relations to several authors (though some of them were non-philosophers). He is aware of the fact that Schopenhauer's influence on Wittgenstein is, in the worst case, ignored or, in the best case, discussed within one brief paragraph. Glombíček's monograph, written in a very readable style, tries to fill this gap.

The first chapter is an attempt to solve a riddle: what was the purpose of publishing Wittgenstein's first work, namely *Tractatus logico-philosophicus*? I appreciate that the author does not wish to give an unequivocal and ultimate interpretation of the *Tractatus*. Rather, he tries to open a discussion. Glombíček partially keeps his distance from the so-called new Wittgenstein supporters who claim that the *Tractatus* is a provocative nonsense. He believes that the above interpretation undervalues and marginalizes the significance of certain key parts of the book. Glombíček presents his own alternative; nevertheless, he acknowledges that the new Wittgenstein supporters have contributed to revealing a therapeutic goal of the book in curing the need to solve philosophical problems.

The first chapter of the book presents several serious problems. One of them concerns the question about who is a possible addressee of the *Tractatus*. The author finds it difficult to determine which reader was supposed to be made happy by the book (based on Wittgenstein's correspondence with the publisher, it was perhaps one particular person). Glombíček thinks it probably was Russell. Nevertheless, Wittgenstein himself claimed that neither Russell nor Frege understood the *Tractatus*.

Besides focusing on obligatory analytic motivations of the *Tractatus*, Glombíček pays attention to non-analytic influences that one can detect in the book as well. Thus, apart from the well-known connections to Frege's and Russell's original ideas, the readers will be surprised by the number of non-analytic

inspirations that Glombíček depicts in his book. To take a somewhat curious ex-
ample, he explains that Wittgenstein's phrases of *misunderstanding of logic of
our language* and *language logic* were taken from P. Ernst's afterword to the
Grimm brothers' fairy tales. He discusses this topic in some detail in the third
and fourth chapter.

The author claims that, according to Wittgenstein, philosophical problems re-
sult from incorrect use of language. Glombíček disagrees with the view that the
Tractatus does not have a meaningful and philosophically beneficial content. After
all, the book aims at determining the boundaries of the language and defining the
right means of the language. The author claims that Wittgenstein clearly distin-
guishes thought (*die Gedanke*), expression of thoughts (*der Ausdruck der
Gedanken*) and act of thinking (*das Denken*). Glombíček reminds us of Wittgen-
stein's confident claim that his implement solved all philosophical problems by
pointing at their language meaninglessness.

The second chapter of the book is devoted to ethics. Glombíček discusses the
Tractatus as well as the well-known *Lecture on Ethics*, which was one of the few
texts that Wittgenstein presented in public. Wittgenstein said that the ethical part
of the *Tractatus* was not written. According to what he wrote in his letter to pub-
lisher Ficker, only a few people will understand his book and the publisher will
certainly not be among them. Glombíček draws our attention to the fact that, in his
letter to Ficker, Wittgenstein put stress on the unwritten part of the *Tractatus*. The
book was supposed to show that any discussion on issues that belong to theoretical
ethics simply makes no sense.

In accordance with Husserl's eidetic reduction and based on some definitions
of ethics, Wittgenstein tried to describe what ethics deals with. He was aware of
the fact that it is impossible to provide a precise definition of ethics and thus, as
Glombíček says, he tried to identify its basis in a Husserl-like way.

Glombíček further claims that one can detect here Schopenhauer's inspiration
too. He presents Wittgenstein's well-known assertions regarding the transcenden-
talism of ethics and aesthetics and the impossibility to express any absolute value
as a mere stating of facts. We are further told that Wittgenstein also draws from
works of G. E. Moore, mainly in connection with the idea that it is impossible to
define goodness verbally. Inspired by Schopenhauer, Wittgenstein pointed to so-
called paradigmatic experiences. In his lecture, he presented three such experiences
– the feeling of guilt, the feeling of being absolutely safe and the amazement at the
existence of the world. However, one can describe such an experience allegorically
at most, without expressing its ethical or religious value.

Glombíček also discusses Wittgenstein's lectures for the Vienna Circle in 1929 and 1930. In one lecture, Wittgenstein claimed that he understood what Heidegger meant by anxiety expressed by a feeling stemming from the ignorance of the meaning of life and being. He believed it was intertwined with the boundaries of the language. The effort to go beyond these boundaries is thus just a blathering attempt at formulating ethical statements. At the same time, Wittgenstein implied that his intention was similar to that of Heidegger and he tried also to compare it to the ideas of Augustine or Kierkegaard. Despite admitting that absolute ethical statements are nonsensical, Wittgenstein did show some understanding for attempts to say what is impossible to express.

Glombíček describes how Wittgenstein explained the notion of miracle. Miracle in the relative sense of the word means that we have not analysed a process or a phenomenon. It remains unknown. This is miracle in the relative sense of the word. In the absolute sense of the word, miracle means undertaking an experience that is similar to the kind of experience he was able to identify on the basis of Heidegger's understanding of anxiety. An absolute miracle can be thus associated with the realm of the mystical, i.e. something that cannot be expressed by language but can only be shown.

The third chapter presents Schopenhauer's ideas that influenced Wittgenstein. The author summarizes certain elements of Schopenhauer's philosophy, his understanding of the subject, the field of interpersonal relations from the viewpoint of subject-other subject relation, and points to several paragraphs Wittgenstein used particularly in the *Tractatus*. The author also points to particular Schopenhauer's formulations that were used by Wittgenstein. He mentions Schopenhauer's words regarding the impossibility to clarify the sense of the world and of one's existence. Similarly, Schopenhauer often uses the metaphor about studying and literature as a ladder to knowledge that becomes useless once knowledge has been achieved. In his work *The World as Will and Representation*, Schopenhauer said that man was his own world, microcosmos, and that his death was the end of the world for him. He also stated that grammar was related to logic just as clothes were related to the body, etc. As Glombíček points out, the list of Schopenhauerian allusions is far more extensive. Schopenhauer's influence is apparent mainly at the end of the *Tractatus*. Wittgenstein's boundaries of the language remind us of Schopenhauer's boundaries of the field of vision. Thus it seems that Schopenhauer indirectly inspired Wittgenstein to develop his theses regarding distinctions between saying and showing. Schopenhauer's influence also is apparent in the case

of understanding the world of the object whereas the subject is not a part of the world.

Glombíček points to further connections between Wittgenstein and Schopenhauer which he explains as an inspiration by Schopenhauer. Schopenhauer said that the meaning of the world was a riddle and cannot be part of the world (the will is blind and human life is meaningless). Wittgenstein was speaking about meaning that cannot be expressed by language and of the impossibility of ethical statements. Schopenhauer's influence is obvious here. In addition, Wittgenstein's understanding of transcendent (and common) nature of ethics and aesthetics has Schopenhauerian origin in Schopenhauer's ethics which stems from aesthetics. Moreover, when explaining the aesthetic viewpoint both Schopenhauer and Wittgenstein use the "*sub specie aeterni*" perspective with almost identical Latin lexis. Glombíček emphasises that both believed in the senselessness of scepticism, though they elaborated different argumentation in this matter.

Glombíček's claim that Schopenhauer's influence on Wittgenstein was enormous and that Wittgenstein even adopted Schopenhauer's phrases in several places is correct. In claiming this, Glombíček joins by G. E. M. Anscombe who endorsed the same view. I believe that Glombíček presented cogent reasons for documenting this influence. Nevertheless, he correctly points out that Wittgenstein did not consider Schopenhauer a master; it is the other way round – he actually challenged Schopenhauer's views.

The fourth chapter is devoted to the influence of several physicists on young Wittgenstein. The author specifically emphasises Hertz and Bolzmann. It was, however, Hertz who played the key role in shaping Wittgenstein's opinions. In his attempt to detect the origins of Wittgenstein's logical isomorphism, Glombíček mentions Hertz as the main inspiration. Wittgenstein referred to him in his early works and sporadically also in his later texts. Glombíček identifies Hertz's influence mainly in connection with the picture theory as well as the effort to disclose pseudo-problems and pseudo-questions. According to Glombíček, there are many common features between the picture theory of meaning from the *Tractatus* and Hertz's picture theory. He further points out that Hertz directly influenced Wittgenstein's views concerning the difference between explanation and clarification.

The final chapter summarizes Glombíček's results. He states that the difference between what is said and what is shown is very important and illustrates various uses of the term "to show" in Wittgenstein's work. He mentions Wittgenstein's understanding of mysticism or his description of the role of philosophy as a

practice concerning clarifying sentences. The author also summarises Wittgenstein's opinions on solipsism. Glombíček lists the differences among the speakable, mysticism, and nonsense. Wittgenstein says that the unspeakable and the mythical is what can be shown. The speakable is expressed by means of meaningful sentences of natural sciences. We can view Wittgenstein's understanding of some of Heidegger's thoughts or some religious statements along these lines. Simply said, in the *Tractatus* one needs to invoke the distinction between *unsinn* and *sinlos*. Glombíček recommends reading the *Tractatus* simultaneously with the *Lecture on Ethics* because it can help to understand the above-mentioned differences better.

I would like to add a few critical remarks. I assume that the author is well acquainted with the facts of Wittgenstein's life. Therefore, I would appreciate a more extensive reference to such facts and a more elaborated analysis of their influence on Wittgenstein's philosophical viewpoints. It is well known that there were many such events (his brothers' death, his exemplary military behaviour in the WWI, his work as a gardener or a teacher, Russell's views on his career, etc.). Wittgenstein's biography surely is crucial to understanding many of his views.

The book extensively describes Schopenhauer's and Hertz's influences. However, I would appreciate if a similar space were devoted to Wittgenstein's relations with Russell's philosophy and to Frege's influence. Though both of them are mentioned in the book, this is done mainly with respect to publishing the *Tractatus*.

This book on the philosophy of young Wittgenstein is by no means introductory, but assumes that readers are familiar with Wittgenstein's issues. It is this fact that makes the book so valuable. Its exceptional contribution consists in that, by analysing relevant texts, it precisely documents the influence of certain thinkers on Wittgenstein. Similarly, it explains many of young Wittgenstein's key ideas and, in doing so, takes into consideration the influence of the above authors and biographical facts. Furthermore, Glombíček emphasises Wittgenstein's impact on analytic philosophy as well as on non-analytic strands. This book is highly beneficial to analytically oriented readers. I assume, however, that it can be fully appreciated only by the readers that are acquainted with at least elementary knowledge of analytic philosophy.

Marián Ambrozy

References

BERAN, O. (2013a): *"Střední" Wittgenstein: cesta k fenomenologii a zase zpátky* [*The Middle Wittgenstein: His Journey to Phenomenology and Back Again*]. Červený Kostelec: Nakladatelství Pavel Mervart.

BERAN, O. (2013b): *Soukromé jazyky* [*Private Languages*]. Filosofia.

DOSTÁLOVÁ, L. & SCHUSTER, R. (eds.) (2011): *Studie k filosofii L. Wittgensteina* [*Studies on the Philosophy of L. Wittgenstein*]. Filosofický ústav AV ČR: Filosofia.

UNESCO Philosophy Day/Night 2016[1]

This year World Philosophy Day was celebrated immediately after International Tolerance Day. This proximity was notable because of the interconnection between tolerance and philosophy. Philosophy is indicative of understanding and respect towards a variety of opinions, views and cultures that can enrich people's lives. Equally, as with tolerance, philosophy also stands for the ability to live next to each other with appropriate respect for the rights and values of other individuals. It is the ability to see the world through a critical eye, to be aware of opinions and views of other people, to enhance the freedom of thought, conscience and faith.

UNESCO Philosophy Night 2016 was a night-time event where philosophy and art met the general public. The Director-General of UNESCO, Irina Boková, stated that "[...] philosophy is more than just an academic subject; it is a day-to-day practice which helps people to live a better life by exercising and applying humanity in everyday situations". One should start learning how to do it at the earliest age and, gradually, perfect oneself. In her view, philosophy can also be understood as an important key to an inspirational public discussion, defending humanity, troubled by violence and tension in the world. In this context, it needs to be stated that philosophy does not offer any immediate, ready solutions, but merely the eternal task to doubt the world and try and make it a better place to live. In the course of this journey, tolerance serves as a moral virtue and a useful tool for dialogue. It has

[1]   ✉ Marta Gluchmanová
        Department of Humanities, Technical University of Košice
        Bayerova 1, 080 01 Prešov, Slovak Republic
        e-mail: marta.gluchmanova@tuke.sk

      ✉ Michaela Joppová
        Institute of Ethics and Bioethics, Faculty of Arts, University of Prešov
        Ul. 17. novembra 15, 080 01 Prešov, Slovak Republic
        e-mail: michaela.joppova@gmail.com

      ✉ Vasil Gluchman
        Institute of Ethics and Bioethics, Faculty of Arts, University of Prešov
        Ul. 17. novembra 15, 080 01 Prešov, Slovak Republic
        e-mail: gluchman@unipo.sk

nothing to do with naive relativism which claims that everything is equally valid; listening is an individual imperative, as it is based on the decisive commitment to fight for general principles of dignity and freedom.

In 2016, UNESCO celebrated the posthumous birthdays of two significant philosophers, Aristotle and Leibniz, who contributed to the development of metaphysics and science, logic and ethics. They both, several centuries apart and in greatly differing cultural contexts, placed philosophy at the heart of public life as the peak of a free and dignified life. It was in this spirit that this entire significant event took place and opened space for free, open and tolerant thoughts. Based on this dialogue it could be assumed that a lasting cooperation between citizens, societies and countries, as a basis of permanent peace, can be built.

UNESCO Philosophy Night 2016, organised within the World Philosophy Day, set the goal to create conditions for philosophy and art to meet the general public. This event has been organised in the Paris UNESCO seat for the seventh time in cooperation with the French National Commission for UNESCO and lasted from the evening of November 18th, 2016 until the early morning of November 19th, 2016.

This event was based on a successful conception of events that preceded it, held by the French philosopher and theatre director Meriam Korichi in Paris, London, Berlin, New York and Helsinki. The night presentations in their standard format were open to the general public and were truly well supported and received. A great number of philosophers offered an overview of contemporary philosophy, as well as new views of UNESCO's mission. Special attention within the programme was paid to the posthumous birthdays of Aristotle and Leibniz, as well as the umbrella topic of tolerance.

The event made full use of the premises of the UNESCO headquarters, providing traditional lectures on philosophy and many innovative activities, such as art installations, live performances, films and video viewings, presenting philosophy in a creative, aesthetic, playful and provoking way. The aim of the event was to encourage each participant to actively think about the contemporary world.

The event offered a great number of interesting presentations dedicated to such personalities as Nietzsche (Hamed Fouladvind), Kierkegaard (Sharon Krishek), as well as to topics such as tolerance (Rainer Forst), moral relativism (Paul Boghossian), building of civil society (Islah Jad), diversity in philosophy (Nele Goutier, Lianne Tijhaar), freedom and safety (Regina Kreide), Buddhist and Confucianist ethics (Eun-su Cho, Wei Xiao), etc. For the very first time, the worldwide

event held in the Paris UNESCO headquarters was attended by philosophers from Slovakia who, in their presentations, talked about Augustín Doležal (Vasil Gluchman), fiction and reality (Martin Vacek), as well as virtual reality (Jozef Sivák).

The UNESCO Chair in Bioethics at the University of Prešov which is a leading institution in bioethics in the Central Europe (Doričová & Pazdera 2015, 233-235; Gluchman 2012, 5-8) also joined the framework of the worldwide events organised by UNESCO on World philosophy Day and, together with the Institute of Ethics and Bioethics and the Institute of Philosophy (both at the Faculty of Arts, University of Prešov), for the very first time in Slovakia, organised an event as part of UNESCO Philosophy Day/Night, which took place in Prešov on November 10th, 2016. The aim was to emphasise and honour the contribution of philosophy to the formation of human thought and culture, as well as appreciate its irreplaceable role in the process of personal individual growth. The common topic of lectures, text-based seminars and workshops organised within UNESCO Philosophy Day/Night 2016 in Prešov was the history of philosophy and ethics.

As part of this day, the organisers prepared for specialists as well as the lay public a series of concurrent lectures, text-based seminars and student workshops. The cycle of lectures began with the topic *Models of philosophical ethics (Aristotle, Kant)*, in which Viera Bilasová presented two essential models of a theoretical approach to ethics – Aristotelian and Kantian. At the core of the contribution was an explanation of the dual concept of practical rationality through the optics of Kantian deontology and Aristotelian ethics of virtue. In the lecture that followed, *Notes on Gabriel Marcel's philosophy: Philosophy of hope,* Marián Palenčár presented the life and work of the above French philosopher, focusing on a number of interesting details from his academic, as well as non-academic, activities.

In the evening-time block of lectures, Vasil Gluchman gave a paper entitled *John Stuart Mill: The art of life*. He introduced to the listeners the broader context of Mill's thought as a representative of classical (non-Benthamian) utilitarianism with regard to the art of life (or its attributes of quality). The follow-up discussion with the audience was directed at an examination of quantitative and qualitative orientations of classical utilitarianism.

In the lecture *Philosophy of history of Scottish and German Enlightenment*, Sandra Zákutná provided an outline of the influence of the German Enlightenment on the formation of contemporary Scottish thinkers. At the same time, she pointed to historical contexts of the given mutual intersections of thought by means of philosophy of history. The block of evening-time lectures was concluded by Pavol

Suchárek's paper entitled *The other as different in the philosophy of Emmanuel Lévinas*, in which he, in detail, explained the basic principles of Lévinas' ethics and metaphysics, mainly with regard to his understanding of difference, transcendence and obligation.

Concurrently with the above lectures, text-based seminars took place focused on the reading and interpretation of (selected extracts from) classical philosophical source works. These were chaired by senior assistants and doctoral students from participating institutes at the Faculty of Arts, University of Prešov. By means of these, the listeners had a chance to familiarise themselves and learn how to work with challenging philosophical texts. In the block of text-based seminars, the first extract came from Aristotle's *Nicomachean Ethics*. The seminar was dedicated to the first volume of the above work and was chaired by Katarína Komenská. Its primary aim was to comprehend Aristotle's understanding of good, based on his critiques of Plato's definition. The second seminar was aimed at Martin Heidegger's, *Being and time*. In spite of its rather challenging nature, thanks to the professional chairing by Stanislav Olejár, the listeners were able to work with the given text. In the text-based seminar aimed at Baruch Spinoza's *Ethics*, co-chaired by Ján Kalajtzidis and Michaela Joppová, the participants analysed Spinoza's critique of anthropomorphism with regard to nature and religion. The final text-based seminar, chaired by Pavol Suchárek, worked with Emmanuel Lévinas' work *Totality and infinity*. The main topic of this seminar was the chapter entitled *Ethics and the face*.

Ethics-related workshops were an important part of the event; prepared by students (under the leadership of their tutors) in the final year of their Master degree at the Faculty of Arts, University of Prešov. The block included the following: a workshop entitled *Homophobia from the viewpoint of moral philosophy* lead by Štefan Oreško, and a workshop on *Morality and forms of hedonism* co-lead by the students Marek Regenda and Tomáš Talpaš. The audience could familiarise themselves with not only philosophical thought in the past but also contemporary philosophical-ethical issues – homophobia and postmodern hedonism. Participants at the workshops could practise their critical thinking and argumentation skills in the context of the above up-to-date topics.

UNESCO Philosophy Day/Night 2016 was full of intellectual input and efforts for critical reflection in a popular form which, we would like to hope, could motivate, in a great number of people, an interest in philosophy not only as cultural heritage but also a dynamic and constantly developing sphere of thought which is no less relevant at present than it was in the past. It could, thus, be concluded that,

in the true sense, the event was a celebration of philosophy and human thought in a complex form.

Marta Gluchmanová, Michaela Joppová, Vasil Gluchman

References

Doričová, I. & Pazdera, M. (2015): Bioethical Activities of the UNESCO Chair in Bioethics. *Ethics & Bioethics (in Central Europe)* 5, 233-235.

Gluchman, V. (2012): UNESCO Bioethics Education Program. *Ethics & Bioethics (in Central Europe)* 2, 5-8.

Two Conferences on Logic Held in Bochum[1]

Two interesting and well-organized conferences on logic took place at the Ruhr-University of Bochum at the beginning of May. The first of them, *PhDs in Logic IX* (2 – 4 May 2017) was the ninth edition of annual conferences for graduate students working in the field of (mathematical, philosophical, computational) logic, this year organized by Christopher Badura, AnneMarie Borg, Jesse Heyninck, and Daniel Skurt. The second one, *Logic in Bochum III* (5 – 6 May 2017) was the workshop organized by the *Research Group for Non-Monotonic Logic and Formal Argumentation* (AnneMarie Borg, Christian Straßer, Dunja Šešelja, Jesse Heyninck and Pere Pardo). To the delight of the author of this report, the latter conference especially focused on deontic logic this year.

The former conference hosted six (!) brilliant keynote speakers, who presented tutorials related to various areas of logic. Christian Straßer gave a tutorial on *Non-monotonic logic*, keeping the talk gentle and touching deep issues in the nonmonotonic logics at the same time. María Manzano gave tutorials *Leon Henkin on Completeness* and *Identity, Equality, Nameability and Completeness*. The former tutorial was concerned with Henkin's proofs of Completeness (for type theory, first

---

[1]   ✉ Daniela Glavaničová
       Department of Logic and Methodology of Sciences, Faculty of Arts
       Comenius University in Bratislava
       Gondova 2, 814 99 Bratislava, Slovakia
       e-mail: dada.baudelaire@gmail.com

order logic, and propositional type theory). The latter tutorial was focused on the notions of identity and equality. Petr Cintula presented *A Gentle Introduction to Abstract Algebraic Logic I & II*, talks not very gentle to philosophers, but nevertheless interesting ones, explaining an important branch of logic. João Marcos introduced us to *Classic-like Analytic Tableaux for Non-Classical Logics* (non-classical understood as many-valued). Gabriella Pigozzi gave two intriguing tutorials on *The Logic of Group Decision: An Introduction to Judgment Aggregation*. How does a group decide on some issue, such as whether to hire a researcher? Various approaches were presented. Heinrich Wansing presented a talk on *Bi-Connexive Variants of Heyting-Brouwer Logic*.

Twenty contributed talks were manifold, faithful to the currently reigning era of logical pluralism. Talks given revolved around topics such as proof mining, intuitionistic logic, many valued logics, metamathematics and metalogic, deontic logic, formal semantics, modal logics, and (hyper)graph theory.

The latter conference hosted four excellent keynote speakers: Gabriella Pigozzi, Allard Tamminga, João Marcos, and Niko Strobach. Gabriella Pigozzi presented a talk on AGM contraction and revision rules using input/output logic: *Changing Norms: A Framework for Changing Rules*. Allard Tamminga in his talk *Collective Obligations and Obligations of Individuals* argued for the need to distinguish individual obligations from collective obligations. The general idea was that a group adopts a group plan coordinating the actions of its members. João Marcos gave a talk *Do Not be Afraid of the Unknown,* presenting a general logical framework based on different cognitive attitudes of agents concerning rejection or acceptance. Niko Strobach presented the last talk of the conference, *Difference – Classical and Paraconsistent*. The talk evolved around the two notions: difference and identity. It was suggested to treat the notion of difference as primitive, thus contrasting the approach from Priest's book *One*.

The first day of the conference was devoted (mainly) to deontic logic. The first contributed talk given by Ilaria Canavotto was concerned with multi-agent dynamic action logic. The second contributed talk presented by Roberto Cuini was focused on modelling obligations-related deliberation by epistemic agents. The third talk by Federico L. G. Faroldi ventured towards developing a first-order hyperintensional deontic logic. Deontic section was closed by Frederik Van De Putte and his talk aiming to model permitted choices of an agent in a generalised stit logic. Mathieu Beirlaen gave the last talk of the first day on abductive and inductive reasoning, using the framework of adaptive logics. Grigory Olkhovikov gave the first contributed talk of the second day on justification stit logic. Heinrich Wansing

talked about modal extensions of first-degree entailment logic. The last contributed talk given by Sergey Drobyshevich was focused on intuitionistic modal logic.

Talks presented on both of the conferences were followed by a lively debate, containing ingenious questions and suggestions. Participants were sharing ideas not only throughout the conference, but also during the coffee breaks, lunch breaks and conference dinners, what serves as an evidence of genuine interest in the topics discussed. In sum, these two conferences were intriguing, pleasant, and important academic events: a pleasure for any logician.

Daniela Glavaničová

Acknowledgments

Department of Logic and Methodology of Sciences, Comenius University in Bratislava

and

Institute of Philosophy, Slovak Academy of Sciences

Current Trends in Deontic Logic 2017

November 23-24, 2017
(Bratislava, Slovakia)

Keynote speakers

SVEN OVE HANSSON (Royal Institute of Technology, Stockholm)
OLIVIER ROY (University of Bayreuth, Bayreuth)

We invite submissions for a 30 minute presentation followed by a 15 minute discussion. Blinded abstracts of 200-500 words should be submitted to **ctdl2017@gmail.com**. Please attach a separate cover page with your name, affiliation and contact information.

Deadline for submission: **June 30, 2017**
Notification of acceptance: **August 15, 2017**

Organizing Committee

Daniela Glavaničová (Comenius University), Tomáš Kollárik (Comenius University), Martin Vacek (Slovak Academy of Sciences), Marián Zouhar (Comenius University, Slovak Academy of Sciences)