Contents ▪ Obsah

Articles ▪ State

Book Reviews ▪ Recenzie

Reports ▪ Správy

Organon F 22 (3) 2015: 282-301

Subjective Theories of Personal Identity and Practical Concerns

RADIM BĚLOHRAD

Department of Philosophy. Faculty of Arts. Masaryk University
Arna Nováka 1. 602 00 Brno. Czech Republic
belohrad@phil.muni.cz

RECEIVED: 24-04-2015 • ACCEPTED: 10-06-2015

ABSTRACT: This paper focuses on three theories of personal identity that incorporate the idea that personal identity is the result of a person's adopting certain attitudes towards certain mental states and actions. I call these theories *subjective theories of personal identity*. I argue that it is not clear what the proponents of these theories mean by "personal identity". On standard theories, such as animalism or psychological theories, the term "personal identity" refers to the numerical identity of persons and its analysis provides the persistence conditions for persons. I argue that if the subjective theories purport to provide a criterion of numerical personal identity, they fail. A different interpretation may suggest that they purport to provide a non-numerical type of identity for the purpose of providing plausible analyses of certain identity-related practical concerns. I argue that the criteria the subjective theories provide fail to capture several of the identity-related concerns. As a result, this interpretation must be rejected as well.

KEYWORDS: Compensation – identification – numerical identity – personal identity – responsibility – self-constitution – self-interested concern.

1. Introduction

In the current theory of personal identity, two strikingly different approaches can be distinguished in the attempts to define the notion of personal identity. On the one hand, there are theories according to which the relation of personal identity holds between persons if and only if there are

some other relations, such as biological or psychological continuity, which connect the persons. The definition of the identity relation takes the following form:

(OI) Necessarily, for any x, if x is a person at t and something y exists at t*, x=y if and only if x at t and y at t* stand in relation R, where R is the relation preferred by the particular theory.

Such a definition of personal identity provides a criterion of *numerical* identity of persons, because it states on what conditions a person identified at one time is the same entity as a person identified at another time, as well as implies answers to the questions of when persons begin to exist, what changes they can persist, and when they cease to exist.

However, there are some theories that use the concept of personal identity differently. These theories include M. Schechtman's, K. Korsgaard's, and C. Rovane's theories, and, for reasons to be specified shortly, I will call these theories *subjective* (as opposed to *objective* theories, outlined above). The idea is that personal identity is not a relation that holds in the lives of persons independently of their beliefs and attitudes. Rather, persons *determine* what their identity is and constitute themselves, and they achieve this by adopting a certain *attitude* to certain actions or mental states such as experiences, beliefs, intentions. As a result of this attitude, these mental states and actions begin to *characterize* the given person, or, in other words, by adopting this attitude the person makes them *her own*, or, in still other words, they become *part of her identity*. Each of the theories I discuss provides a unique analysis of this identity-constituting attitude. However, before I introduce the particular analyses, I will refer to the attitude as the attitude of *identification*.

We can now express the general idea behind subjective theories more formally as follows:

(SI) Necessarily, for any x, if x is a person at t and there is a set of mental states and actions M at t*, M will be part of x's identity at t if and only if x at t identifies with M.

According to the proponents of subjective theories, they offer much more plausible grounding for certain practical concerns that have traditionally been taken to presuppose personal identity. It is widely believed, for in-

stance, that the notions of *responsibility*, *compensation* and *self-interested concern* presuppose personal identity. Proponents of subjective theories claim that it is their theories, as opposed to the objective ones, that best explain these concerns.

Most importantly, some statements of the proponents of subjective theories even seem to suggest that their criteria have implications for the *persistence* of persons, that is, for the *numerical* identity of persons. If the claims are taken seriously, subjective theories are further committed to the following thesis:

(SIP) Necessarily, for any person x at t and any set M of mental states and actions at t*, if M is part of x's identity, then for any y at t*, y is the subject of the mental states and actions M if and only if x=y.

This principle states that if a set of (possibly past or future) mental states and actions is *part of a person's identity* (in the sense defined by the individual theories), then whoever is (was, will be) the subject of those mental states and actions must be *numerically* identical to the person and vice versa. This statement seems like a truism, because it is hard to imagine a situation in which some characteristics were *mine* without their bearer being *me*. But it is, actually, an open question whether subjective theories are committed to it. This question along with the question if a person can make an action or a mental state part of her identity merely by identifying with them in the way that the discussed theories propose, will be the subject of this paper.

I will first explain what it means to say that personal identity is a subjective relation. Next, I will provide examples of subjective theories with detailed descriptions of their claims concerning the concept of personal identity. We will see that their proponents sometimes speak as if they were addressing the issue of *numerical* personal identity, because they seem to suggest that their proposed criteria have implications for the existence and persistence of persons. I will argue, however, that defining numerical identity by means of the criteria proposed by the theories leads to problems and paradoxes. I will then suggest that subjective theories might be addressing a *different* concept of personal identity: one which does not have implications for our persistence and is not committed to (SIP). I will call this notion *practical identity* and provide textual evidence to support this interpretation.

However, I will also argue that while this interpretation avoids some of the problems of the first interpretation, it faces new problems: the concept of practical identity covers a range of characteristics that cannot all be accounted for by means of identification. Thus, we should reject the conviction that the subjective theories provide an analysis of personal identity of any sort, even if they may be useful in analysing aspects of some particular practical concerns.

2. Subjective theories of personal identity

Before presenting the details of subjective theories it is necessary to define what I mean by the term *subjective*.

(S) The relation of personal identity is *subjective* iff its exemplification constitutively depends on an attitude that the person has towards certain actions or mental states, such as beliefs, desires, experiences.[1]

In the following sections we will look in a greater detail at the mechanism of identity constitution that the individual theories describe, including the particular forms of identification they propose as identity-constituting.

2.1. Narrative identity

I will begin my illustration of subjective theories with Marya Schechtman's narrative self-constitution view (see Schechtman 1996). Schechtman's motivation is to develop a theory of personal identity that captures our intuitions about the identity-related practical concerns listed above. Schechtman provides a *characterization* criterion of identity, which specifies under what conditions a mental state or an action characterizes, or is attributable to, a person. She develops a *narrative self-constitution* theory of characterization. This theory is based on the idea that persons are self-creating beings and that persons' lives have a narrative form. Persons constitute themselves by coming to think of themselves as persisting subjects who have had experiences in the past and will continue to have experiences

[1] This is an adapted version of Huemer's definiton of a subjective property. See Huemer (2005, 2).

in the future, taking certain experiences as theirs by incorporating them into a self-told story of their lives (cf. Schechtman 1996, 94). An experience or an action characterizes a person as long as and to the degree that it is incorporated in a story that the person creates about her life.

This theory is committed to principle (SI), which takes the following form:

(SIN) Necessarily, for any x, if x is a person at t and there is a set of mental states and actions M at t*, M will be part of x's identity at t if and only if x at t incorporates M into the narrative of her life.

According to this criterion, personal identity consists in the existence of a coherent story of a person's life, and that fact further presupposes that the person adopts a certain attitude towards certain experiences and actions, in which she incorporates them into her narrative and, thus, makes them her own. This makes it a subjective theory according to my criterion.

This conclusion can be supported by several of Schechtman's claims: According to Schechtman, personal identity is a product of a person's action:

> I develop a view according to which a person creates his identity by forming an autobiographical narrative – a story of his life. (Schechtman 1996, 93)

That action has the form of attitude adoption:

> An identity in the sense of the characterization question, is not, I claim, something that an individual has whether she knows it or not, but something that she has *because* she acknowledges her personhood and appropriates certain actions and experiences as her own. (Schechtman 1996, 95)

> Personhood and personal identity thus rely crucially on an individual's inner life and her attitude toward her actions and experiences. (Schechtman 1996, 95)

So, personal identity is a subjective relation in Schechtman's theory. But what exactly does Schechtman mean by *personal identity*? In particular, does the concept have implications for the persistence of persons? Textual evidence suggests that it at least purports to do so. Schechtman demonstrates

this goal with an illustration of someone who has long been the victim of a violent, abusive spouse, which results in deep personality changes – the person becomes timid and fearful, supresses her own desires and characteristics, severs crucial relationships, and may have trouble identifying with the teenager she sees in a high school photo. Schechtman argues that it would be appropriate to say that the person has "lost her identity", that she "is no longer the same person", and that "the person we knew is gone" (cf. Schechtman 1996, 88). She adds that these claims need not be considered entirely metaphorical (cf. Schechtman 1996, 88), and finally concludes that "the degree to which a person is alive, and hence survives, seems linked to the degree to which her actions, experiences, and characteristics are her own …" (Schechtman 1996, 89).

The idea seems to be that if a person does not incorporate her past actions and mental states into her current narrative, those actions and mental states are not her own, that is, they are not part of her identity (according to (SI)), and, therefore, she cannot be numerically identical to the person who had the experiences and carried out the actions (according to (SIP)).

When I have presented further examples of subjective theories, we will see whether this idea is coherent.

2.2. The unity of agency

In an influential paper (see Korsgaard 1989), Christine Korsgaard develops a theory to challenge Derek Parfit's theory of personal identity and its implications for our identity-related practical concerns.[2] A central notion in Korsgaard's agential theory of personal identity is the notion of *authorial connection*. She states that the difference between actions and choices on the one hand and mere behaviour determined by biological and psychological laws on the other is the fact that the former require agents and choosers, i.e. they have a subject. The relationship of agents to actions and choices is essentially *authorial*. Unlike mere happenings, our actions and choices are essentially *our own* (cf. Korsgaard 1989, 121). Authorial connections stratify the class of our mental states into levels of differing impor-

[2] It would be beyond the scope of this paper to map the whole dispute. For an exposition of the main differences and objections see Shoemaker (1996) and Bělohrad (2014).

tance. According to Korsgaard, the mental states that we have an authorial connection to are much more relevant for personal identity:

> This is because beliefs and desires you have actively arrived at are more truly your own than those which have simply arisen in you... (Korsgaard 1989, 121)[3]

Korsgaard illustrates the role of authorial connections in personal identity through a popular thought experiment involving a mad surgeon who drastically manipulates a person's memory and character. On a standard psychological theory, the severe discontinuity in the person's psychology causes the person's demise and her replacement by another person. Korsgaard, however, maintains that it is not the discontinuity itself that causes the demise of the person. Rather, it is the fact that the intervention is external and unauthorized by the person. In other words, even severe changes in psychology are consistent with personal identity, as long as those changes are the product of the person's own choice. Korsgaard concludes that "the sort of continuity needed for what matters to me in my own personal identity essentially involves my agency" (Korsgaard 1989, 123).

Korsgaard's theory has all the elements characteristic of a subjective theory of personal identity. The relevant identity-constituting attitude is authorial connection. Personal identity is a product of agency, and agency consists in authorizing mental states and actions. Only the mental states and actions that are authorized in this manner become characteristic of who the person is, that is, part of her identity. Criterion (SI) takes the following form in Korsgaard's theory:

(SIA) Necessarily, for any x, if x is a person at t and there is a set of mental states and actions M at t*, M will be part of x's identity at t if and only if x at t authorizes the metal states and actions in M.

Korsgaard discusses the implications of her view for several practical concerns, including self-interested concern and compensation. But again,

[3] The subtle shift of focus from actions and choices to beliefs and desires in this paragraph is not my mistake in interpreting Korsgaard. I believe that it can be explained by the fact that Korsgaard states that one can view certain mental states as forms of action. See Korsgaard (1989, 103).

our question is whether the criterion purports to define numerical identity. It seems that if authorial connectedness is the identity-constituting attitude, it marks our boundaries and is a condition for our persistence. The mad surgeon case supports this claim, for it shows that Korsgaard believes that if person does not authorize anticipated experiences, those experiences are not her own. This means, according to (SI), that they are not part of her identity, and, according to (SIP), she cannot be numerically identical to the subject of the experiences. In contrast, a person can survive even drastic psychological changes as long as these are the product of the person's decisions, that is, authorized (cf. Korsgaard 1989, 123).

2.3. Identity as a choice

My final example of a subjective theory of personal identity is a theory developed by Carol Rovane (see Rovane 2009).

The central claim that Rovane defends is that the existence of a person is never a metaphysical or a biological given but is always bound up with the exercise of effort and will (cf. Rovane 2009, 96). To argue for the claim, Rovane first explains what she means by the concept of person.

Persons, according to Rovane, are subjects with the capacity for *rational agency* – they are able to deliberate about the reasons for action and to present reasons to others and thus influence their decisions. Further, the ability to respond to reasons constrains persons' behaviour in accordance with the *normative requirements of rationality*. These require of persons that they arrive at an act on the basis of an all-things-considered judgment about what it would be best to do in the light of all their beliefs, desires and attitudes. This requires that persons resolve contradictions among their beliefs, work out their implications, and rank their preferences in transitive order. In sum, persons must strive to achieve what Rovane calls *an overall rational unity* within themselves (cf. Rovane 2009, 105).

According to Rovane, the normative requirement to achieve overall rational unity makes an implicit reference to personal identity, because it defines what it is for an *individual person* to be fully or ideally rational. After all, we do not consider it a failure of rationality if several people have incompatible preferences; we only do so if one does (cf. Rovane 2009, 105). Thus, one can approach the issue of personal identity by considering under what conditions a commitment to meeting the normative requirements of rationality arises, because "[t]his is the condition in which we have a *person*

in the sense that goes together with the ethical criterion of personhood" (Rovane 2009, 105).

Unfortunately, Rovane does not explicitly state her criterion of personhood, so we must rely on an interpretation of her claims. Rovane says that the existence of a person is bound up with the exercise of effort and will. The effort and will seem to be related to the person's activity of unifying her mental states into a coherent and consistent set. Thus, Rovane is committed to the following thesis:

(EP) Necessarily, for any x, x is a person if and only if x seeks an overall rational unity within the set of her mental states.

Rovane's criterion of personhood has interesting implications. Rovane claims that the commitment to achieving rational unity can also transcend the boundaries of a single human being.

> [Human beings] can exercise their rational capacities together so as to achieve rational unity within groups that are larger than a single human being, and they can exercise their rational capacities in more restricted ways so as to achieve rational unity within parts that are smaller than a single human being. (Rovane 2009, 106)

In other words, there may be *group persons*, comprising several human beings, and *multiple persons* within a single human being.

The concept of *multiple persons* is illustrated by an imaginary situation in which we visit a friend at a company headquarters and see that our friend has "become" a bureaucrat who cannot recognize the demands of friendship (cf. Rovane 2009, 112). According to Rovane, his life seems to take up less than the whole human being and the rest of it literally belongs to the life of the corporation. Importantly, Rovane states that "this may not be mere 'role playing'. This may be, literally, a fragmentation of the human being into relatively independent spheres of rational activity, with separate rational points of view that can be separately engaged" (Rovane 2009, 112). Rovane elaborates that we typically try to live our lives in rationally unified ways for the sake of more specific projects, such as relationships and careers. She stresses, however, that these are just projects and that they are optional.

> It is possible for human beings to strive for much less rational unity than these projects require and still be striving for rational unity. And,

> sometimes, the result may be relatively independent spheres of rational unity with a significant degree of segregation. (Rovane 2009, 112)

I believe that Rovane's arguments for the claim that personal identity is a matter of choice, rather than a metaphysical or biological given, prove that her theory is another instance of a subjective theory of personal identity. If, according to Rovane, personal identity consists in the commitment to achieving overall rational unity, it presupposes adopting an attitude towards a set of mental states, leading to their adoption or rejection. I believe that criterion (SI) takes the following form in Rovane's theory:

> (SIC) Necessarily, for any x, if x is a person at t and there is a set of mental states and actions M at t*, M will be part of x's identity at t if and only if x at t seeks to bring the mental states and actions in M into rational unity with her other mental states and actions existing at t.

This is, then, how Rovane uses the concept of personal identity. And again, what we want to consider is whether this concept purports to be the concept of numerical identity, implying the persistence conditions of persons. Rovane's theory is less explicit about this, but her comment about the friend-bureaucrat example, according to which it may *literally* involve a fragmentation of the human being into separate rational points of view that can be separately engaged, at least seems to be addressing numerical identity. That is, some of Rovane's claims seem to commit her to the view that if a person at t does not seek a rational unity among the set of mental states she has at t and a set of states existing at t*, she cannot be numerically identical to whoever is subject of the mental states at t*.

I thus conclude my presentation of three subjective theories of identity. In what follows I will assess the plausibility of their claims. For brevity, I will often use the term (*an attitude of) identification* to refer to the mechanism of identity-constitution the theories employ.

3. Interpreting "identity"

We have seen that there is textual evidence that supports the belief that the subjective theories purport to define numerical personal identity. How-

ever, the belief that numerical personal identity could be defined by means of a subjective attitude to a set of mental states or actions leads to grave difficulties.

3.1. The presupposition of numerical identity

The first problem is that the theories presuppose that persons can be defined independently of the criteria that the theories propose. To see that, consider again the general form of the subjective criterion (SI). Obvious counterexamples show that, as such, this statement cannot be true, because a person is limited in the range of mental states and actions that she can plausibly identify with. For example, consider the desire to have a third child, the belief that whales are fish, or the action of executing Saddam Hussein, which may all have occurred at some places on December 30, 2006. Any theory that claims that if I identify with these mental states and actions, they are mine must be seriously mistaken. I never had that desire or belief, and I did not carry out that action, so their incorporation in my narrative, my authorization of them, or my effort to rationally unify them with my current beliefs and actions would not show that they are mine, but rather that I am seriously confused. This shows that I can only reasonably identify with a subset of all the mental states and actions that there are. These are presumably those that I have *objectively* had and carried out. But in that case there must be another criterion of personal identity that will establish which mental states and actions are objectively mine, and only then can I legitimately adopt an evaluative attitude and identify with them. But this further shows that identification cannot make these mental states and actions literally mine and, conversely, if I fail to identify them, they do not really cease to be mine.

This point can be elaborated with respect to the concept of existence. As I have stated, a criterion of numerical identity will imply an answer to the question of when persons begin to exist. If the subjective theories define numerical identity, they are committed to the claim that persons only begin to exist as a result of their identification with their mental states and actions. But if persons do not exist prior to the act of identification, they cannot have any mental states, let alone identify with them. Conversely, persons already have to exist and have mental states in order to be able to identify with them.

In a later work, Korsgaard takes up this "paradox of self-constitution" and argues that it is not a paradox at all (see Korsgaard 2009). Cases of self-

constitution are common in the natural world and there is no mystery about them. A giraffe's nutritive processes turn food into matter that replaces tissue that is in need of renewal. Thus, a giraffe constitutes itself by its own activity. "Being a giraffe is *doing* something: a giraffe is, quite essentially, an entity that is always *making* herself into a giraffe" (Korsgaard 2009, 36). Similarly, persons can constitute themselves by their own activity.

But it is questionable whether this analogy dissolves the mystery. The case shows that living systems can *maintain* themselves once they are alive. It does not show that they can *bring themselves into existence* by their own activity. A giraffe does feed itself, but a giraffe embryo does not start its own existence through its activity. But if the theories purport to address numerical identity, they must explain how persons begin to exist, not just how they maintain their existence. And it is clear that persons cannot begin to exist by identifying with their mental states and actions.

3.2. A paradox of identification

Another argument against the claim that personal identity is the result of a person's identification with certain mental states and actions is that it leads to a paradox. We have seen that if subjective theories purport to address numerical identity, they are committed to (SI) as well as (SIP). Taken together, these principles say:

(SISIP) A person x at t is identical to someone y at t* if and only if x at t identifies with the mental states and actions of the y at t*.

The problem is that identification is not a symmetrical relation. As a result, the person at t may identify with the anticipated mental states and actions of a person at a t*, but the person at t* may fail to identify with the mental states and actions of the person at t. This supposition generates a paradox: the person at t is identical with the person at t*, while the person at t* is not identical with the person at t.[4]

[4] If subjective theories purport to define numerical identity, they also share a host of problems with psychological theories of personal identity, such as *the fetus problem* and *the thinking animal problem.* See Olson (2008).

3.3. The wrong interpretation?

The above arguments show that even though some claims by the authors reveal their intention to define numerical personal identity, this project cannot succeed. But perhaps my interpretation is mistaken. Perhaps when Rovane says that a human being can literally be fragmented into different persons, she does not mean *literally*. Perhaps when Schechtman states that our claims that a person who has been the victim of abuse is "no longer the same person" or "is gone" "need not be considered entirely metaphorical" (Schechtman 1996, 88), she is not implying they are to be taken *literally*. Taken at face value, these claims are about numerical identity. But, in any case, the authors also make claims that seem to contradict the above claims. Rovane, for instance, writes:

> ... we needn't infer that personal identity is distinct from human identity, in the sense that there is a distinct thing – the person – whose life is shorter than a given human being's life. We can suppose instead that personhood is a *status* that is sometimes achieved by a given human being and sometimes not, without introducing any distinct existence. (Rovane 2009, 101)

In a later work, Schechtman, describing several cases in which a person changes so much that she can no longer identify with her past mental states, comments:

> ... we might say that she has become a different person, but there is some sense in which we clearly do not mean it. The change is only remarkable *because* she also remains the same person. (Schechtman 2001, 98)

As a result of these confusing remarks, it is quite unclear whether or not numerical identity is the focus of these theories. But since the numerical interpretation leads to obvious difficulties, we should try to find a more plausible interpretation.

What could be a more plausible interpretation? I believe we can propose an interpretation according to which the theories attempt to define what could be called *practical identity*. All of the authors point out that the concept of identity they develop is deeply related to our practical lives. Schechtman proposes the notion in her attempt to analyse, besides survival, the concepts of responsibility, self-concern, and compensation. Both Kors-

gaard and Rovane maintain that the concept is closely tied to the notion of agency, which is crucial in our conception of ourselves as moral beings (e.g., Korsgaard 1989, 132). I believe that we could find an interpretation that emphasises the practical dimension of persons while it does not entail any claims about persons' existence and persistence.

In this interpretation the basic entities are human beings, whose existence and persistence is determined by a criterion of numerical identity. (It does not matter now which criterion it is). The concept of personhood refers to a status or a role that a human being may or may not assume. When a human being assumes this role, no new entity begins to exist; the human being merely becomes a person, that is, acquires an important cluster of properties which make it an appropriate target of our practical concerns. And this happens, according to the subjective theories, when the human being identifies with its mental states.

This interpretation deals with the paradox of self-constitution discussed above. Persons do not bring *themselves* into being, but are brought into being when human beings identify with their mental states. The act of identification in which persons begin to exist does not presuppose the existence of persons. It only presupposes the existence of human beings that have mental states and capacities to identify with them.

Further, on this interpretation, the concept of personal identity refers to the unity of this role. That is, it refers to a relation which has to hold in the life of a human being in order for the human being to be the same person in time. If the human being is the same person in time, we may legitimately attribute responsibility to it for past actions, compensate it for past harms, and it may legitimately express self-concern for its future well-being. And on this interpretation of the subjective theories, one would be the same person in time as long as one identified with roughly the same mental states and actions.

This interpretation enables us to provide benign paraphrases to the troubling implications of the numerical interpretation of subjective theories. For example, saying of someone who does not identify with her past mental states that "she is no longer the same person" does not mean that an entity has ceased to exist and a new one has arisen. It simply means that the human being that she is has begun to exemplify a different set of practically relevant characteristics. Saying that the past person has "not survived" simply means that the human being no longer has the set of practi-

cally relevant characteristics that she used to have, and that, as a result, we may not legitimately blame her for her past actions or needn't compensate her for past harms. On this interpretation, using the vocabulary of existence and persistence is merely a misleading way of speaking about (continuing) property exemplification. It is as misleading as saying of a president of a corporation who has become the president of a country that "she is no longer the same president" or "the president we knew is gone".

It is worth emphasising how this interpretation differs from the numerical one. The major difference is in that, on the numerical interpretation, if person P at t is not identical to person Q at t*, there are two entities. But on the practical interpretation P and Q denote complex properties which may be exemplified in time by a single entity, a single human being. And saying of a single human being that it is no longer the same person as it used to be only means that the practical concerns and attitudes that were legitimate with respect to the former one do not carry over to the latter one.

In what follows I would like to show that even if this interpretation is logically and ontologically more innocent, its practical implications are still extremely implausible.

4. Identification and practical identity

As I have indicated, the subjective theories are committed to the claim that being the same person amounts to identifying with the same set of mental states. At the same time, being the same person is a necessary condition for the legitimacy of the identity-related practical concerns. However, I will attempt to show that identifying with the same set of mental states is not a necessary condition for the legitimacy of the concerns. I contend that identification is only relevant for some *aspects* of *some* of our identity-related practical concerns, while others are not affected by it.

Let me first argue that identification is not a necessary condition for the legitimate application of several practical concerns. Rovane's example of multiple persons in one body is a particularly suitable example to illustrate this. As we have seen, Rovane claims there may be separate spheres of rational unity within one body, resulting in the existence of multiple persons. A single human being can be a friend (person 1), while on other occasions

she can be a bureaucrat who cannot recognize the demands of friendship (person 2). If personal identity in subjective theories is practical identity, we are led to the conclusion that a single human being can have multiple practical identities, that is, multiple roles relevant for the attribution of responsibility, compensation and legitimacy of self-concern.

However, this supposition is hard to believe. Suppose that my friend makes a promise to me. Are we really ready to accept that once he enters the headquarters and becomes a bureaucrat who cannot recognize the demands of friendship, it is not legitimate for me to insist that he keep the promise? Or suppose that the bureaucrat embezzles a large sum of money. Would the police really be unjustified in arresting my friend? Would we be willing to accept the friend's excuse that he is not responsible for the bureaucrat's actions because he is not seeking rational unity among his mental states and the mental states of the bureaucrat? Or take another example. Suppose the bureaucrat is compensated for work-related health problems. Would we really consider it a theft if my friend enjoyed the benefits resulting from the compensation? I doubt that we are ready to accept these claims. And since we are not, it shows that we do not think that the fact that a person does not identify with certain mental states or actions means that we may not legitimately compensate her or hold her responsible with respect to those mental states or actions.

But Rovane's example may be an easy target, because it is very difficult to take seriously the idea that one human being could really change her practical identities by walking into and out of her office. Also, Rovane's example seems to be special in that it is meant to be an example of a human being alternately assuming the role of two persons. The examples offered by Schechtman and Korsgaard seem to suggest, rather, the idea that a human being may forever cease to have the status of one person and, instead, assume the status of another. In such cases it may seem more plausible that our practical concerns directed at the former person may no longer apply to the latter one.

Even so, I would like to insist that these cases still fail to illustrate a situation in which a human being's practical identity changes as a result of her non-identification with her past mental states. To see this, it will be useful to consider a few more examples to stimulate our intuitions.

Schechtman gives an example of a carefree young woman who eventually settles down and becomes a serious matron. Schechtman describes her

as someone who can remember her wild days, but who cannot recapture the emotions and desires she once had. She is someone who fails to see how she could have made the choices she made and who is completely alienated from the past reasons that motivated the choices (cf. Schechtman 2001, 101). In the terminology we have been using, she cannot identify with her past mental states.

Korsgaard suggests that a person whose mental life has been changed by external forces in a way that has not been authorized by that person leads to a loss of identity. Her example of a mad surgeon is a far-fetched thought experiment, which may not satisfy those who are sceptical about using such examples as evidence. But some actual cases come close to this hypothetical scenario. A case that comes to mind is the life story of Phineas Gage. Gage was a construction worker who suffered a serious brain injury when a metal rod was driven through his skull, resulting in profound changes in his pro-social behaviour. Even though the extent of the mental changes in this case is controversial, what is important is that these changes were not self-induced. Let us suppose, without any pretence of historical accuracy, that a complete lack of identification with his past occurred after the accident. Let us now consider what our practical attitudes in these cases should be.

Take first the notion of responsibility. If the serious matron fails to identify with the mental states of the carefree young woman, then, according to the subjective theories, we should be inclined to say that she cannot be responsible for any acts the young woman carried out. But suppose that the carefree woman badly hurt the feelings of a good friend of hers and they now meet after many years. It seems quite obvious to me that an apology is appropriate and rightly expected and, therefore, that the matron is still responsible for the act. But if, as a result of her non-identification with the young woman, the matron is a different person, as subjective theories say, any expectation of apology is unwarranted. Even if the serious matron is the same human being as the carefree woman, they are different for the purposes of attribution of responsibility, so there should be no reason for the matron to deal with the consequences of the young woman's reckless behaviour. I find this implication of subjective theories hard to accept.

But it may be objected that I have ignored the fact that responsibility comes in degrees. Some actions can be attributed to an individual in the minimal sense that they occur in her history and it is true that an act of non-identification cannot erase them. The individual is responsible for

them in the sense that the actions are still hers. That fact, however, shows very little about the extent to which she should currently be blamed – and this aspect of responsibility *is* determined by the degree of the current person's identification with the acts. After all, the matron is mentally so different that she may not even understand the reasons that led to the past act, let alone identify with that act. She does not deserve the same degree of blame as someone who truly identifies with an evil act.

This argument has some force. It does seem that at least *the degree of* responsibility interpreted as blameworthiness depends on whether or not the blamed subject identifies with the act for which she is blamed. After all, we recognize this distinction in our different attitudes towards the online murders committed by ISIS as opposed to cases of unintentional manslaughter in traffic accidents, for instance. But responsibility is not the only practical concern that attaches to personal identity and I would like to show that non-identification does not play a role in other such concerns.

Take self-concern, for instance, and consider the idea that a lack of identification justifies a corresponding lack of self-concern. Suppose I do not identify with the person who will be in my body in 30 years' time, because I am a *bon vivant* who lives by the motto of *carpe diem*. Suppose, further, that due to my love of food, alcohol and tobacco I am badly damaging the health of that person. It does not seem obviously true to me that the fact that I do not identify with the future person who I (as a human being) will become makes it justifiable for me to ignore her well-being. But then, self-concern is a practical concern that subjective theories fail to fully account for, because identification is not necessary the appropriateness of self-concern.

Let us now turn to compensation, which both Korsgaard and Schechtman discuss. Schechtman, for instance, focuses on the question of *what constitutes adequate compensation* for past harm (cf. Schechtman 1996, 157), and she argues, convincingly, that the answer does depend on the degree to which one identifies with the mental states and actions that compensation affects. If I take myself to be a football-despising opera lover and someone compensates me for a past insult with Premier League tickets, I am not likely to consider that adequate. The more closely compensation targets mental states I identify with, the more of a compensation it is for me.

But there is a more fundamental question relating to compensation. It is the question of *when compensation is legitimate* in the first place. It does

not seem identification has any say here. We believe that a person is legitimately compensated for a past harm only if she is the same person as the one to whom the harm was done. If personal identity was analysed by means of identification, it would follow that a person is only a legitimate target of compensation if the person identifies with the mental states of the person to whom harm was done. But then, if we are supposing that the changes resulting from Gage's injury could have been so severe as to lead to his non-identification with his previous desires, actions, beliefs and decisions, we must conclude that he does not deserve any compensation, because, while being the same human being, after the accident he was a different person. If the reader finds this conclusion hard to accept, as I do, it may be because we actually believe that the legitimacy of compensation does not presuppose identification. Thus, there is another aspect of our practical identity that cannot be captured by subjective theories.

5. Conclusion

Personal identity is both a metaphysical and a practical concept. It provides the existence and persistence conditions for human persons, and it is presupposed in our attributions of responsibility, compensation, expressions of self-concern, and other everyday practical attitudes and concerns. Defining identity by means of an attitude towards mental states and actions fails to provide existence and persistence conditions for human persons. In several instances, it also fails to account for our intuitions as to when the identity-related practical concerns and attitudes are appropriate. Thus, subjective theories of personal identity fail in both of the interpretations and owe us a clear explanation of what concept of personal identity they purport to define.[5]

[5] I would like to thank the Czech Science Foundation (GAČR) for supporting my research by grant no. 13-00624P. This paper was also supported by the Faculty of Arts, Masaryk University. I am also indebted to Tomáš Hříbek and Eric Olson for their comments to drafts of this paper.

References

Bělohrad, R. (2014): Can We Do Without a Metaphysical Theory of Personal Identity in Practice? *Prolegomena* 13, No. 2, 315-334.

Huemer, M. (2005): *Ethical Intuitionism.* Houndmills (UK): Palgrave Macmillan.

Korsgaard, C.M. (1989): Personal Identity and the Unity of Agency: A Kantian Response to Parfit. *Philosophy & Public Affairs* 18, 101-132.

Korsgaard, C.M. (2009): *Self-Constitution: Agency, Identity, and Integrity.* Oxford: Oxford University Press.

Olson, E. (2008): Was I Ever a Fetus? (New Version). In: Nichols, R. – Smith, N.D. – Miller, F. (eds.): *Philosophy through Science Fiction.* New York: Routledge, 339-348.

Rovane, C. (2009): Personal Identity and Choice. In: Mathews, D.J.H. – Bok, H. – Rabins, P.V. (eds.): *Personal Identity and Fractured Selves: Perspectives from Philosophy, Ethics, and Neuroscience.* Baltimore: John Hopkins University Press, 93-128.

Schechtman, M. (1996): *The Constitution of Selves.* Ithaca – London: Cornell University Press.

Schechtman, M. (2001): Emphatic Access: The Missing Ingredient in Personal Identity. *Philosophical Explorations* 4, No. 2, 95-111.

Shoemaker, D. (1996): Theoretical Persons and Practical Agents. *Philosophy & Public Affairs* 25, 318-332.

Organon F 22 (3) 2015: 302-324

Structural Realism without Metaphysics: Notes on Carnap's Measured Pragmatic Structural Realism

MAJID DAVOODY BENI

Philosophy of science department. AmirKabir University of Technology
No. 424 Hafez Street. P.O. Box 3313-15875. Tehran. Iran
m.davoody@aut.ac.ir

RECEIVED: 23-02-2015 • ACCEPTED: 30-05-2015

ABSTRACT: Carnap's reinvention of the Ramsey-sentence approach to scientific theories has been at the center of a new debate in recent years. Following Grover Maxwell, Psillos (2000a) argued that Carnap's re-invention of the Ramsey-sentence had failed to lead to the desired neutral stance in the realism-instrumentalism debate, and ended, instead, to a form of structural realism which happened to be liable to Newman's objection to Russell's version of structural realism. The objection held that without putting suitable restrictions on the range of the variables of the Ramsey-sentence, a Ramsey-sentence approach to theories renders trivial and a priori true all ontological commitments to unobservable entities issued by scientific theories. By arguing that Carnap achieved the neutral stance, Friedman (2011) counter Psillos claim. He denied that any form of realism could be attributed to Carnap. In this paper, I provide a middle ground, where an unorthodox form of structural realism could be attributed to Carnap. I highlight parts of Carnap's work which deal with the problem of designation of abstract terms and the relation of the language to the facts of the matter (in Carnap 1934; 1950; 1966), to argue that it was Carnap's view about the practical methodological considerations, being at work in the construction (or choice) of the linguistic systems, which led him to the unorthodox form of structural realism. I also claim that the same practical considerations constitute the nub of a viable Carnapian answer to Newman's objection.

KEYWORDS: Carnap – conventionalism – structural realism – metaphysics – Newman's objection – pragmatics – Ramsey-sentence approach – semantics – truth.

0. Introduction[1]

Carnap's reinvention of the Ramsey-sentence approach to scientific theories has been in the centre of an interesting debate in recent years. The credit of bringing back the subject to the foreground goes to Stathis Psillos (1999, 2000a, 2000b). While Psillos' work gave rise to a number of studies and assessments (e.g. Creath 2012, Cruse 2005, and Demopolous 2008), it finally fell to Friedman (2011) to make an attempt for answering Psillos' challenge.

Psillos' (2000a) claim was that Carnap's re-invention of the Ramsey-sentence had failed to result in the desired neutral stance in the realism-instrumentalism debate, and led, instead, to a form of *structural realism*, which happened to be liable to Newman's objection (which had been originally aimed) to Russell's version of structural realism. The objection held that without putting suitable restrictions on the range of the variables of the Ramsey-sentence, a Ramsey-sentence approach to theories renders trivial and a priori true all ontological commitments to unobservable entities issued by scientific theories (see Psillos 2000a, 254).

Friedman countered Psillos' view by arguing that Carnap's conception of a scientific theory, as the conjunction of its Ramsey-sentence and Carnap-sentence, had indeed resulted in the desired neutral position (see Friedman 2011). Consequently, Friedman claimed that Newman's objection, raised in the context of the recent debates about the structural realism, is no problem for the Carnapian metaphysically-neutral structuralism (cf. Friedman 2011).

My aim is to find a middle ground, to state that Carnap's structuralism is accompanied by an unorthodox but fulfilling form of realism, which rests on the functioning of *the practical methodological considerations*. These considerations were contrived to work at the basic level of the construction (or choice of the rules) of the linguistic systems. I will develop this to suggest that the Carnapian stance is an elaborated extension of realism, because, in spite of Friedman's (2011) discord, at least as far as the *physical systems* are concerned, there are indeed some robust *factuality-conducive referential links*

[1] For coming to the final version of this paper I am indebted to Richard Creath, Steven Elliot, Hassan Khodawerdian, Aboturab Yaghmai, and the two anonymous referees of *Organon F*. All of these debts are gratefully acknowledged.

in the Carnapian system. These links are forged by the pragmatic-practical factors, and they subtly prevail between the variables of Carnap's structures to connect them to the facts of the matter. The referential links, therefore, are soiled with some pragmatic taint, and the semantical relations at the root of Carnap's irenic[2] form of structural realism are pragmatically enriched.

The referential connections are settled pragmatically and methodologically. Therefore, interestingly enough, Carnap's realist stance is not at odds with the metaphysical neutralism which Friedman has underlined in his interpretation. It is not resting on the standard semantics of metaphysical realism. Nor does it hinge on some contentious metaphysical arguments such as No Miracle Argument (NMA) and Inference to the Best Explanation (IBE)

But to obtain its full-legitimacy, Carnap's structural realism needs to survive the Newman's challenge. Friedman's answer to Newman's objection came in terms of depriving Carnap's approach from any commitment to the factual or synthetic content of the existentialised terms of the theory, beyond what is conveyed by their empirical adequacy. My response is developed in a different direction, and it indicates that the problem of finding the appropriate structures, or equivalently, setting restriction on the sets of the existentialised variables of the Ramsey sentences, could be sorted out plausibly enough, in a pragmatic, rather than in a syntactic or semantic way. This answer to Newman's objection is an extension of the moderate pragmatic realism which I read into Carnap's anti-metaphysical structural realism.

1. Carnap's empirical structural realism

Existence of an unfathomable chasm between observational and theoretical domains in Carnap's conception of scientific theories has been questioned by Creath (1985, 2012). By arguing that the ontological commitment, which has been readily made with regard to the observational entities, could be extended into the adjacent (and in Creath's view, entwined)

[2] The term has been coined by Creath (1985), who assigned a form of irenic realism to Carnap.

domain of theoretical sentences, he stated that Carnap's philosophy is prone to be understood as a subtle form of (irenic) realism. Although Creath's interpretation is persuasive, neither Psillos nor Friedman accepted that Carnap's endeavour had led to a solemn form of realism. My construal of Carnap's reinvention of Ramsey-sentence approach was spelled out to meet Psillos and Freidman's recent non-realist interpretations.

Creath's remark about the blurredness of observational-theoretical border notwithstanding, it is customary to assume that the received view of theories indicates that there is a distinction between these two parts of the language of science. The Ramsey-sentence approach to theories (developed by Ramsey 1929), had been supposed to be efficient in dealing with the troublesome theoretical parts (i.e., conveyed by secondary terms, in Ramsey's terminology), via explaining their meaningfulness solely in virtue of their connection to the observational domain.[3] It was received as an empiri-

[3] There already exists a prosperous literature formed around Ramsey-sentence approach and Carnap's innovation. All necessary technical details are articulated in the mentioned works (Ramsey 1929 and Carnap 1966 besides, particularly Psillos 2000a; 2000b; 2006 and Creath 2012 and Cruse 2005 seem to be the most instructive ones). Therefore I am only spelling out Ramsey-sentence approach and Carnap's innovation as briefly as possible. Let us take TC as our theory, holding theoretical (T_i) and observational (O_j) terms and postulates (in Ramsey 1929, secondary and primary terms respectively). This is the standard (received) form of a theory:

$$TC(\ldots T_1 \ldots O_1 \ldots T_2 \ldots O_2 \ldots T_n \ldots O_m \ldots)$$

where T_n are theoretical predicate constants and O_m are observational predicate constants. From this we can derive the Ramsey-sentence R(TC):

$$R(TC): (\exists U_1), \ldots, (\exists U_n)TC(U_1, \ldots, U_n; O_1, \ldots, O_m)$$

In R(TC) the observational terms (O_j) are preserved, and the theoretical constants (T_i) which occur in TC are replaced by distinct higher-order predicate variables (U_k) which do not occur in TC, and then the variables are prefixed by existential quantifiers. This is the realized form of the theory, because, according to Ramsey, R(TC) would be observationally equivalent to TC, and would preserve the empirical content of the theory, (that is $T(C)\rightarrow O$ if and only if $R(TC))\rightarrow O$). And according to Carnap R(TC) would be *semantically* equivalent (L-equivalent) to TC. That is, (speaking in model-theoretic terms) if there exist a class of entities which satisfy the Ramsey-sentence, then there is a denotation between theoretical terms (T_i) and the class members. Carnap-sentence of the theory namely ($R(TC) \supset TC$) works as an analytic part of the reformulation of the theory to provide the necessary interpretation of the theoretical terms to the necessary extent.

cist solution for the problem of the meaning of the theoretical terms. Carnap reinvented the approach in mid-1950s (the story of reinvention has been mentioned in Carnap 1963 and Psillos 2000a) and made some clarifications about it in mid-1960s (cf. Carnap 1963; 1966).

It was with regard to this chapter of the history of empiricism, and after examining the Carnapian differentiation between (with regard to the borders of linguistic frameworks) internal and external questions (elaborated in Carnap 1950; 1956), that Psillos asked "why isn't Carnap's position realist enough?" (2000a, 256). It is true that Carnap had declared that any question concerning the reality of the system of entities as a whole, is an external (or metaphysical), and hence an illegitimate (*pseudo)*-question. But questions could be asked about the reality of particular entities, questions which were raised and answered after the acceptance of a certain Linguistic Framework (LF). These were internal questions, which their answer might be found, legitimately enough, by either purely logical or purely empirical methods, depending on whether the framework is a logical or an empirical one (cf. Carnap 1950).

The looseness in fixing LFs in a cognitively meaningful and theoretical (i.e. logical) way makes the approach inapt for being considered as a form of orthodox scientific realism which is based upon the watertight semantics of correspondence theory. But it could be construed as a limited or internal form of realism all the same. This eccentric form has been traced back by Psillos to Feigl's (1943; 1950) "empirical" or "semantic realism", which held that scientific theories imply commitments to unobservable entities no less than to observable ones. The claim is, of course, empirical (in Feigl's sense) rather than metaphysical" (Psillos 2000a, 257). This much could be conceded to almost unarguably.

But Psillos went even further. He claimed that Carnap's empirical realism had been taking some *structuralist* turn, in the course of Carnap's reinvention of the Ramsey-sentence approach. I explained Ramsey's approach in the previous endnote. Carnap followed the same track in his "The Methodological Character of Theoretical Concepts" (see Carnap 1956): The language of science was supposed to be divided into two sub-languages. The observational language L_O which is completely interpreted (in virtue of referring to observable domain) and L_T, whose vocabulary V_T consists of theoretical terms. Carnap's Ramsey-wise move, which was evolved at first independently (and in ignorance of Ramsey's achievement)

by Carnap, was to suggest that the vocabulary of V_T could be conceived as ranging over the class of natural numbers which are representing mathematical, rather than theoretical, entities. To make the mathematical parts adequate for the representation of the physical concepts, some C-postulates had been contrived to connect the theory, which was presented as exemplifying certain logico-mathematical structure, to the observable world. It was how the scope of Carnap's structural realism was spread.

It is a historical fact that Carnap's reinvention of Ramsey-sentence approach had been subjected to criticism from the very beginning. The objections were raised to challenge the aptness of the representational (or referential) function of the logical structure, from two opposite fronts. It has been argued that concerning the existentialised variables of the Ramsey-sentence, either they serve their purpose and inferentially refer to the theoretical entities, and therefore do not undertake any fewer ontological commitments than the original theory (as remarked by Hempel 1958), or they refer to nothing beyond the abstract set-theoretic mathematical notations which conveys them, and therefore the approach would lead to a form of "syntactical positivism" (this was remarked by Feigl 1958). Taking the dilemma in either way, the demise of the Carnapian peculiar form of structural realism would be inexorable: it is doomed to collapse either to the orthodox scientific realism, or to syntactical positivism, which strives to stay limited to formal notations, without taking the risk of assigning semantical interpretation to the formulas.

More recently, Psillos and Friedman raised similar issues. Although Psillos did not regard Carnap as an advocate of the orthodox scientific realism, he argued for the necessity of reducing Carnap's version of structural realism into traditional scientific realism. Friedman, on the other hand, maintained that the theoretical parts of language of science could not be interpreted in terms of the standard Tarskian semantics.

For structural realism, to overcome such qualms and become a significant metaphysical and ontological thesis, theories should be primarily conceived as abstract mathematical structures, and then, by application of a semantics which permits the interpretation of the theoretical parts, the main ontological commitments have to be undertaken primarily with regard to these structures. In more precise words, within the context of the received view of the theories, "within which a theory is taken to be a set of sentences, realism amounts to the commitment to standard (correspon-

dence) referential semantics, and to truth, for the whole theory" (Ladyman 1998, 416). Making unswerving ontological commitments to existence of unobservable phenomena (being recognized as structures in ontic structural realism, or theoretical entities in traditional scientific realism) seems to be at the heart of the standard understanding of realism. And it is an unchallenged presupposition that in the syntactic period, this metaphysical realism (MR) ran through the semantical machinery of the correspondence theory (hereafter CT).

But Carnap's philosophy is bereft of any such semantical and metaphysical compartments. He was allegedly unwilling to appeal to the standard correspondence semantics to assert that the theoretical sentences refer to the unobservable entities or structures of extra-linguistic domain.

2. The factuality-conducive referential link

The *referential relation* between the structures of the existentialised variables and their referents in the extra-linguistic domains was supposed to be formed by application of CT. But Carnap's anti-metaphysical agenda was urging him to be reluctant to undertake any such ontological commitment with regard to unobservable entities or concede to the standard referential semantics. Scientific realism rests on the standard referential semantics, and the requisite referential links could not be forged within the framework of this limited realism. According to Friedman (2011), in absence of *a direct referential link* between theoretical terms and unobservable physical phenomena, we should "keep firmly in mind the fact that theoretical terms, for Carnap, are semantically uninterpreted: we assign no designata to them in our semantical meta-language, and so Tarskian semantics (as Carnap understands it) literally assigns no truth-values at all to purely theoretical sentences" (Friedman 2011, 256). This is the most serious piece of evidence that Friedman has offered in the way of ruling out the viability of the realist interpretation of Carnap's structuralism.

So, there is an essential question that the advocate of the realist interpretation of Carnap's structuralism should answer: abiding by the limitations of Carnap's internal realism, how the factuality-conducive referential links could be established between the existentialised variables and their referents in the extra-linguistic domains. If it could be shown that Carnap

had contrived the appropriate means for making a theory of factual reference, albeit without giving way to MR or CT, then it would be easy to argue that Carnap's enterprise does not collapse into a version of strict empiricism or even a neutral stance with regard to realism-instrumentalism debate, but would instead lead to an interesting and elegant though unconventional version of structural realism, provided that we could accept that realism does not necessarily amount to the commitment to standard (correspondence) referential semantics.

The problem on the way of establishment of the referential links is that, as Friedman remarked, the theory of "factual reference", which had been assumed to link the theoretical terms to their unobservable referents (as CT demands), has been replaced in Carnap's thought by the question of which form of language we should prefer – and prefer for "purely pragmatic or practical rather than theoretical reasons" (Friedman 2011, 257). This connotes that purely pragmatic reasons do not count as justifications, or at least as *epistemically viable justifications*, in accounting for the choice of the realist stance which conveys the referential links, and they could not be used in construction of a theory of factual reference. Let's see why.

2.1. Carnap's conventionalism

From the early 1930s onward, in his so called syntactical episode, conventionalism about language and logic has been the kernel of Carnap's thought (see Carnap 1934, §17, which contains Carnap's famous principle of tolerance[4]). The conventional elements have survived the semantical turn and were transferred to Carnap's studies about the nature of truth and semantical relations, reference and designation (in 1940-50s). Conventionalism was, therefore, the enduring essence of Carnap's philosophy. For example in "Empiricism, Semantics, and Ontology" (see Carnap 1950), Carnap continued the same conventionalist vein to suggest that "the question of the admissibility of entities of a certain type or of abstract entities in general as designata is reduced to the question of the ac-

[4] "In logic, there are no morals. Everyone is at liberty to build up his own logic, i.e. his own form of language, as he wishes. All that is required of him is that, if he wishes to discuss it, he must state his methods clearly, and give syntactical rules instead of philosophical arguments" (Carnap [1934] 1937, 52).

ceptability of the linguistic framework for those entities" (Carnap 1950, 92). That is, the question of the designation of, say, the theoretical term "electron", depends on accepting LF of modern physics. The question of acceptance of the LF and its constitutive rules, on the other hand, is not an authentic logico-philosophical problem, but a matter of convention and hence, at most, a practical question of expedience. Notice that this was precisely the context in which Carnap reinvented the Ramsey-sentence approach. Because in Carnap (1956), along with all of the technical elaborations, he kept up to speak in terms of the distinction between the inside and outside of frameworks (as had been initiated in Carnap 1950), to state that there are two kinds of existential questions and two senses of "real".

As Carnap remarked in Carnap (1956), you can accept the reality of an event, or assert the truth of the statement which describes it, only after acceptance of the general logical system, or a body of rules and postulates which rule over the theory which conveys it. But as the postulates and rules do not yield themselves easily to semantical interpretation, the question concerning the existence of the general system of entities should be taken as a question of framework principle. It is true that, as Carnap declared, "for an observer to 'accept' the postulates of T means here not simply to take T as an uninterpreted calculus, but to use T together with specified rules of correspondence C for guiding his expectations by deriving predictions about future observable events from observed events with the help of T and C" (Carnap 1956, 45). But the rules of correspondence work as parts of the inductive systematization to organize and interpret the theoretical expressions in accordance with the observational outcome. As Friedman (2011, 258) has remarked, this does not mean that there is any referential (correspondence) semantics at work in connecting the formal structures to the unobservable events and structures of the world. The rules and postulates of the system are generally contrived in a conventional and arbitrary manner, to (arbitrarily) assign a sequence of semantical values to theoretical terms so that the general outcome of the theory could obtain its empirical adequacy.

Therefore, Carnap's semantics is incapable of offering any ontological indications about the existence of unobservable entities or the modal relations between them, of the kind that the metaphysical realist expects of CT to contribute.

2.2. The theory of factual reference and the theoretical sentences

Carnap's stress on the role of pragmatic factors in latching the logical system into the objectivity of the factual world tends to be examined more carefully as an important chapter in the Carnapian studies in recent years (see Mormann 2007, Richardson 2003; 2007 and Uebel 2013 among a few others). The aim of this section is to show how the pragmatic or practical reasons which are at work in preferences of linguistic forms, could play a decisive role in forging the factuality-conducive referential links which, according to Friedman, could not be accounted for theoretically in Carnap's philosophy. Moreover, I will specifically show that the theoretical sentences of language refer on a par with the observational sentences, in a Carnapian system. Finally I will build my argument on this, to conclude that the existence of the theory of factual reference is enough for founding a peculiar form of realism. Let me elaborate.

There are of course more things conveyed in the vast frameworks of Carnap's ocean of logical systems, than are dreamt of in the narrow scope of traditional philosophy. Even so, when it comes to systems which should be used for accommodating the *language of natural sciences*, LFs could not be produced in some arbitrary and whimsical conventional ways. The language of natural sciences should be useful for communication of reports and predictions, and not every arbitrary language is convenient for accomplishing the task. It is the language of sciences which the philosophers of sciences are mostly concerned about. Now, in spite of his profuse conventionalism, as early as in his 1934 book Carnap remarked that:

> The construction of the physical system is not effected in accordance with fixed rules, but by means of conventions. These conventions ... are, however, *not arbitrary*. The choice of them is influenced, in the first place, by certain *practical methodological considerations* (for instance, whether they make for simplicity, expedience, and fruitfulness in certain tasks). This is the case for all conventions, including, for example, definitions. (Carnap [1934] 1937, 320, my emphasis)

And after three decades he still observed that:

> Factual knowledge is necessary in order to decide which kinds of conventions can be carried out without coming into conflict with the facts

> of nature, and various logical structures must be accepted in order to avoid logical inconsistencies. (Carnap 1966, 68)

Thus the objectivity of the referential relations laid at the foundation of the linguistic system (devised for communication about what natural sciences convey) was preserved against the conventional elements, and the factuality has been neatly interwoven into part and parcel of Carnap's conventional approach, via what was called methodological practical considerations in 1934. The choices of these LFs were not dislodged of the factuality of the world of experiences, and the construction of LF did not take place in an unrestrained and arbitrary way. The upshot is that although, as Friedman emphasized, the question of the reality of the theoretical entities has to be reduced to the question of the preference and practical decision about the language of science (cf. Friedman 2011, 250), yet the frameworks were not devoid of factual content, and the designation relations which were established within the framework had been evolved to be factuality-conducive: as these were pragmatic considerations which were appointed to rule over the choice of LFs to vouchsafe the connection to the factual domain, we may conclude that the designation relations and truths which were formed and conveyed within the framework were pragmatically encroached as well, and by the same token, were attached to the facts of the matter.

Let me summarize. It is true that the ontological commitments of Carnap's internal realism are frame-relative. Normally, this may appear to be at odds with the traditional realist position that seeks to establish the objective and theory-independent reality of unobservable entities. As the links which were forged within Carnap's system were not contrived to work as *direct* referential links to channel between theoretical terms and unobservable physical phenomena, it may be claimed that, there were no ordinary semantic rules of designation in Carnap's system. This may represent Carnap's enterprise as fitting within an anti-realist position. But considering the possibility of choosing and constructing physical linguistic systems in a non-arbitrary manner and in consistency with the facts of nature, it could be agreed that the referential links which have been carved out in the Carnapian physical systems were subtly ushered by the objectivity-preserving considerations to carry factual content within them, albeit in a holistic and non-literal manner. I argue that this provide some footing for launching a subtle form of realism. I should emphasize that this is true about the ref-

erential relations of the theoretical sentences, in an equal footing with the designation of the observational statements of the system: the sequence of the semantical values that make the theory come out true from among the values ranged over by the theoretical variables are not assigned more *arbitrarily* than the designations of the observational parts of the language. The explanation is as follows.

In Carnap (1950), in unfolding the philosophical implications of his semantical enterprise, unlike a man who in his everyday life does with qualms many things which are not in accord with the high moral principles he professes on Sundays (or the physicist who is suspicious of theoretical entities and tries to mark a part of the language of science as uninterpreted and uninterpretable), he did not make a difference between abstract and concrete terms of a theory (cf. Carnap 1950, 85). In contrast to such double-dealers, Carnap conceded to the possibility of assigning truth-value to the theoretical sentence on a par with the observational ones, the postulates and rules of inference of his system permitting (i.e. if "electron" was supposed to designate electron according to the rules of designation of the system, see Carnap 1950). Carnap gave a clear and decisive reason for his impartial behaviour: in certain scientific contexts, it seems hardly possible to avoid referring to the abstract entities (the mathematical and theoretical entities involved). Particularly in physics, Carnap declared, it is more difficult (than mathematics) to shun referring to theoretical entities, for the language of physics serves for the communication of reports and predications, and cannot be taken as a mere calculus (cf. Carnap 1950, 85). Thus, by something like an indispensability argument, Carnap came to the conclusion that acceptance of a language referring to the theoretical entities is completely consistent with empiricism and strict scientific thinking. And as the acceptance of the language is guided by the objectivity-preserving factors, the semantical referential links carved out therein do not run against the grain of the factuality of the world of experience.

Carnap's sophisticated and pragmatically contaminated system of semantics is adequately apt for being used in the way of interpretation of the theoretical statements. The designation relations and truth are impartially assignable to theoretical as well as observational statements in interpretation of physical systems. The tradition has it that the technical features of Carnap's structuralist approach have been contrived to explain the meaningfulness of the theoretical statements in virtue of their relation to the ob-

servational counterparts. But when examined against the vaster context of Carnap's unprejudiced semantics, which includes the subtle technicalities, the claim that Tarskian semantics (as Carnap understands it) literally "assigns no truth-values at all to purely theoretical sentences" (Friedman 2009, 256) would appear to be incorrect. Therefore it is only on grounds of this minute point that I argue that Carnap's philosophy slightly bends toward a form of unorthodox realism. This construal could still be challenged: are the presence of merely indirect ontology and the absence of correspondence theory still within the lines of realism? Well, obviously these are not in line of an up-front standard realism. But even in absence of CT, it could still be argued that Carnap's semantics assigns truth-values to theoretical sentences in an equal footing with the observational ones. And this provides the necessary foundations of a subtle form of unorthodox realism. Of course the primary distinction between the orthodox and unorthodox forms of realism is a mere matter of classification rather than argument. But it does not turn the debate to a verbal issue. For, there are historical pieces of evidence and philosophical arguments to be produced to show how this unorthodox form of realism could obtain its legitimacy and plausibility.

Regrettably there is little space for a detailed historical survey of the invention of semantics in hands of Tarski, Carnap, and a few other gifted logicians (for Carnap's account of this history see Carnap 1963, 29-36). To make a long story short, there is no denying that, according to some understanding, Tarski's semantics is a lair to CT. But this does not mean that Tarski's correspondence referential semantics, being constructed around his definition of truth in formal systems, has to be necessarily understood in terms of metaphysical realism. As Tarski himself explicitly acknowledged, "the semantic definition of truth implies nothing regarding the conditions under which a sentence like ... snow is white can be asserted .... Thus we may remain naive realists, critical realists or idealists, empiricists or metaphysicians-whatever we were before. The semantic conception is completely neutral toward all these issues." (Tarski 1944, 362) The truth-value of the sentences, in a Tarski's system, would be decided in the fit between the object-language and meta-language, without giving way to any metaphysical indications about the referents of the statements of the object-language or the ontological state of the meta-language. Neither truth nor the referential relations were articulated in terms of metaphysical realism any more than, say, a pragmatic or deflationary account (for an ex-

tended explanation on this see Wilfrid Hodges' (1985-86) "Truth in a Structure").

Accordingly, even Carnap's investment in Tarski's semantics did not persuade him to add some metaphysical realist flavour to his logic of science. In such circumstances, the notions of pragmatic truth and designation, defined within LFs which were pragmatically picked and formed, could very well play the role of the underlying semantical foundation of the Carnapian form of realism. That is, staying in the metaphysically neutral grounds does not prevent the approach from bringing about realist fruits in philosophy of science. As the theoretical sentences within Carnap's system are capable of conveying truth-values – in terms of Tarski's unfamiliar understanding of Tarski's semantics – Carnap's structuralism is prone to be interpreted in terms of a sophisticated and untraditional form of structural realism.

The untraditional aspect is not by itself a gap in the Carnapian view. Many a peculiar form has been developed in parallel to the orthodox trend of scientific realism. One tends to think that there should be a common essence to these (sometimes remotely) resembling forms of realism, which have all of the properties of the members of an unruly family. But an unshakable loyalty to the standard referential semantics of CT does not seem to be either the essence or the necessary requirement of realism. It is true that, in a realist understanding of the theories, the scientific theories and models should represent the world in one way or another. But as French's (2003) interesting inquiry on the nature of representation shows, the existence of an isomorphic relation (the model-theoretic relative of CT in semantic view) is neither the necessary nor the sufficient condition for the representation of the world within the models.[5]

[5] Perhaps as French suggested, to outline representation in holistic and nonliteral ways, the idea of denotation could be appealed to, as a suitable relation for showing how a model stands for physical system and explaining how theoretical conclusions correspond to the phenomena and decides whether the theory is empirically adequate (see French 2003, 1478). Denotation, embodied in form of partial isomorphic account, is much more flexible and modest than the idea of total isomorphism or linguistic correspondence. To take the discussion back to the context of Carnap's so-called received view, it seems that the loose conventional relation which is pragmatically restricted, is akin enough to French's notion of denotation, to equip Carnap's structuralism with the appropriate means for channelling between theories and the world.

Perhaps we still can concede that after accepting the theory, believing in truth of what the theory says – in whatever imaginative way that the belief in truth may bloom – could be maintained as a handy but provisional characteristic of realism. Mind that the belief in truth of the theory may flourish in quite a number of imaginative ways (see Boyd 1999; Ellis 1988; Fine 1990, French – Saatsi 2006; Hacking 1982; and Quine 1981). But whether truth should or shouldn't be characterized as correspondence with reality (as was taken for granted in the standard scientific realism) is a separate question, which as Horwich (1991) persuasively argued, has a little bearing on the question of realism.

2.3. Purely pragmatic reasons?

There is another significant point to be remarked before going to the next part of the paper. The pragmatic reasons, which play a significant role in loading the system with objectivity-preserving factual elements, are not, in spite of Friedman's remark, "PURELY pragmatic or practical rather than theoretical reasons" (Friedman 2011, 257). If they had been of purely pragmatic nature, then, at least according to the advocates of the orthodox epistemology, they could not assume epistemic roles in stabilizing the foundations of knowledge.[6] It's true that they certainly were not staged to play the role of purely epistemic factors which partake in the cognitive nature. But (at least in Carnap's 1950 and some later works such as his answer to Abraham Kaplan in 1963 Schlipp's volume) this was not taken to mean that they were totally detached from the domain of theoretical justifications and cognitively meaningful expressions either. According to Carnap:

> The decision of accepting the thing-language, although itself not of a cognitive nature, will nevertheless usually be *influenced by theoretical knowledge*, just like any other deliberate decision concerning the acceptance of linguistic or other rules. ... The efficiency, fruitfulness, and simplicity of the use of the thing-language may be among the decisive factors. And the questions concerning these qualities are indeed of a *theoretical nature*. (Carnap 1950, 87, my emphasis)

[6] There are of course the advocates of pragmatic encroachment in epistemology (e.g. Fantl and McGrath, Stanley, Hawthorne, Weatherson), who are arranging a revolt against this dominant orthodox view. But I try to stay in the framework of the orthodox view for the time being.

Therefore, if it is the want of the theoretical touch which keeps the practical reasons back from being considered as viable candidates for vindicating the choice of the realist framework, then by remarking that the influence of the theoretical knowledge on practical considerations is strong enough to boost them to the level of (even epistemologically) plausible justifications, it could be shown that the choice of the realist LF which conveys the factuality-conducive referential links is quite reasonable in spite of not being based on metaphysical speculations. The practical and the theoretical deliberations work together in dealing with the problem of the choice of linguistic frameworks, as Carnap declared some years later (see Carnap 1963, 539).

3. Newman's challenge

Here I attend to Psillos' qualm about the plausibility of Carnap's structural realism. As Psillos' "Choosing the Realist Framework" (2009) implies, he was primarily somewhat interested in the moderate and measured form of realism which had some "pragmatic ring to it" and was "free from metaphysical anxiety".[7] Unlike Friedman, Psillos did not altogether dismiss the aptness of pragmatic reasons for founding an interesting and unorthodox form of realism. But eventually, it turned out that Carnap's irenic position was not realistic enough for Psillos either. For, although it did not give way to a negative form of instrumentalism, it was "not a fully realist position either, since asserting what these entities are is no longer a substantive assertion, but instead it reduces to adopting a meaning postulate" (Psillos 2000a,

[7] Although this paper mostly deals with Feigl's empirical realism, Psillos in short and to the point remarks explains how Carnap's thought is connected to Feigl's endeavor. One of his hints, is so remarkable that we would quote it right in here:

> In fact, in his *Empiricism, Semantics and Ontology*, Carnap (1950, 214) refers the reader to Feigl (1950) piece "for a closely related point of view on these questions [how do we adopt a framework?]". Conversely, in his own defence of semantic realism, Feigl refers the reader to Carnap's (1946, 528), where Carnap says: "I am using here the customary realistic language as it is used in everyday life and in science; this use does not imply acceptance of realism as a metaphysical thesis but only what Feigl calls 'empirical realism'". (Psillos 2009, 308, footnote 4)

270). And this is quite true. For Carnap, in elaborating the technical aspects of Carnap (1956) did indeed assert that only the observational parts of the theory are semantically interpreted in his approach. The semantically uninterpreted theoretical parts are defined implicitly through the postulates of the system. But as I discussed in the previous section, the choice of the meaning postulates (as well as definition and any other kind of convention) could be supplied with some viable pragmatic reasons, to guarantee that they are justified enough to be laid at the foundation of a realist framework (see Carnap 1934). It was how the metaphysical realist semantics of CT had been replaced by pragmatic vindications of methodological naturalism in Carnap's thought. So I have to confess that I feel very tempted to wave away Psillos' objections as relic of some misplaced royalty to the dogma of MR. But the objection carries a vicious technical feature which could not be dismissed without doing an injustice to Psillos' endeavour.

Carnap's reinvention of the Ramsey-sentence approach has been formed around a structuralist idea: "the structure can be uniquely specified but the elements of the structure cannot. Not because we are ignorant of their nature; rather because there is no question of their nature" (Carnap 1956, 46). But by the same token, the view is liable to Newman's objection. So at the end of his paper, Psillos noted that there is a challenge that the Carnapian should face to obtain the viability of her structural realism:

> If it is not to become a trivial thesis, nor to collapse to scientific realism, then at least a story needs to be told as to how it can survive the Newman challenge. (Psillos 2000a, 275)

Psillos did not think that the approach could, in its present formulation, face the challenge. In a nutshell, Newman's objection holds that:

> Any collection of things can be organised so as to have the structure W, provided there are the right number of them. Hence the doctrine that *only* structure is known involves the doctrine that *nothing* can be known that is not logically deducible from the mere fact of existence, except ('theoretically') the number of constituting objects. (Newman 1928, 144)

And obviously, merely knowing about the number of constituting objects is not enough for maintaining a realist stance. To overcome the objection, the Carnapian should set a restriction on the range of the variables

which set up the theoretical structures. Otherwise, in confrontation with the experience, the theoretical structures would be multiply realizable: having a formal structure is not enough for determining the uniquely true set of the referents of the structure. In other words, as Psillos indicated, there is a dilemma that the advocate of the Carnapian structural realism has to face:

> Either they should choose to avoid addressing the issue of which structures are specified by theories and their Ramsey-sentences, thereby making the claim that theories are true empty and a priori true. Or they should have to appeal to non-structural considerations in order to say which structures are important, thereby undermining the distinction between knowledge of structure and knowledge of nature upon which they base their epistemology and their understanding of theories. (Psillos 2000a, 274)

Psillos has even offered a solution to the objection: the structures should be restricted by contriving a stipulation about the necessity of ranging the variable over the natural classes. This part of solution does not *per se* contradict the structuralist approach. But to fulfil this task, Psillos suggested, the structural realist should be able to make a distinction between natural and non-natural classes, and she has to appeal to some "non-structural knowledge": "the only way to do that is to rely on interpreted scientific theories and to take them as their guides as to which properties and relations are the natural constituents of the world" (Psillos 2000a, 274). Carnap, of course, could not possibly comply with such modifications. The anti-metaphysical allegiance accompanying his structuralism nips anything like appealing to pre-[linguistic]-existing natural kind structures in the bud. So Psillos' solution is not a viable option for the advocate of the Carnapian structural realism.

The solution that I am going to suggest in order to resolve the problem expectedly amounts to appealing to the role of the pragmatic factors in restraining the number of structures and fixing the actually feasible vessels of conveying the factual content. The explanation is simple enough and could be spelled out briefly: the *methodological practical considerations*, or (in semantical period) the practical-pragmatic reasons, which have been the substantial ingredients in the establishment of the Carnapian realism, could very well be appealed to in restricting the range of the constitutive variable

(as well as relations) of the structures of the theory. Of course, these methodological considerations are not to be understood as some formal logical properties attached to structural relations.[8] They are the *meta-logical* practical considerations which fix the relevant relation between the structure and the nature. Sorting the structures according to their (say, computational and empirical) simplicity, expedience, efficiency, fruitfulness, etc., would remarkably help in constraining the number of the appropriate candidates for representing the modal relations between certain domains of objects. Theoretically, it is still possible for two or several structures to organize the same number of things with an equal simplicity, efficiency, etc., but practically, finding even one appropriate structure which could do the job appropriately enough would be quite rewarding.

The meta-logical information (about simplicity, expedience, etc.) does not need to be encoded within the structures, and so, and to Psillos' delight, we can say that there are indeed "*non-structural* considerations" at work in setting restrictions on variables and relations of the existentialised structures. We may even go so far to add that the distinction between knowledge of structure and knowledge of nature is to some extent encroached in this reading of the Carnapian stance. We already saw how the objectivity-related elements, working in the capacity of pragmatic factors, leave their impressions on the choice of the rules and postulates of the system, and penetrate into LFs to influence the semantical and syntactical relations therein. In this way, we can assert that the distinction between knowledge of structure and knowledge of nature is as flimsy as

---

[8] In his examination of Carnap's possible answer to Newman's objection, Ainsworth (2009) disapproved Carnap's approach, and blamed him for inventing logical predicates at whim. According to Ainsworth's reading (based on what Carnap said in his *Aufbau*):

> The essence of the proposal is the suggestion that we should take importance (or as Carnap [1967] calls it, 'foundedness') as a primitive (second-order) logical property that attaches to some relations (in the way that identity is sometimes taken as a primitive logical relation that holds between some pairs). (Ainsworth 2009, 163)

My point is that if perhaps not in *Aufbau*, but at least in Carnap's later syntactical and semantical endeavours, there were the pragmatic-practical reasons, which could be used in the capacity of an unfailing license, and applied in singling out certain logical relations, and highlighting them against the background of the others. And these were not formal logical properties attached to relations, but meta-logical considerations.

the conventional border which has been traditionally drawn between the domains of semantics (concerning the rules of formation and inference of an artificial language) and pragmatics (which in general sense of the term, has been defined as the study which involves speakers of scientific languages... from methodology to the sociology of science (and beyond); see Uebel 2013, 530). Whether the distinction is or is not completely undermined remains beyond the scope of this study. Be that as it may, I showed that Psillos' dilemma is resolvable in Carnapian terms and Newman's challenge does not seem to be a threat to this form of structuralism anymore.

4. Concluding remarks

Let's grant that factual and the conventional elements, or as Quine (1936, 125) once described them – the white and the black threads of the lore – are not quite separable from one another. Although the links are smeared with conventionalism, yet there actually survives a theory of "factual reference" which is strong enough for linking the language to the real, empirical, and objective domain in an indirect way, and yet is subtle enough to not entrap us in the burdensome metaphysical speculations about the nature of the external world or an unexplainable correlation between language and reality. It was with regard to this later point that Carnap said that these questions [of efficiency, fruitfulness, and simplicity] cannot be identified with the question of realism. For the factuality-conducive links which had been carved out within the pragmatically encroached frameworks were not designed to be as cumbersome as metaphysical chains. There is no straightforward semantical story about the hidden access strips between language and reality, nor has any ontological record been presented to account for the pre-existence of the real entities as the blue-prints of the terms of the theory. Carnap's thesis should not be understood as implying that "those who accept and use a language are thereby committed to certain "ontological" doctrines *in the traditional metaphysical sense*" (Carnap 1956, 45, my emphasis). But if we could accept that for obtaining the plausibility of our view, we cannot appeal to methods other than the intellectual tools used in scientific practice, as methodological naturalism persuade us to believe, then we can enjoy all of the benefits of the realistic stance without

paying any unreasonable metaphysical price.[9] As Richardson (2003) indicated, the notion of methodological naturalism had been inspired to Carnap by the pragmatists of the day.[10] And I believe that a respectable though modest version of realism could be built upon this common legacy.

Far from being ashamed on account of its metaphysical poverty, this form of realism, which was founded upon the referential links fostered by pragmatic factors, can stare any other form of realism out of the countenance in a debate over its philosophical richness, any day of the week. There are worse things than being in poverty, after all. Being uncared for, unloved and unwanted are such things. Being in endless and fruitless meta-

---

[9] This point is mainly inspired by Richardson (2003, 21), who explained how Carnap had embraced the methodological naturalism, without making any commitment to metaphysical naturalism. There is some similarity in Richardson and Laudan's conception of methodological naturalism as an empirical discipline of regularities which govern the research (cf. Laudan 1996, 110). But in the present context, the concept is calibrated according to the concerns that Carnap had originally shown about the problem of choice of LF and the factors that rule the choice.

[10] Richardson showed that the term that Carnap and Charles Morris (i.e. the pragmatist of the day) actually used for "methodological naturalism" was "scientific philosophy" (see Richardson 2003, 21). Scientific Empiricism was also the title of Charles Morris' speech at the mentioned meeting, a speech which was planned for reviewing and cherishing the affinities between the aims, methodologies and working plans of logical empiricism and American pragmatism of the day. Participation of Morris (a fervent pragmatist and loyal advocate of Mead and Dewey) to a program which was originally planned by logical empiricists was indeed an early instance of the realization of the very aim of the program.

The concept of "scientific empiricism" was used by Carnap (who perhaps was the original architect of the plan), a few years later, in his "Testability and Meaning" (see Carnap 1936) in an illuminating footnote which was presented to define the main characteristics of philosophical approach of the philosophers who were allegedly called logical positivists:

> It has sometimes been called Logical Positivism, but I am afraid this name suggests too close a dependence upon the older Positivists, especially Comte and Mach. We have indeed been influenced to a considerable degree by the historical positivism, especially in the earlier stage of our development. But today we would like a more general name for our movement, comprehending the groups in other countries which have developed related views... The term 'Scientific Empiricism' (proposed by Morris [i] p. 285) is perhaps suitable. (Carnap 1936, 422)

physical feuds is more undesirable than paying the price of realism by the pragmatist coin.

References

Boyd, A. (1999): Kinds as the "Workmanship of Men": Realism, Constructivism, and Natural Kinds. In: Nida-Rümelin, J. (ed.): *Rationalität, Realismus, Revision: Proceedings of the Third International Congress. Gesellschaft für Analytische Philosophie*. Berlin: de Gruyter.

Carnap, R. (1934): *Logische Syntax der Sprache*. Wien: Springer (Translated as *The Logical Syntax of Language*, 1937, London: Kegan Paul).

Carnap, R (1936): Testability and Meaning. *Philosophy of Science* 3, No. 4, 419-471.

Carnap, R. (1950): Empiricism, Semantics, and Ontology. *Revue Internationale de Philosophie* 11, 20-40. Reprinted in Boyd, R. – Gasper, P. – Trout, J.D. (eds.) (1991): *The Philosophy of Science*. MIT Press.

Carnap, R. (1956): The Methodological Character of Theoretical Concepts. In: Feigl, H. – Scriven, M. (eds.): *Minnesota Studies in the Philosophy of Science. Vol. I: The Foundations of Science and the Concepts of Psychology and Psychoanalysis.* Minneapolis: University of Minnesota Press, 38-76.

Carnap, R. (1958): Observation Language and Theoretical Language. In: Hintikka, J. (ed.) (1975): *Rudolf Carnap, Logical Empiricist*. D. Reidel.

Carnap, R. (1963): Replies and Systematic Expositions. In: Schilpp, P.A. (ed.): *The Philosophy of Rudolf Carnap.* LaSalle, IL: Open Court, 859-1013.

Carnap, R. (1966): *Philosophical Foundations of Physics: An Introduction to the Philosophy of Science*. New York: Basic Books.

Cei, J. – French, S. (2006): Looking for Structure in all the Wrong Places: Ramsey Sentences, Multiple Realizability, and Structure. *Studies in History and Philosophy of Science* 37, 633-655.

Creath, R. (1985): Carnap's Scientific Realism: Irenic or Ironic? In: Rescher, N. (ed.): *The Heritage of Logical Positivism*. Lanham: University of America Press, 117-131.

Creath, R. (1990): *Dear Carnap, Dear Van: The Quine-Carnap Correspondenceand Related Work*. Creath, R. (ed.): Los Angeles: University of California Press.

Creath, R. (2012): Analyticity in the Theoretical Language: Is a Different Account Really Necessary? In: *Carnap and the Legacy of Logical Empiricism*. Vienna: Springer, 65-74.

Ellis, B. (1988): Internal Realism. *Synthese* 76, 409-434.

Fine, A. (1990): Piecemeal Realism. *Philosophical Studies* 61, 79-96.

French, S. (2003). A Model-Theoretic Account of Representation (Or, I Don't Know Much About Art... But I Know It Involves Isomorphism). *Philosophy of Science* 70, No. 5, 1472-1483.

French, S. – Saatsi, J. (2006): Realism about Structure: The Semantic View and Nonlinguistic Representations. *Philosophy of Science* 73, 548-559.

FRIEDMAN, M. (2011): Carnap on Theoretical Terms: Structuralism without Metaphysics. *Synthese* 180, 249-263.

HACKING, I. (1982): Experimentation and Scientific Realism. *Philosophical Topics* 13, 71-87.

HEMPEL, C.G. (1958): The Theoretician's Dilemma. In: Feigl, H. – Scriven, M. – Maxwell, G. (eds.): *Minnesota Studies in the Philosophy of Science*. Vol. II. Minneapolis: University of Minnesota Press, 37-98.

LADYMAN, J. – DOUVEN, I. – HORSTEN, L. – VAN FRAASSEN, B.C. (1997): A Defence of van Fraassen's Critique of Abductive Inference: Reply to Psillos. *Philosophical Quarterly* 47, 305-321.

MORMANN, T. (2007): Carnap's Logical Empiricism, Values, and American Pragmatism. *Journal for General Philosophy of Science* 38, 127-146.

PSILLOS, S. (1996): On van Fraassen's Critique of Abductive Reasoning, *The Philosophical Quarterly* 46, 31-47.

PSILLOS, S. (1999): *Scientific Realism: How Science Tracks Truth*. London – New York: Routledge.

PSILLOS, S. (2000a): Carnap, Ramsey Sentence and Realistic Empiricism. *Erkenntnis* 52, 253-279.

PSILLOS, S. (2000b) (ed.): Rudolf Carnap's 'Theoretical Concepts in Science'. *Studies in History and Philosophy of Science* 31, 151-172.

RAMSEY, F. (1990): *Philosophical Papers*. Edited by D.H. Mellor. Cambridge: Cambridge University Press.

RICHARDSON, A. (2003): Logical Empiricism, American Pragmatism, and the Fate of Scientific Philosophy in North America. In: Hardcaste, G. – Richardson, A. (eds.): *Logical Empiricism in North America*. University of Minnesota Press, 1-24.

RICHARDSON, A. (2007): Carnapian Pragmatism. In: Friedman, M. – Creath, R. (eds.): *The Cambridge Companion to Rudolf Carnap*. Cambridge: Cambridge University Press, 295-315.

UEBEL, T. (2013): Pragmatics in Carnap and Morris and the Bipartite Metatheory Conception. *Erkenntnis* 78, 523-546.

VAN FRAASSEN, B. (1989): *Laws and Symmetry*. Oxford: Oxford University Press.

WORRALL, J. (1989): Structural Realism: the Best of Both Worlds? *Dialectica* 43, 99-124.

ZAHAR, E. (1994): Poincaré's Structural Realism and his Logic of Discovery. In: Heinzmann, G. et al. (eds): *Henri Poincaré: Akten Des Internationale Kongresses, Nancy, 1994*. Berlin: Akademie Verlag, 45-68.

Organon F 22 (3) 2015: 325-339

Some (Philosophical) Problems for Consciousness as a Neural Capacity for Objectivity

JOHN IAN K. BOONGALING

Department of Humanities. College of Arts and Sciences
University of the Philippines Los Baños. College, Laguna 4031. Philippines
johnianboongaling@gmail.com

RECEIVED: 15-02-2015 • ACCEPTED: 30-05-2015

ABSTRACT: This paper is a critical appraisal of the most recent attempt from cognitive science in general, developmental and evolutionary biology in particular, to understand the nature and mechanisms underlying consciousness as proposed by Anton J.M. Dijker. The proposal, briefly stated, is to view consciousness as a neural capacity for objectivity. What makes the problem of consciousness philosophically and scientifically challenging may be stated as follows: If consciousness has a first-person ontology and our best scientific theories have a third-person ontology, how can we come up with a satisfactory theory? Moreover, if the reduction of one to the other is impossible, what are we supposed to do? By neglecting what Chalmers calls the "hard problem" of consciousness, Dijker's proposal seems unable to respond to the foregoing questions, and these questions, I maintain, are the very motivations that most of us have when we inquire about consciousness.

KEYWORDS: Consciousness – objectivity – subjectivity – qualia – mind.

0. Introduction

The *mind* is interesting both as a *phenomenon* and as a *problem* – not only for philosophy but also for the empirical sciences. One might say that it is both *familiar* and *strange*. It is familiar in the sense that the activity of thinking constitutes a huge portion of our lives. It is strange in the sense

that we find it difficult to provide definitive answers to our most important questions about it. The same observation might be said about *consciousness*. For instance, what could be more familiar than the fact that I am conscious right now and that I am writing this paper? What could be more familiar than the fact that I am experiencing something, e.g. seeing the distinct greenness of the leaves of the mango tree (at this time of the year) just outside my study? It is important to note that for philosophers and reflective persons in general, the *science* behind the process of visual perception is not the problem since most of us are already aware of it (e.g. how vision requires light, how light passes to the different parts of the eye (e.g. cornea, lens), the role of photoreceptors in gathering visual information which is then sent to the brain via the optic nerve as electrical signals). Moreover, the science behind visual perception and many other physical/biological processes is not in any way a potential source of perplexity for most of us. What *can* be perplexing about all this may best be summarized by a question: "Why should any experience emerge from molecular-biological processes?" (Kim 2011, 4) At this point, we find ourselves confronted with conflicting intuitions – an experience that is characteristic of the intellectual activity we call *philosophy*.

This paper is an assessment of the most recent attempt from cognitive science in general, developmental and evolutionary biology in particular, to understand the nature and mechanisms underlying consciousness as proposed by Anton J.M. Dijker. The proposal, briefly stated, is to view consciousness as "a neural capacity for objectivity" (see Dijker 2014). For philosophers in general, the idea that scientists can now confidently venture into studying consciousness is a breath of fresh air. Searle, for instance, recounts his personal experience when he first became interested in the problem of consciousness. He says that "most people in the neurosciences did not regard consciousness as a genuine scientific question" (Searle 1997, 193). Indeed, times have changed, and this is a good thing. At present, studies about the mind and consciousness are now done in a more inter/multidisciplinary manner which brings together people from different fields (e.g. biology, neuroscience, psychology, philosophy).

The paper consists of three main parts. The first part is expository. It provides a summary of Dijker's proposal to view consciousness as a neural capacity for objectivity, its theoretical underpinnings and some of its alleged achievements (e.g. the explanation and integration of intelligence, morality,

and esthetics). The second part constitutes the analysis and appraisal of Dijker's proposal. It identifies some philosophical problems and provides arguments that need to be addressed if Dijker's proposal is to count as an acceptable (or at the very least, a coherent) account of consciousness. It also includes responses to some anticipated objections to the arguments that have been presented. The third part – the conclusion – provides a synthesis and a criterion/condition that any theory of consciousness (whether scientific or philosophical) must meet in order to be considered acceptable.

1. Consciousness as a neural capacity for objectivity

Dijker offers a new way to look at consciousness – as "the brain's most adaptive property" which may be described as "a neural capacity for objectivity" (Dijker 2014, 1). As might be expected, how Dijker defines a "capacity for objectivity" is crucial not only for a fuller understanding of his proposal but also for properly assessing it. How then does Dijker define a "capacity for objectivity?" Dijker clearly states it in the following:

> The answer proposed here is: a capacity for objectivity, to be defined as the capacity to produce *states* of objectivity that internally represent objects and their dispositional properties (as well as movements and behaviors predicted by these dispositions) in relatively stable, accurate, increasingly complete, perceiver-independent and neutral ways, unbiased by specific needs, motives, and anticipation of instrumental aspects and rewards. (Dijker 2014, 2)

The foregoing passage highlights the idea that for Dijker, a "state of objectivity" is a state where "subjective aspects are *absent* and one is "just looking" at the world as it really is and can be" (Dijker 2014, 2). For a fuller understanding of viewing consciousness as a neural capacity for objectivity, it is imperative that we discuss its theoretical underpinnings and identify some of the arguments that support it. *First*, as might be noticed from the foregoing definition of a capacity for objectivity, it appears that in general, Dijker adopts a *realist* framework. If we want to be more specific, Dijker adopts a *naïve realist* framework. It is important to note that naïve realism is usually associated with *common sense*. To help us better understand naïve realism, Audi provides us with the following example and description of the view:

> One natural thing to say about what it is for us to see the green field is appealingly brief. We simply see it, in an ordinary way: it is near and squarely before us; we need no light to penetrate a haze or a telescope to magnify our view. We simply see the field, and it may normally be taken to be pretty much as it appears. This sort of view, called *naïve realism*, has been thought to represent common sense: it says roughly that perception is simply a matter of the senses telling us about real things ... it is a form of realism because it takes the objects of perception to be real things external to the perceiver, the sorts of things that are "out there" to be seen whether anyone sees them or not. (Audi 2011, 38)

That Dijker adopts a naïve realist framework is unexpected (but I will discuss the reasons why in the next part of the paper). The *second* important theoretical component of Dijker's proposal involves a combination of a developmental and an evolutionary view on a capacity for objectivity. This theoretical component is important because it allows Dijker to identify the underlying *mechanisms* that can help explain human beings' ability "to integrate intelligence, morality, and esthetics" (Dijker 2014, 3). It is important to note that this integration is supposedly one of the important *achievements* of Dijker's proposal. This is done by linking together the capacity for objectivity with various behavioral manipulations such as exploration, play, and a mechanism of care (Dijker 2014, 6). Consider what Dijker says in the following:

> [S]tates of objectivity are not only realized by brain mechanisms of a subject trying to make sense of a pre-existing objective world, but also by behavioral attempts to make objects themselves permanent by preserving, protecting, perhaps even constructing and beautifying them. These attempts most likely are motivated and controlled by a specific motivational mechanism with a social origin. (Dijker 2014, 6)

As the foregoing passage shows, our initial observation is correct (i.e. that Dijker adopts a naïve realist framework). However, what needs to be emphasized in the foregoing passage is the mechanism itself which links together the various behavioral attempts mentioned (e.g. preserving, protecting, beautifying). This mechanism is *care*. How exactly does the care mechanism work and how does it link together intelligence, morality, and esthetics according to Dijker's account? To see how such a mechanism works, Dijker needs another important concept: *vulnerability*.

> From an evolutionary perspective, vulnerability can be defined as the disposition or likelihood of living things to change into a state of lowered fitness (a state inconsistent with their "design specification") when exposed to certain conditions. (Dijker 2014, 7)

For Dijker, the vulnerability of both the perceiver and the object being perceived, for instance, in a state of exploration or play, allows for *modifications* in the perceiver's behavior. For example, we tend to be gentle or careful in handling things or other living things which we perceive to be fragile.

The *third* crucial element in Dijker's proposal involves ideas presented by Merker (2013) concerning significant interactions of three things: brains, bodies, and their world. Dijker notes that for Merker, a conscious state "allows the organism to be primarily concerned with the objective aspects of its environment and not to be bothered by the sensations that might be produced by underlying perceptual and behavioral mechanisms" (Dijker 2014, 6). As noted by Dijker himself, his proposal is distinct from Merker's in the sense that it further adds that "a conscious state requires awareness of the possibility of multiple looks or behavioral manipulations, and the inhibition of motivational systems that could bias perception" (Dijker 2014, 6).

Mindful of the underpinnings of consciousness as a neural capacity for objectivity, we are now in a better position to describe how such a capacity, according to Dijker, can *integrate* intelligence, morality, and esthetics. As Dijker optimistically remarks:

> Perhaps, a capacity for objectivity and its foundation on a care mechanism are the key to the century-old philosophical puzzle of how judgments of truth, moral goodness, and beauty are related. (Dijker 2014, 8)

For a rough sketch of the idea, it is important to note that the integration is made possible by the following: vulnerability, care mechanism, and the distinct aspect of Dijker's proposal: *multiple looks*. Let us begin with intelligence:

> States of objectivity are necessary for the kinds of problem solving that we tend to consider intelligent and creative. When in a state of objectivity, one tries to be as complete as possible, by looking at objects from multiple perspectives and performing small, virtual what-if experiments, thereby coming to understand or "grasp" the many relationships among

> objects and their properties that are possible ... To illustrate, briefly consider an experiment performed with crows to demonstrate how previously acquired knowledge of object or tool properties and corresponding skills are used in a novel context, suggesting perceiver-independent or objective internal representations. (Dijker 2014, 8-9)[1]

One of the results (although admittedly controversial) of the research in the abovementioned passage is that the success of the crows can be attributed to their cognitive ability which involves knowledge (of some sort) of abstract causal rules. Relying on recent data on the problem solving skills of certain animals, Dijker maintains that "it is very difficult to imagine how this ability is possible without the birds having acquired a perceiver-independent and objective representation of the total configuration of objects and their individual but interrelated physical properties" (Dijker 2014, 9). Take note that in the foregoing passage, Dijker includes the ability of looking at objects from multiple perspectives. This suggests that (at least as Dijker sees it), the capacity for objectivity is a *necessary* condition for both intelligent and adaptive behavior. In the following, Dijker attempts to explain the concept of *conscience* using his proposal (take note of the employment of vulnerability, care mechanism, and multiple looks (or perspectives) in the overall explanation):

> It may be proposed that the concept refers to the accurate or objective perception of a vulnerable object (i.e., to being conscious of a vulnerable object), activation of a care mechanism, and perception or anticipation of the different negative consequences of one's own behavior for the object's well-being or fitness, typically experienced as the emotion of guilt. Thus while tenderness is a response to observing that a vulnerable object is in the desirable state of good health, guilt implies the causal attribution to the self of an observed or anticipated decrease in health.

---

[1] The experiment that Dijker mentions concerns crows' successful performance of obtaining food (meat in particular) through a hole inside a box that could only be obtained by using not just one but *several* tools. The said experiment is setup in the following way: The meat is placed inside a box. The meat can only be obtained by inserting a stick through a hole in the box that is long enough to reach it. Such a stick is available but it is visibly contained in another box. The stick can only be reached by using another tool – a shorter stick – which is attached to a string from a branch. For further details, see Taylor et al. (2010).

> Other moral emotions more strongly focus on the harmful behavior of third parties (e.g., moral anger) or the undesirability of the object's lowered fitness and suffering. (Dijker 2014, 10)

As is well-known, human beings are toolmakers and users. Dijker capitalizes on this idea and explains how states of objectivity integrate *esthetic experience* in the following:

> A state of objectivity integrates esthetic experience, tenderness, care, and specific motor aspects. Hence there may be a close association between making beautiful things (art), craft, and tool making. In particular, during the initial stages of tool making, the tool is perceived as a vulnerable object that needs to be treated with care and brought into a less vulnerable and more mature shape by allowing it to "grow" or develop according to its inherent material properties, with the tool maker facilitating this with a gentle and protective attitude (involving activities such as cleaning, polishing, inspecting, touching, testing, and reshaping). (Dijker 2014, 10)

Earlier, I mentioned that one of the supposed achievements of Dijker's proposal (i.e. to view consciousness as a neural capacity for objectivity) is the integration of intelligence, morality, and esthetics. In general, Dijker accomplishes this by combining a naïve realist framework, a combination of a developmental and an evolutionary view, and Merker's work on the interactions of brains, bodies, and their world with an additional requirement: the possibility of multiple looks (or perspectives).

I hope that the foregoing discussion clearly shows the significant concepts that Dijker's proposal employs: vulnerability, care mechanism, and multiple looks (or perspectives). At this point, the expository part of the paper is complete. The next part is concerned with the appraisal of Dijker's proposal and its proper place in our continuous attempts to understand (or make sense of) the nature and underlying mechanisms of consciousness.

2. Some (Philosophical) Problems for Consciousness as a Neural Capacity for Objectivity

Novel theories are always welcome in our continuous efforts to understand the nature and underlying mechanisms of consciousness, but not, we

hope, at the expense of oversimplifying or ignoring important *theoretical*, sometimes *philosophical*, questions that make it a difficult problem in the first place. While past and current empirical researches on various aspects of consciousness may prove to be helpful, it is important to be able to weave together their various results under a cogent theoretical framework. As usual, in both philosophy and science, we need both theory and evidence to mutually support each other.

For a short but helpful background on the issue, it is best to begin with *conscious mental states*. That an organism has conscious mental states means that "there is something it is like to be that organism" (Nagel 1979, 166). In contrast, there is nothing, in the relevant sense, of what it is like to be a book, a table or a chair. If this is correct, then conscious mental states are characterized by a kind of *qualitative feeling* or by "the subjective aspects of experience" (Campbell 2005, 189). This is what philosophers mean by the term *qualia* (in singular form, *quale*).

Philosophers and psychologists also distinguish between two levels of consciousness: (1) *simple awareness* (i.e. nonreflective conscious functioning) and (2) *reflective consciousness* (i.e. reflective conscious functioning) (cf. Martí – Rodríguez 2012, 103-104). The first level involves *representations* (e.g. percepts, consciousness of an object's properties (e.g. 'red')). The second level involves *metarepresentations* (e.g. reflection about the experience of 'red'). The difference between the two levels is that on the first level, "the subject is a mere spectator of his functioning" whereas on the second level, the subject is "also an observer of his functioning" (Martí – Rodríguez 2012, 104).

Before I provide some theoretical or philosophical problems with consciousness as a neural capacity for objectivity, let me state that the attempt by itself of integrating intelligence, morality, and esthetics is commendable. I do acknowledge that it is a difficult task. While I acknowledge these things, I think Dijker's proposal still needs further refinement for it to be considered as a tenable position to take. Let me state the reasons why. *First*, an acceptable theory of consciousness should be responsive to the "hard problem of consciousness" (Chalmers 1996, xii) and this problem takes the *subjectivity* (e.g. first-person account) of consciousness seriously. Unfortunately, Dijker's proposal *neglects* them both. Neglecting this problem has important implications for viewing consciousness as a capacity for objectivity: (1) consciousness is usually understood as having a "first-person

ontology" (cf. Searle 1997, 212); (2) that consciousness has a first-person ontology poses a difficult problem for any theory that seeks to explain consciousness in physicalist (or materialist) terms (This is because physicalist accounts have a third-person ontology. It is important to note that even our usual understanding of knowledge makes use of the third-person perspective.); (3) we cannot reduce first-person subjective experiences to third-person phenomena, and vice versa (cf. Searle 1997, 212). It is important to note that it is precisely for these reasons that the problem of consciousness is perplexing in the first place. Mindful of these points, we can understand, for example, why philosophers of mind observe that "we are entirely in the dark about how consciousness fits into the natural order" (Chalmers 1996, xi). Many philosophers of mind will agree that it is precisely the *subjectivity* of consciousness and the supposed *objectivity* of the natural order which makes it difficult for us to come up with a satisfactory theory of mind (whether *scientific* or *philosophical*).

Dijker's neglect of such an important aspect of consciousness is unfortunate because consciousness is a "natural phenomenon" (Chalmers 1996, xiii) or a "biological phenomenon" (Searle 1997, 6). My complaint about Dijker's proposal is simple: If consciousness is primarily characterized by subjectivity, then our theory about consciousness should be able to accommodate it and not neglect it. Dijker's neglect of subjectivity is also unfortunate for another reason: There is an available option which actually tries to accommodate subjectivity in describing a consciousness like ours:

> By 'consciousness like ours,' we mean *the subjective experience of a suitably neurobiologically complex living organism*. Such consciousness is subjective insofar as it necessarily involves an egocentrically centered, single point of view that is spatio-temporally located wherever and whenever one's body is located. (Maiese 2011, 11)

At this point, let me provide some possible objections that might be raised against the argument that I have presented so far. It might be argued that Dijker actually discusses a certain kind of subjective experience in his work: the experience that "one is 'just looking' at the world as it really is and can be" (Dijker 2014, 2). This brings me to my *second* point: Dijker's account of subjective experience *deviates* from our usual understanding of the kind of subjectivity that is involved in theorizing about consciousness. Such deviancy therefore needs to be *justified* (or at the very least, explained).

In the article, Dijker clearly describes "states of objectivity" as states where "subjective aspects are *absent* and one is "just looking" at the world as it really is and can be" (Dijker 2014, 2). If this is the kind of subjective experience that Dijker is discussing, then this is simply problematic. It is difficult, even to imagine, a subjective experience where "subjective aspects are *absent*." This entitles us to say that even if such an account of subjective experience is included in Dijker's work, it is plausible to maintain that such an account is prone to the charge of being internally *inconsistent*. The import of the discussion so far is that we need a theory which can maintain the subjectivity of consciousness and the sort of objectivity that is required by our best scientific theories in accounting for conscious phenomena.

Another objection that might be raised against the points presented so far concerns the two levels of consciousness discussed earlier. Will such a distinction help Dijker's proposal? To a certain extent, it can, but only if we do not take the hard problem of consciousness seriously. If we take the hard problem of consciousness seriously, we cannot easily appeal to the familiar distinction that we have between *appearance* and *reality*.

> For example, the sun appears to set but the reality is that the earth rotates. But you cannot make this move for consciousness, because where consciousness is concerned the reality is the appearance. (Searle 1997, 212-213)

This means that we cannot isolate *qualia* from *consciousness*. "There are not two types of phenomena, consciousness and qualia. There is just consciousness, which is a series of qualitative states" (Searle 1997, 9). If this is correct, then what do we mean by "just looking at the world as it really is" as described by Dijker? The most charitable interpretation of the aforementioned phrase from Dijker is an interpretation which contextualizes it in a naïve realist framework. It is in that framework, we might say, where it does have (or makes) sense.

Here then is the *third* point: It is clear that Dijker adopts naïve realism and this, as I mentioned earlier, is unexpected. I take it as uncontroversial (i.e. that it is common knowledge) amongst philosophers in general, epistemologists and philosophers of mind and science in particular, that naïve realism is problematic. It is prone, for instance, to problems that range from the simpler (e.g. problems associated with visual perception and hal-

lucination) to the more complicated ones (e.g. several experiments in quantum mechanics (e.g. double-slit, quantum erasure, EPR pairs)). Clearly, Dijker is *silent* on these problems. On the extreme, some philosophers might even say that naïve realism has been discredited already.

The *fourth* point is devastating and it is a corollary of the arguments that have been presented so far: If we start with the Searlean premise that "[t]here are not two types of phenomena, consciousness and qualia. There is just consciousness, which is a series of qualitative states" (Searle 1997, 9), then does it not follow that we have a genuine problem for viewing consciousness as a neural capacity for objectivity, more especially so given that such a view is grounded in naïve realism?

The *fifth* point may be summarized as follows: Dijker's proposal makes use of *multiple looks* but this strategy seems to get the order of explanation *backwards*. What this means is that the very possibility of multiple looks is intelligible only through the prior recognition of my point of view as a view among many other points of view. This means that it is the concept of subjectivity that can help explain objectivity and not the other way around. If this is correct, then subjectivity (in the relevant sense) must be incorporated (and not neglected) in our theory of consciousness. It is important to note that I do not intend to show that human beings have no capacity for objectivity. Indeed, we have such a capacity. But such a capacity is only possible because consciousness is subjective by *default* (e.g. in visual experience, it is precisely because of my *situatedness* and *physical constitution* that I see an object as thus-and-so).

Let me expound on the fifth point. We can begin by taking note of two familiar facts about beings like us: (1) that we have certain views or perspectives, and (2) that our thoughts always have certain objects. In a sense, we can say that our thoughts are always directed at something (or they are always about something). Let me begin by expounding on (1). What does it mean to have a view or a perspective? In order to make sense of this question, we have to recognize how it is even possible for beings like us to have a view or a perspective. The answer seems readily available to us: It is possible for us to have views or perspectives precisely because we are in such a position that we can have them. This means that to have a view or a perspective entails a prior recognition that we are occupying a particular position in the world (or the universe) – like a particular dot in a coordinate system. Being situated in this sense allows for the possibility of (1) and

thus serves as some kind of grounding for it – comparable but not entirely identical to Kant's categories of space and time and their significant roles to fulfill in the very possibility of experience (see Kant 1992). There is more to be said about (1) and it is deeply connected to (2). The previous analogy concerning a dot in the coordinate system does not really tell us the whole story. It certainly provides us with a picture but it is obviously an incomplete one which can easily lead us into error if we are not careful. Being situated, by itself, certainly would not be sufficient for something to have a view or perspective. The being in question, must therefore be configured in a particular way – or have some sort of functional organization – such that it can have a view or a perspective. In other words, the being in question, must possess a mind (or anything which functions like one) or if we want to make a bolder claim, the being in question must be a mind. We can ignore the other difficult issues concerning the previous remark (perhaps we can deal with them in another paper). For now, it is enough that when we think about what it means to have a view or a perspective and appreciate the intentional character of our thoughts, these familiar facts about ourselves point us directly to the complex phenomenon that is the *mind*.

If, as the foregoing discussion suggests, we can only make sense of the idea that to have a view entails being situated, does it mean that the mind will always be trapped in its own subjectivity, that it can only know, for instance, the world or the self from its own subjective point of view and experiences? No, it does not in any way mean that. (If that is what it means, then we commit ourselves to solipsism and I think that there are better positions to take than that of the solipsist.) Even if our primary means for experiencing or even discovering the world or the self is our own point of view (and thus, subjective), we can (and with good reasons) say that we are capable of achieving something more – an objective view of the world and of the self as a point of view among many others that are included in our conception of the world.[2] In my estimation, this is made possible by the mind's capacity for *imagination* and *abstraction*. It is not difficult to see that we can and we do place ourselves in the place of others (e.g. when we want to understand the reasons why a person acted in a particular way). These are cases that adults like us are all familiar with, and in these cases, we can say that it is possible for us to *transcend* our subjective point of view and

[2] Here, I am following Nagel (1989).

think *as if* we are the other person. These cases demonstrate that the mind is capable of stepping back, and moving from a mere subjective standpoint to a more objective one. It is important to note that our capacity for *empathy* shares the same general feature.

The foregoing discussion lays the basis for the *sixth* point: the employment and requirement of multiple looks in Dijker's proposal (which for him constitutes the distinct aspect of his proposal) appears to be more suited for characterizing objectivity not as a neural capacity but as "a method of understanding" *à la* Nagel (cf. Nagel 1989, 4). Surely, we cannot equate consciousness with a method of understanding. It is important to note that Dijker is not simply saying that consciousness has a feature, call it a neural capacity for objectivity. He is saying that consciousness *is* a capacity for objectivity. The relevant use of 'is' in Dijker's proposal is therefore the 'is of identity' and not merely the 'is of predication.' As such, I am expecting to find a set of necessary and sufficient conditions from Dijker's discussion. Such conditions however are nowhere to be found.

Another important point worth emphasizing is that language might provide us with a clue as to how it is possible for beings like us to achieve an objective view about the world or the self in relation to that world. (Perhaps *language* does not merely provide us with a clue but actually serves as the vehicle in which we are able to achieve an objective view of the world or the self.) For instance, I might start with my subjective views and experiences. From these subjective views and experiences, I am able to abstract that all the *impressions* (or *sense data*, if we like) that I encounter always involve the 'I' (*á la* Kant's transcendental unity of apperception) as the 'subject' of those impressions. In other words, these impressions are all subsumed under one consciousness. In these examples, I wish to highlight something that might easily go unnoticed: The fact that I can think about myself 'as if' I am not myself shows that the 'I' can be the 'object' of my inquiry (again, 'as if' the 'I' (which is the object of inquiry) is distinct from the other 'I' (which conducts the inquiry)). If this is not a manifestation of a human being's capacity for objectivity (in the relevant sense) made possible by language and our capacity for imagination and abstraction, then it is difficult to see what can count as one. I hope that it is clear from the foregoing discussion that the capacity for objectivity is only made possible because consciousness is subjective by default. Unfortunately, it is this same characterization of consciousness that Dijker's proposal neglects.

Finally, the foregoing points (or problems) taken collectively, are what I have in mind when I said at the outset that we should not ignore important theoretical, sometimes philosophical, questions that make the problem of consciousness a difficult problem in the first place. Surely, the foregoing problems are theoretical (or philosophical) problems that most philosophers will encounter when they read Dijker's proposal to view consciousness as a neural capacity for objectivity.

3. Conclusion

Any philosophy (or theory) of mind worth taking seriously must include two important things: (1) the phenomenon of consciousness and (2) a satisfactory explanation (or solution) to the hard problem of consciousness. These requirements, I maintain, should not be neglected. To expound on these requirements, it is important to note that (1) entails the recognition that consciousness is to be treated as part of this world and not something outside it. This requirement sits well with science in general. As might be expected, the situation is different with philosophy. I can only hope that people from both science and philosophy can begin to realize that they cannot continue ignoring each other. In addition, (2) entails the recognition that it is the subjectivity of consciousness that is responsible for our current inability to fit consciousness into the natural order. While Dijker's proposal might not have significant problems with the first requirement, I hope that it is clear from the arguments that have been presented that the proposal suffers from significant problems with the second requirement.

Let me end this paper with the following remark about the problem of consciousness. If we appreciate the problem of consciousness in its full complexity, then we are left with the difficult problem of choosing between two standpoints that stand in diametrical opposition with each other: the *subjective* and the *objective*. The prospect of a *rapprochement* between these standpoints seems to be the first business of any serious philosopher of mind because simply choosing one and leaving out the other seems incorrect (or at the very least insufficient) for what we seek in general is *understanding*. Since the early beginnings of the philosophy of mind in the 20th century, we have become more knowledgeable about many things, ourselves

and the world included. It is therefore surprising that now, more than ever, we feel the great burden of trying to make sense of the apparent conflict between our best scientific theories on the one hand, and our conception of ourselves on the other.

References

Audi, R. (2011): *Epistemology: A Contemporary Introduction to the Theory of Knowledge*. New York: Routledge.

Campbell, N. (2005): *A Brief Introduction to the Philosophy of Mind*. Toronto: Broadview Press.

Chalmers, D. (1996): *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

Dijker, A. (2014): Consciousness: A Neural Capacity for Objectivity, especially Pronounced in Humans. *Frontiers in Psychology* 5, No. 223, 1-15.

Kant, I. (1992): *Critique of Pure Reason*. Smith, N. K. (trans.) London: The Macmillan Press.

Kim, J. (2011): *Philosophy of Mind*. Boulder: Westview Press.

Maiese, M. (2011): *Embodiment, Emotion, and Cognition*. London: Palgrave Macmillan.

Martí, E. – Rodríguez, C. (2012): *After Piaget: History and Theory of Psychology*. New Brunswick: Transaction Publishers.

Merker, B. (2013): Body and World as Phenomenal Contents of the Brain's Reality Model. In: Pereira Jr., A. – Lehmann, D. (eds.): *The Unity of Mind, Brain, and World: Current Perspectives on a Science of Consciousness*. Cambridge: Cambridge University Press, 7-42.

Nagel, T. (1989): *The View from Nowhere*. New York: Oxford University Press.

Nagel, T. (1979): What is it Like to be a Bat? In: Nagel, T.: *Mortal Questions*. Cambridge: Cambridge University Press, 165-180.

Searle, J. (1997): *The Mystery of Consciousness*. New York: The New York Review of Books.

Taylor, A. – Elliffe, D. – Hunt, G. – Gray, R. (2010): Complex Cognition and Behavioral Innovation in New Caledonian Crows. *Proceedings of the Royal Society of London B: Biological Sciences* 227, 2637-2643.

Organon F 22 (3) 2015: 340-344

Did the Past Really Change in 2012?

TERENCE RAJIVAN EDWARD

The University of Manchester. School of Social Sciences
The University of Manchester. Manchester M13 9PL. United Kingdom
t.r.edward@manchester.ac.uk

RECEIVED: 11-04-2015 • ACCEPTED: 11-05-2015

ABSTRACT: There is an intuition that the past does not ever change. In their paper 'The puzzle of the changing past', Luca Barlassina and Fabio Del Prete argue that in 2012 the past changed. I show that we are not in a position to accept their argument.

KEYWORDS: Change – competent authority – past – winner.

In a recent paper entitled 'The puzzle of the changing past', Luca Barlassina and Fabio Del Prete reject the impression that the past cannot change (Barlassina – Del Prete 2015). Lance Armstrong was declared the winner of the Tour de France cycling race on 23rd July 2000 by *Union du Cyclisme International* (UCI). On 22nd October 2012, UCI withdrew all of Armstrong's Tour de France wins, because they found out that he had made use of banned substances while competing. Barlassina and Del Prete believe that this withdrawal changed the past. It was once true that Armstrong won the Tour de France, because of what UCI declared on 23rd July 2000, but the withdrawal means that from 22nd October 2012 it is false that Armstrong won the Tour de France. Barlassina and Del Prete portray an aspect of the past, who won the Tour de France, as determined by an authority in such a way that a declaration years later changed the past. However, there is an objection to their view and their attempt to reject this objection is at present inadequately justified.

The objection I have in mind is that Armstrong was never the winner, whatever UCI might once have declared, because he cheated by using banned substances. Barlassina and Del Prete respond as follows:

> This objection rests on a confusion, by which the property *being the winner* is conflated with the property *being the person who deserves to win*. True enough, one cannot enjoy the latter property if one cheated; however, one can enjoy the former even if one cheated, since the possession of the property of *being the winner* is determined solely by a declaration of a competent authority, and a competent authority may, for one reason or another, declare a cheater a winner. (Barlassina – Del Prete 2015, 62)

In order to support this point, they appeal to a case from another sport: football (soccer). The 1986 World Cup match between Argentina and England was won by Argentina, who scored two goals to England's one, but one of Argentina's goals was scored by violating a rule. Diego Maradona scored a goal with his hand. The referee did not see this, we are told. Since Maradona's rule violation was intentional, Argentina won despite cheating. This is meant to show that the property of being the winner is determined solely by a declaration of a competent authority (cf. Barlassina – Del Prete 2015, 62).

I will identify three obstacles to endorsing this rejection of the objection. By 'obstacles', I mean things that Barlassina and Del Prete, or someone else, must do before we are in a position to endorse this rejection. The obstacles are in italics below. Note that in this paper, I use 'determine' in the following sense: for X to determine that Armstrong is the winner is for X to make it the case that Armstrong is the winner. There is another sense of 'determine' in which if something determines who the winner is, then it provides a good means of finding out who the winner is.

1. Barlassina and Del Prete do not define 'competent authority'. They put considerable effort into being precise regarding other points within their article, but regarding this matter the reader is left to guess the meaning from the two sporting examples. *However, we require a clarification before we can endorse their response to the objection.*

There are two reasons why we require a clarification. One reason is that it is unclear how exactly we are to understand the term 'competent author-

ity'. I suspect that a football referee can be very bad at their job and still count as a competent authority for Barlassina and Del Prete. I wonder if the referee can even act on a bribe and be counted as a competent authority, as they are using the term. If so, their use of the word 'competent' is potentially misleading.

Another reason why we require a clarification is this: there is a way of understanding what a competent authority is for Barlassina and Del Prete which leads to a regress. What determines who the Tour de France winner is for them? It is the competent authority on who the winner is which determines this. But what determines who the competent authority is? What determines that it is UCI? Is it some other authority, an authority on who the competent authority is regarding the Tour de France winner? If so, what determines who that other authority is? Is it yet another authority? There is a danger of a regress: a regress of background authorities. Barlassina and Del Prete need to either clarify what a competent authority is in a way that avoids this regress or else acknowledge the regress and explain why it is not a vicious regress.

2. In order to support their view that it was once true to say that Armstrong won the Tour de France, Barlassina and Del Prete appeal to the case of Argentina's victory over England in the 1986 World Cup. Barlassina and Del Prete think that on the basis of this World Cup case, we should agree that being a winner is determined by the declaration of a competent authority (Barlassina – Del Prete 2015, 62). For convenience of expression, I will often omit the declaration element when evaluating this view below.

Barlassina and Del Prete appear to make the following argument: if the winner of one particular competition is/was determined by a competent authority, then the winner of any competition is determined by a competent authority; there is one particular competition in which the winner was determined by a competent authority; therefore the winner of any competition is determined by a competent authority. This argument can be disputed by presenting a single example in which it does not seem as if the winner is determined by an authority. If the impression is correct, then the argument must have gone wrong somewhere. (And since the first premise is an assumption, rather than something argued for, Barlassina and Del Prete are not in a strong position to insist that their conclusion applies to a proposed example.)

Consider the following situation, which was possible once upon a time. Bobby Fischer and Boris Spassky agree with a certain chess authority to play a chess match. Fischer demands that the match be played behind closed doors, with no cameras present and no witnesses other than the players and members of the authority. Fischer wins by checkmate. But the authority decides to teach Fischer a lesson for being so demanding and it declares that Fischer lost. Fischer protests to the world at large. It seems to me that in this hypothetical situation, Fischer is still the winner, even if he cannot prove it. Maybe it will be said that in this situation, the chess authority does not count as competent. But even if the chess authority had spoken honestly, I do not see why that would change what determines the winner, i.e. what makes it the case that Fischer is the winner. Why would it not be the same thing that determines that Fischer is the winner either way – in short, the fact that he checkmated Spassky? If it is, then Barlassina and Del Prete's argument must have gone wrong somewhere. *They need to counter this challenge.*

3. In some competitions, perhaps in all competitions, the rules specify the conditions that need to be met in order to be the winner. If there were no such specification, how would competitors know what to do? The rules of a race say that the winner is the person who meets certain conditions. Those conditions might not include or entail that the winner is whoever some authority declares to be the winner. For example, the rules might specify that the winner is the competitor who has not taken certain substances and has finished ahead of all other competitors who have not taken such substances. If the rules specify this, then there is potentially a clash between who the rules entail is the winner versus who the winner is according to the 'competent' authority. Barlassina and Del Prete say that the objection we are considering is based on confusing who deserved to win with who won, but this is not necessarily true. If the rules of the Tour de France specify that the winner is the person who meets conditions X, Y and Z and Armstrong does not meet those conditions, even though he was once judged by the 'competent' authority to meet those conditions, then the objector may appeal to the rules as determining who the winner really is. *Barlassina and Del Prete need to show that this appeal is mistaken.*

Note that it is not obvious that the winner, going by the rules, is always identical to the person who deserves to win. If you play an opponent in

a one-on-one competition and they are expected to win, with good reason, yet they suffer an unlucky injury, forcing them to resign, do you deserve to win? Presumably, some people will say, 'Yes,' while others will say, 'No, you were just lucky.'

References

BARLASSINA, L. – DEL PRETE, F. (2015): The Puzzle of the Changing Past. *Analysis* 75, No. 1, 59-67.

Organon F 22 (3) 2015: 345-357

Searle on the Intentional Content of Visual Experiences

ANAR JAFAROV

Institute of Philosophy, Sociology and Law. Azerbaijan National Academy of Sciences
Azerbaijan Republic
anarjafarov@hotmail.com

RECEIVED: 04-03-2015 • ACCEPTED: 31-05-2015

ABSTRACT: There is no unique idea regarding the form of the (Intentional) content part of visual experience in the specification. The philosophers' approaches diverge as to whether the content of visual experience is equivalent to a sentence expressing proposition or not. Some of them (mainly philosophers from the phenomenological tradition) consider that one must use a proposition for the specification of the content only when the subject, while having a visual experience, exercise a concept or judge. For the other cases, which can be called *simple seeing*, a noun phrase is preferable. I argue that, holding that the specification of Intentional content of the visual experience should be in the form of a proposition, John Searle gives up the first-person Intentionality and therefore bypasses the first-person important distinction between simple seeing and judgmental seeing. The specification of the content only in the form of the proposition does not allow making such a distinction on the level of description. Then I argue that the feature of the causal self-referentiality of the visual experience belongs to its psychological mode but not, as Searle holds, to the Intentional content of the visual experience.

KEYWORDS: Intentional content – Searle – visual experience.

1. Introduction

Every theory of Intentionality is to explain how the Intentional state must be individuated. For this purpose, most theories of Intentionality distinguish the psychological mode of the Intentional state from the objects

or state of affairs it is directed at. These are two crucial points in the individuation of the state. The Intentional states such as *a visual experience* that there is a yellow wagon there and *a remembrance* that there is a yellow wagon there are directed at the same state of affairs, whereas their psychological modes are different. Yet a visual experience *that there is a yellow wagon there* and a visual experience *that a man is walking in the garden* are psychologically the same, but are different Intentional states; because the state of affairs at which they are directed are different.

However, it is not sufficient for the individuation of an Intentional state, since there can be cases where the psychological mode and the object are the same, but the Intentional states are still different. For the same object I might have a belief *that there is a yellow wagon there* and *that there is a yellow cubic form thing there*. Therefore, most theories of Intentionality distinguish still one point – Intentional content of the state. The Intentional content of the state contains the mode of presentation of objects or state of affairs. When we specify the content, we make explicit how our Intentional state is directed at its object.

To this we should add that, when the content is complex, its structure order also becomes important for the individuation of the Intentional state. For example, my visual experiences *that the pen is on the paper* and *that the paper is on the pen* are different. Because though the contents of the states have the same constituents, their structure orders are distinct.

The content is brought into light in reflection on the Intentional state by a person who directly experiences it. So the specification of the content should be committed to the subject's mode of (re)presentation of objects or states of affairs toward which the Intentional state is directed. If I see an apple tree and specify the content of my visual experience as *an apple tree*, or *that this is an apple tree*, the specification of the content is exactly committed to my mode of presentation of that object. Even if I know much more about apple trees, I cannot specify my knowledge on the content, substitute *an apple tree*, say, with the description *a deciduous tree whose fruits I like to eat*, which might be implicit at the moment of the visual experience, or add any implicit knowledge to the content. If I do so, then it means that my specification does not express how the content exactly is.

However, there is no unique idea regarding the form of the content part of visual experience in the specification. The philosophers' approaches di-

verge as to whether the content of visual experience is equivalent to a proposition or not. Some of them (mainly philosophers from the phenomenological tradition) consider that one must use a proposition for the specification of the content only when the subject, while having a visual experience, exercise a concept or judge (cf. Mulligan 1995, 170). For the other cases, which are called *simple seeing*, a noun phrase is preferable. Searle, however, maintains that the linguistic correlate of the content should be a sentence expressing proposition.

In what follows I will argue that, holding that the specification of Intentional content of the visual experience should be in the form of a proposition, Searle gives up the first-person Intentionality and therefore bypasses the first-person important distinction between simple seeing and judgmental seeing. Moreover, if we hold the view that the role of the content in the individuation of the state consists in its containing the mode of presentation of objects or state of affairs, then Searle's theses that the specification of the content of the visual experience is propositional and that the specification of the causal self-referentiality of the visual experience is to be made explicit in the content of that state do not match that function of the content of the Intentional state. Then I will argue that the causal self-referentiality and the other features of the visual experience belong to its psychological mode but not to the Intentional content of the visual experience.

Before touching these problems in a detailed way, let us first briefly consider some ideas from Searle's conception of the visual experience, which are apt for this paper.

2. Searle's conception of the visual experience

Searle's conception of the visual experience is a part of his theory of Intentionality whose main idea is that "[...] every Intentional state consists of an *Intentional content* in a *psychological mode*. Where that content is a whole proposition and where there is a direction of fit, the Intentional content determines the *conditions of satisfaction*[1]" (Searle 1983, 12). The visual expe-

[1] The conditions of satisfaction is not an unambiguous notion. In *Intentionality,* where this notion is clarified, Searle writes:

rience as a kind of the Intentional state also bears these features. For Searle, "[i]t does not just make reference to an object", but its "[...] content requires the existence of a whole state of affairs if it is to be satisfied" (Searle 1983, 40). Therefore, he holds that the content of the visual experience has to be propositional. For example, if "I have a visual experience of a yellow station wagon", the content of my visual experience is *that there is a yellow station wagon there*, but not merely *a yellow station wagon*. But this is not all; for he additionally holds that some kind of Intentional states (intentions, orders, and visual experiences) have more complicated content than what one can specify by directly reflecting on it. The visual experiences, like intentions and orders, for Searle, are causally self-referential, and this feature should be specified in the content of those states. So, according to him, the Intentional content of the visual experience of a yellow station wagon has to be made explicit in the following form:

> Conditions of satisfaction are those conditions which, as determined by the Intentional content, must obtain if the state is to be satisfied. For this reason the specification of the content is already a specification of the conditions of satisfaction. Thus, if I have a belief that it is raining, the content of my belief is: that it is raining. And the conditions of satisfaction are: that it is raining – and not, for example, that the ground is wet or that water is falling out of the sky. (Searle 1983, 12-13)

From this passage we can see that, by holding that they have the same specification, Searle endorses the conditions of satisfaction to be depended on the Intentional content. According to his "it is raining" example – since it holds that if the same state of affairs was believed under a different aspect, the conditions of satisfaction of the corresponding Intentional state would be different – one can even maintain that, like the content, the conditions of satisfaction contain a mode of representation of the state of affairs. R. McIntyre, considering a similar interpretation, suggests that "conditions of satisfaction must be stated from the *subject's* point of view" (McIntyre 1984, 472). However, when we deal with this notion in the context of Searle's theory of perception, we see that his use of this notion is equivalent to the third-person notion of state of affairs, rather than to the subject's view-point. Here, for Searle, "conditions of satisfaction are always that such and such is the case." In what follows, corresponding to the context in which Searle uses it, we will take the notion of conditions of satisfaction to be equivalent to the notion of state of affairs.

> I have a visual experience (that there is a yellow station wagon there and that there is a yellow station wagon there is causing this visual experience). (Searle 1983, 48)

The idea here is that, in the perceptual cases, it is necessary for the visual experience to be caused by the conditions of satisfaction of this very visual experience. Otherwise, the visual experience is not a perception;[2] it might be a hallucination, or another kind of misperception. And for Searle, that the visual experience must be caused by its conditions of satisfaction is part of the conditions of satisfaction of this visual experience; therefore, it has to be specified in the content of that experience.

3. Is the content of the visual experience equivalent to a proposition?

Searle's main reason to state that the content of visual experience is propositional is that it "[...] is an immediate (and trivial) consequence of the fact that they have conditions of satisfaction, for conditions of satisfaction are always that such and such is the case" (Searle 1983, 41). As mentioned above, the Intentional content, for Searle, always requires a state of affairs for its satisfaction (cf. Searle 1983, 41).

---

[2] Searle makes a distinction, which seems to me superfluous, between the visual experience and the perception. According to this distinction, the perception involves the notion of succeeding. The visual experience, however, might be unsuccessful either. The hallucination or illusion, for example, do possess a visual experience, but these are not perception cases. Yet this distinction violates the intuitive idea that Intentional states with the direction of fit must have two values regarding the satisfaction of their conditions; they can be either satisfied or unsatisfied. The beliefs, for example, can be either true or false. According to Searle's visual experience/perception distinction, visual experiences can have these two values, but perceptions, in order to be perceptions, must only be satisfied; because, for Searle, when they are not satisfied, they are *not* unsatisfied perceptions, what they logically must be if they are Intentional state with a direction of fit, rather unsatisfied visual experiences. So it turns out that the perception must only be true of its conditions of satisfaction – which is logically not the case. Though Searle uses the mentioned notions interchangeably, this can show that the visual experience/perception distinction is superfluous.

Two understandings of this view can be put forward. The first understanding is possible in terms of the assumption that every particular object (since it has properties) can be treated as a state of affairs; the second can be suggested in terms of the basic characteristic of the situation of seeing. Let us begin with the first.

Suppose my seeing a station wagon. The station wagon I see has properties. It has a color, weight, a certain place in the space; it can be old, without repair, and so on. That is, the object seen is, so to speak, a bundle of different states of affairs. Nevertheless, from Searle's example we can see that the phrases *there* and *in front of me* refer to the space relation between particulars, between the subject and the object, rather than to any property of the object perceived. Therefore, to hold the first view would be incorrect.

What Searle means is presumably the second, which can be called the situation of seeing. It is simple to note that in the situation of seeing there must be at least two particulars – one of them must necessarily be the person who sees and the other(s) must be object(s) of seeing – and a spatial relation between them, so that the person can see the object(s). This is a sufficient reason to state that seeing requires a state of affairs, but to my mind, not sufficient to state that the Intentional content of seeing (or visual experience) requires a state of affairs. Because the notion of situation of seeing we use here is a notion of the third-person view, whereas the notion of Intentional content of the visual experience indicates a phenomenological fact, the person's mode of presentation of that situation, but not refer to it from the "outside".[3] Accordingly, they should have different specifications.

However, one might argue that, insofar as everybody has a belief that seeing entails the existence of an object in the field of vision, there must be an implicit sense in the content of every seeing that what is seen is always before the person experiencing this visual experience. Therefore, it would be reasonable to hold that the spatial indexicals such as "there" or "here" can be made explicit in the content of seeing. So, for example, it might be suggested that the content of my visual experience of a station wagon should be described by the noun phrase *a station wagon there*. Below, I will

[3] That might be seen as a bare stipulation, but Searle seems to agree with this, for he holds that he deals with "first-person Intentionality".

show that such a specification of the content would be incorrect as well. But now I want to emphasize that the specification of the content with the proposition *that there is a station wagon there* would be also indefinite. Because if the situation was described from the third-person view, we could hold that, for the description is the description of a state of affairs, it must be in the form of proposition. Yet if we hold that this is a description from "first-person Intentionality", then it must be committed to the person's view of the situation which, depending on the person, can be seen either as an object, or as a state of affairs. Thus, if the person sees the situation as a state of affairs, it is preferable to describe it with a proposition; if not, then, for it would be different way of seeing, the description should be conducted with a noun phrase. In other words, the specification of the content must be the specification of how the subject experiences it.

If we give up this thesis and hold the view that the specification of the content of seeing must be only propositional, then ambiguities in the description will be inevitable. Because from the intuitive level of experiencing our visual experiences we know that we should distinguish between *simple seeing* and *judgmental seeing*. This distinction

> [...] is evident from the fact that at any given moment we perceive many more objects, and features of objects, than we make judgments about. When I look out of my window at my garden and judge on the basis of what I see that the tree in my garden is blooming, I see at that moment other plants besides the tree. I also see at that moment many more features of that tree besides its blossoms, features, about which I judge (at that moment) nothing at all. (Miller 1984, 34)

To put it otherwise, in most cases when we have visual experiences, we do not judge; we simply perceive. A judgmental seeing, however, is entertained when we make judgments on the basis of our visual experiences which are concurrent with those judgments.

If we make such a distinction between simple seeing and judgmental seeing, it is simple necessity that descriptions of the content of these experiences must also be distinguished from each other. And it is useful to employ noun phrases for the first cases, and propositions for the second cases, in order to show these two different ways of seeing on the descriptive level.

Searle's neglection of this convention derives from the fact that, though he deals with "first-person Intentionality", he passes over the distinction

between simple seeing and judgmental seeing, which is an evident phenomenological fact, and specifies the contents of visual experiences with a whole proposition in terms of the third-person view, which does not allow noticing this distinction.

4. The causal self-referentiality as a feature of the psychological mode

Now let us consider Searle's thesis that the causal self-referentiality of the visual experience has to be specified in the content of that experience. For this purpose, let us put aside his view that the content of seeing is equivalent to a proposition. Because whether the content is propositional or not is irrelevant here; for both cases we observe the same fact: what is initially specified in the content is enriched by the contentual constituents which are step by step made explicit. For example, while seeing a station wagon, we first realize that the content of my visual experience is not simply *a station wagon*, but *a station wagon there* (or *there is a station wagon there*). Then, analyzing more deeply, it is figured out that the visual experiences have the feature of causal self-referentiality which should also be specified in the content.

Here I am going to show that the features of the visual experiences such as causal self-referentiality, or *thereness,* do not belong to the content and therefore cannot be specified in the content of the visual experiences.

The fact that does not let specify the content of my seeing *that there is such and such there* as *that there is such and such there and that there is such and such there is causing this visual experience* (cf. Searle 1983, 48) could be that the part of content which describes the causal self-referentiality of my visual experience is not accessible to my reflection on the content from the first person view. No matter how much I would reflect on my content of the visual experience, I cannot find in it "that there is a yellow station wagon there is causing this visual experience." This is evident from the fact that whenever we see we do not exercise the concept of the causality in the content of our experience.

Yet Searle himself foresees this problem. He adds that "[...] the sense in which the visual experience is self-referential is simply that it figures in its

own conditions of satisfaction", but "[...] not that it contains a verbal or other representation of itself [...]" (Searle 1983, 49). By this Searle means that the part specifying the causal self-referentiality in the content is not a representation as the other normal part. Rather, it is the specification of conditions of satisfaction what requires the causal self-referentiality to be added into the content.[4]

However, to my mind, if we hold that the content's function is to determine how the state exactly relates itself to the world, and accept the intuitive idea in terms of which is made a distinction between the content and psychological mode of the Intentional state – namely, the idea that, while the psychological mode is held, the content of the state can vary depending on constituents (and their structure order) of which it consists or vice versa – then the specification of the causal self-referentiality in the content seems to be incorrect.

To see this, let us pose a question: What makes the visual experience a psychologically distinct kind of the Intentional state? We can answer this question by observing that the Intentional state called the visual experience is intuitively distinct from other kinds of Intentional states. That is, as we have a visual experience, we experience typologically special Intentional state whose distinctness is evidently noticed from the first-person view when we compare it with the other kinds of Intentional states; in visual experiences we experience the Intentional object as sensuously self-given. We can represent the same object in different Intentional states; we can remember, desire it (cf. Husserl 1970, LI V, §20). However, it is easy to observe that when we give up seeing and change the psychological mode, though the content can remain, the self-givenness of the object also disappears.

It is the feature of self-givenness of the object what compels philosophers to specify the visual experience by distinct ways. Most philosophers, including Searle – as we have seen – specify this basic feature of seeing with using different indexicals in the content part of the specification. D.W. Smith, for example, uses the indexicals *this* or *that*, arguing that the content of seeing has a demonstrative element. By this he means "[...] that feature of a visual experience – that part of its intentional character – which

[4] This view is the continuation of Searle's idea that the Intentional content and the conditions of satisfaction have the same specification (cf. Searle 1983, 13).

consists in its being a presentation of a particular object visually before the subject" (Smith 1989, 41). D.W. Smith mentions two features which for him are derived from the demonstrative content: 1) "a *singular* presentation of a particular object 'itself'"; and 2) "a presentation of a particular object *sensuously before* the subject" and its "causing this very experience." In other words, the demonstrative content *this* engenders two features which is described by still two indexicals: *itself* for the singularity, and *there* (or *here*) for an object's being before the subject so that the object can cause this experience. If we explicate all this features, the content of our visual experience of a station wagon, according to Smith, will be: *this* station wagon *itself there and causing this very experience* (cf. Smith 1989, 45).

However, what Searle firstly emphasizes is not *that-* or *thisness* of the content, but *thereness* of the object, which, according to Smith, is derivable from *thisness* of seeing. But at any rate, for Searle, the causal feature is also derivable from *thereness*, in other words, from the fact that the object seen is sensuously before the subject. In a nutshell, both philosophers agree that the indexical feature of the visual experience should be specified in the content. That seems to me unjustifiable.

One basic reason that makes me suspicious of the specification of *thereness* in the content of the visual experience is its relatedness or reducibility to the self-consciousness. Self-consciousness "happens for the experiencing subject in an immediate way and as part of this immediacy, it is implicitly marked as my experience" (Gallagher – Zahavi 2005). However, it is the easiest explicable feature of Intentional states. If I make explicit the content of my visual experience together with the self-consciousness, I would have to use the proposition *I see a station wagon*, which refers to "a whole state of affairs". Now, if the causal feature is reducible to *thereness*, considering that the situation of seeing involves at least two particulars (subject and object) and a spatial relation between them, then one can assimilate *that I see a station wagon* into *that there is a station wagon there*. Because in the latter proposition *a station wagon* refers to the object seen, but the subject – I – is hidden or implicit under the spatial indexical *there*. In other words, *there* here refers to a certain place which is only *there* from the perspective of the person who sees that object. It is in virtue of the self-consciousness that the subject is conscious of the object's being in front of her, or *there,* in the visual experience. If there were no self-consciousness, then, whether it be implicit or explicit, she could not have the sense *there*. Therefore, *there*

necessarily implies the second particular as an implicitly self-conscious subject.

This speculation shows that *thereness* and the causal feature of seeing is derived from deeper structures of consciousness and that is why it cannot be made explicit in the specification of Intentional content, which is the surface, more vulnerable to changes, aspect of consciousness with the function to determine how the Intentional state exactly relates itself to the world at the given moment. *Thisness* or *thereness,* as well as causal self-referentiality, however, are features that, independent of the content, belong to each satisfied visual experience; if one has a satisfied visual experience, she has these features necessarily. They always recur regardless of the content, making that kind of Intentional states be identical to themselves (from the standpoint of their psychological mode).

The other reason which does not allow specifying *thereness* and the causal feature in the content, like in the case of propositional/noun phrase specification of the content, is related to the conventional side of the description. Namely, if we specify the features of seeing in the content, then the ambiguities can take place. Suppose a wagon station to be near on my road to the railway station and suppose that it is first time that I see it. While seeing it, I simply pay attention to its being a station wagon. If somebody asked me "what do you see?" my answer would be "I see a station wagon" or "I see that this is a station wagon"; because the other properties of the station wagon are irrelevant for me. Now suppose another case where, while seeing a station wagon, I am interested in seeing a station wagon with considering its place, but I have no any description for its place. Then the content of my visual experience would be *a station wagon there*. Now, this *there* is distinct from Searle's (or Smith's) specification of *there* in the content in the sense that it is taken to be explicit together with *a station wagon*, but is not made explicit in terms of the philosophical meditation on the visual experience as a kind of Intentional state. Yet, by specifying the content with *there* for all (re)presentationally different cases of visual experience of the same object, we cannot distinguish whether *there* in the content expresses the explicitness of subject's considering the place of the object or is made explicit by the meditation as an implicit feature which is general for all visual experiences independent of what the content explicitly represents. Hence, if we want to make explicit these features, it would be correct to ascribe the part specifying *thereness,* or the causal self-

referentiality (*...that there is such and such there is causing this visual experience*), not to the content of the state but to its psychological mode. However, since the verbs that stand for the psychological mode perform this function, we do not usually do such specifications. When I specify my seeing an object, say, as *I see that this is such and such,* the verb *see* here, for it refers to the psychological mode of my Intentional state, stands for the feature of *thereness* and the causal self-referentiality as well. To put it otherwise, to use the relevant (verb) expressions as shorthand for the psychological modes, or for their features, together with the expressions specifying the content, is sufficient for the whole specification of the visual experience.

5. Conclusion

I have considered that Searle gives up the first-person Intentionality when he analyses the specification of Intentional content of the visual experience and therefore bypasses the first-person important distinction between simple seeing and judgmental seeing. The specification of the content only in the form of the sentence expressing proposition does not allow making such a distinction on the level of description. Then I have argued that the causal self-referentiality and the other features of the visual experience such as *thereness* belong to its psychological mode but not to the Intentional content of the visual experience.

References

GALLAGHER, S. – ZAHAVI, D. (2005): Phenomenological Approaches to Self-Consciousnes. In: Zalta, E.N. (ed.): *Stanford Encyclopedia of Philosophy*. Available at: http://plato.stanford.edu/entries/self-consciousness-phenomenological/.

HUSSERL, E. (1970): *Logical Investigations.* Vol. 2. Translated by J. N. Findlay. London: Routledge.

MCINTYRE, R. (1984): Searle on Intentionality. *Inquiry* 27, 468-483.

MILLER, I. (1984): *Husserl, Perception, and Temporal Awareness.* Cambridge: The MIT Press.

MULLIGAN, K. (1995): Perception. In: Smith, B. – Smith, D.W. (eds.): *The Cambridge Companion to Husserl.* Cambridge: Cambridge University Press, 168-238.

SEARLE, J.R. (1983): *Intentionality: An Essay in the Philosophy of Mind*. New York: Cambridge University Press.
SMITH, D.W. (1989): *The Circle of Acquaintance. Perception, Consciousness, and Empathy*. Dordrecht: Kluwer Academic Publishers.

Organon F 22 (3) 2015: 358-377

The Semantics of Empirical Unverifiability

IGOR SEDLÁR[1]

Institute of Philosophy. Slovak Academy of Sciences
Klemensova 19. 811 09 Bratislava. Slovakia
igor.sedlar@savba.sk

RECEIVED: 08-04-2015 • ACCEPTED: 08-06-2015

ABSTRACT: Pavel Cmorej has argued that the existence of unverifiable and unfalsifiable empirical propositions follows from certain plausible assumptions concerning the notions of possibility and verification. Cmorej proves, it the context of a bi-modal alethic-epistemic axiom system AM4, that (1) p *and it is not verified that* p is unverifiable; (2) p *or it is falsified that* p is unfalsifiable; (3) every unverifiable p is logically equivalent to p *and it is not verifiable that* p; (4) every unverifiable p entails that p is unverifiable. This article elaborates on Cmorej's results in three ways. Firstly, we formulate a version of neighbourhood semantics for AM4 and prove completeness. This allows us to replace Cmorej's axiomatic derivations with simple model-theoretic arguments. Secondly, we link Cmorej's results to two well-known paradoxes, namely Moore's Paradox and the Knowability Paradox. Thirdly, we generalise Cmorej's results, show them to be independent of each other and argue that results (3) and (4) are independent of any assumptions concerning the notion of verification.

KEYWORDS: Completeness – epistemic logic – knowability– verifiability.

1. Introduction

Cmorej (1988; 1990) argues that the existence of unverifiable and unfalsifiable empirical propositions is a consequence of certain plausible

[1] This work has been supported by the VEGA grant no. 2/0019/12, *Language and the Determination of Meaning in Communication*. I am grateful to Pavel Cmorej for clarification, comments and encouragement.

assumptions concerning the notions of possibility and verification.[2] His argument is proof-theoretic and employs an alethic-epistemic axiom system. Cmorej's main result is that schemas

(1) $\sim MV\,(\alpha \wedge \sim V\alpha)$

(2) $\sim MF\,(\alpha \vee F\alpha)$,

are provable in the axiom system in question, where M stands for 'it is possible that', V stands for 'it is verified that' and F stands for 'it is falsified that' ($F\alpha$ is defined as $V\sim\alpha$). If α is a hitherto unverified empirical proposition, then $\alpha \wedge \sim V\alpha$ is empirical as well. Yet, according to (1), it is unverifiable. Similarly, if α is empirical and not falsified, then $\alpha \vee F\alpha$ is empirical and, according to (2), not falsifiable.

Cmorej then goes on to establish two further results concerning unverifiable propositions. Firstly, each unverifiable proposition α is necessarily equivalent to $\alpha \wedge \sim V\alpha$. In other words,

(3) $\sim MV\alpha \supset L\left(\alpha \equiv (\alpha \wedge \sim V\alpha)\right)$

is provable (where L stands for 'it is necessary that'). Secondly, each unverifiable proposition α entails a proposition saying that α is unverifiable, i.e.

(4) $\sim M\,V\,\alpha \supset L(\,\alpha \supset \sim MV\alpha)$

is provable. (Similar results are established for falsifiability, but these are easily derivable form the results stated above by applying the definition of F.)

This article elaborates on Cmorej's results and sets them into a wider philosophical context. Firstly, Cmorej's arguments are simplified by replacing the complex axiomatic proofs of the results concerning (1) – (4) by simple model-theoretic arguments. Secondly, Cmorej's result concerning (1) is linked to two well-known paradoxes, namely Moore's Paradox (Green – Williams 2007; Moore 1942) and the Knowability Paradox (Fitch 1963; Salerno 2009). Thirdly, the results are generalised and shown to be independent. In particular, we set up a weak bi-modal logic that validates

---

2 Cmorej (1990) is a translation of the Slovak original Cmorej (1988). I'll refer to the internationally accessible Cmorej (1990) for the rest of the article.

(1) and (2) without validating (3), (4) and most Cmorej's assumptions concerning V and M. We also formulate a bi-modal logic that validates (3) and (4) without validating (1) or (2). The precise nature of the latter logic suggests that the results concerning (3) and (4) are independent of any assumptions concerning the notion of verification.

The article is organised as follows. Section 2 introduces Cmorej's axiom system AM4 and establishes completeness with respect to a specific class of modal neighbourhood frames (Chellas 1980; Segerberg 1971). This allows us to formulate simple model-theoretic arguments establishing (1) – (4). Section 3 relates Cmorej's result concerning (1) to Moore's Paradox and the Knowability Paradox. Section 4 shows that the results concerning (1) and (2) are independent from the results concerning (3) and (4), and that the latter two are independent of any assumptions concerning the notion of verification. The final section sums up the main points of the article.

2. Semantic arguments

This section introduces the axiom system AM4 (see 2.1.), discusses models (2.2.), proves completeness (2.3.) and provides simple model-theoretic arguments establishing (1) – (4) (2.4.).

2.1. AM4

Let us fix a denumerable set Var of propositional variables. Every propositional variable $p, q, \ldots$ is a *formula*. If α and β are formulas, then so are $\sim\alpha, \alpha \wedge \beta, L\alpha$ and $V\alpha$. Other Boolean connectives are defined in the usual fashion. $L\alpha$ is read as 'it is necessary that α' (or 'α is necessary') and $V\alpha$ as 'it is verified that α' (or 'α is verified'). $M\alpha$ is defined as $\sim L\sim\alpha$ and is read as 'it is possible that α' (or 'α is possible'). $F\alpha$ is defined as $V\sim\alpha$ and is read as 'it is falsified that α' (or 'α is falsified'). A formula is *tautologous* if it is a substitution instance of a tautology of classical propositional logic.

Definition 2.1 (AM4, Cmorej 1990). The axiom system AM4 is given by the following axiom schemas and rules of inference. Every tautologous formula is an axiom. Other axioms are all formulas of the form:

(A1) $L\alpha \supset \alpha$

(A2) $L(\alpha \supset \beta) \supset (L\alpha \supset L\beta)$

(B2) $V(\alpha \wedge \beta) \supset (V\alpha \wedge V\beta)$

(B3) $(V\alpha \wedge V\beta) \supset V(\alpha \wedge \beta)$

(A3) $\sim L\alpha \supset L{\sim}L\alpha$ (B4) $V\alpha \supset VV\alpha$

(B1) $V\alpha \supset \alpha$ (C) $L(\alpha \supset \beta) \supset (V\alpha \supset V\beta)$

There are two rules of inference, namely Modus Ponens and L-Necessitation ('If $\vdash \alpha$, then $\vdash L\alpha$'). Proofs and derivations are defined as usual. ■

The choice of 'alethic' L-axioms and rules and 'methodological' V-axioms makes it clear that L is a *normal* modality governed by axioms of the system S5 (see Hughes – Cresswell 1996), while V is a *regular* modality governed at least by the axioms of the system RT4 (see Chellas 1980). We shall see later on that, in fact, V is a non-normal modality as the rule of V-Necessitation is not a derivable rule. In other words, verification is not closed under admissible zero-premise inference rules. However, as the 'interaction axiom' (C) suggests, verification is closed under admissible one-premise rules. In fact, a consequence of the inclusion of (B3) among axioms entails that verification is closed under admissible multi-premise rules as well.

Lemma 2.2. *$L\alpha \supset LL\alpha$ is derivable in AM4.*

Proof. Folklore (see Hughes – Cresswell 1996, 58).□

Lemma 2.3. *If $\alpha \equiv \beta$ is provable in AM4, then so is $V\alpha \equiv V\beta$.*

Proof. We make use of some obviously admissible S5-rules. If $\vdash \alpha \equiv \beta$, then $\vdash L(\alpha \equiv \beta)$, then $\vdash L(\alpha \supset \beta) \wedge L(\beta \supset \alpha)$, then $\vdash (V\alpha \supset V\beta) \wedge (V\beta \supset V\alpha)$. □

2.2. Models

The models of our choice are neighbourhood models, where neighbourhoods (to be defined shortly) are closed under intersection. The assumption of closure under supersets, standard when regular systems are dealt with, is simulated by a non-standard truth-condition for $V\alpha$. L is treated as a universal modality.

Definition 2.4 (Frames). A *frame* is a couple

$\mathcal{F} = \langle \mathcal{W}, \mathcal{N} \rangle$,

where $\mathcal{W}$ is a non-empty set ('states' or '(possible) worlds') and $\mathcal{N}$ is a function from $\mathcal{W}$ to subsets of the power-set of $\mathcal{W}$ ('neighbourhood function'). Hence, $\mathcal{N}(w)$ is a set of sets of worlds ('neighbourhoods of w'). It is assumed that

- (c) If $X, Y \in \mathcal{N}(w)$, then $X \cap Y \in \mathcal{N}(w)$;
- (t) If $X \in \mathcal{N}(w)$, then $w \in X$;
- (iv) If $X \in \mathcal{N}(w)$, then $\{v \mid X \in \mathcal{N}(v)\} \in \mathcal{N}(w)$. ■

Sets $X \in \mathcal{N}(w)$ can be thought of as propositions 'directly' verified at w. The assumption (c) guarantees that (B3) is valid in every frame (to be defined shortly); (t) ensures (B1) and (iv) ensures (B4), (see Chellas 1980).

Definition 2.5 (Models and Truth-Sets). A *model* based on $\mathcal{F}$ is a couple

$\mathcal{M} = \langle \mathcal{F}, \mathcal{V} \rangle$,

where $\mathcal{V}$ is a function from Var to subsets of $\mathcal{W}$ ('valuation'). The *truth-set* $|\alpha|_{\mathcal{M}}$ of a formula α in model $\mathcal{M}$ is defined recursively as follows:

- $|p|_{\mathcal{M}} = \mathcal{V}(p)$;
- $|{\sim}\alpha|_{\mathcal{M}} = \mathcal{W} \setminus |\alpha|_{\mathcal{M}}$;
- $|\alpha \wedge \beta|_{\mathcal{M}} = |\alpha|_{\mathcal{M}} \cap |\beta|_{\mathcal{M}}$;
- $|V\alpha|_{\mathcal{M}} = \{w \mid X \subseteq |\alpha|_{\mathcal{M}} \text{ for some } X \in \mathcal{N}(w)\}$;
- $|L\alpha|_{\mathcal{M}} = \mathcal{W}$ if $|\alpha|_{\mathcal{M}} = \mathcal{W}$; $|L\alpha|_{\mathcal{M}} = \emptyset$ otherwise. ■

$w \in |\alpha|_{\mathcal{M}}$ is read as 'α is true in w (in the context of $\mathcal{M}$)'. This will be written also as $\mathcal{M}, w \vDash \alpha$. Informally, $V\alpha$ is true in w iff there is a proposition directly verified at w that 'entails' α. $L\alpha$ is true at any world iff α is true in every world. $\mathcal{M}$ will not be mentioned when the identity of the model in question is clear from the context or immaterial.

Definition 2.6 (Consequence). α is a $\mathcal{M}$*-consequence* of a set of formulas Γ iff

$$\bigcap_{\beta \in \Gamma} |\beta|_{\mathcal{M}} \subseteq |\alpha|_{\mathcal{M}}$$

('Γ $\mathcal{M}$-entails α'). α is $\mathcal{M}$-valid iff it is an $\mathcal{M}$-consequence of the empty set. α is a $\mathcal{F}$-*consequence* of Γ iff it is a $\mathcal{M}$-consequence of Γ for every $\mathcal{M}$ based on $\mathcal{F}$. If $\mathcal{C}$ is a class of frames (models), then α is a $\mathcal{C}$-consequence of Γ iff it is an $\mathcal{F}$-consequence ($\mathcal{M}$-consequence) of Γ for every $\mathcal{F}$ ($\mathcal{M}$) in $\mathcal{C}$. Similarly for $\mathcal{C}$-validity. ■

Γ $\mathcal{M}$-entails α iff there is no world in $\mathcal{M}$ where all the 'assumptions' in Γ are true, but α is false. α is $\mathcal{M}$-valid iff it is true 'throughout the model $\mathcal{M}$'.

Example 2.7. Let us consider an example. Let the set of worlds be $\{v, u\}$ and assume that the truth-set of p is $\{v\}$, while the truth-set of q is $\{v, u\}$. In addition, let $\mathcal{N}(v) = \{\{v\}\}$ and $\mathcal{N}(u) = \emptyset$. It is easy to check that this model satisfies the conditions (c), (t) and (iv). q is valid in the model and, hence, Lq holds in both worlds. So does $\sim L(p \wedge q)$. Vq holds in v, because $\{v\} \in \mathcal{N}(v)$ and $\{v\} \subseteq \{v, u\}$, the truth-set of q. However, Vq does not hold in u. Note that even $V{\sim}q$ is false in u. The truth-set of $\sim q$ is $\emptyset$, but, obviously, $\emptyset \notin \emptyset$. In fact, $\sim V\alpha$ holds in u for every formula α. In conjunction with our completeness proof of Section 2.3, this example shows that Cmorej's V is not a normal modality (as such, it would have to satisfy V-necessitation). ■

Neighbourhood semantics has a wide range of applications, including models of coalitions within games (Pauly 2002). Neighbourhood models have recently been applied to an epistemic language with both normal an non-normal modalities within the project of *evidence logics* (see van Benthem – Fernández-Duque – Pacuit 2014; van Benthem – Paciut 2011). In view of our completeness result established below, Cmorej may be credited with an early contribution to evidence logic.

2.3. Completeness

The goal of the present subsection is to show that α is derivable from a set of assumptions Γ in AM4 iff Γ $\mathcal{F}$-entails α in every $\mathcal{F}$. One half of the claim is established easily.

Proposition 2.8 (Soundness). *If α is derivable from a set of assumptions Γ in AM4, then Γ $\mathcal{F}$-entails α in every $\mathcal{F}$.*

Proof. It is sufficient to show that every axiom is valid in every frame and that the rules of inference preserve validity. All cases are straightforward. Nevertheless, let us prove the validity of (B3) and (B4). First, (B3). Consider any $\mathcal{M}, w$. If $\mathcal{M}, w \vDash V\alpha \cap V\beta$, then there is $X \in \mathcal{N}(w)$ such that $X \subseteq |\alpha|$ and there is $Y \in \mathcal{N}(w)$ such that $Y \subseteq |\beta|$. But then $X \cap Y \in \mathcal{N}(w)$ by (c). Obviously, $X \cap Y \subseteq |\alpha \wedge \beta|$. Hence, $\mathcal{M}, w \vDash V(\alpha \wedge \beta)$. Next, (B4). If $\mathcal{M}, w \vDash V\alpha$, then there is $X \in \mathcal{N}(w)$ such that $X \subseteq |\alpha|$. By (iv), $\{v \mid X \in \mathcal{N}(v)\} \in \mathcal{N}(w)$. It is plain that $\{v \mid X \in \mathcal{N}(v)\} \subseteq |V\alpha|$. In other words, there is $Y \in \mathcal{N}(w)$ such that $Y \subseteq |V\alpha|$. Con-sequently, $\mathcal{M}, w \vDash VV\alpha$. □

To establish the other half of the main claim, we employ the standard canonical model technique; see (Chellas 1980), for example. A specific feature of our situation is the presence of the universal modality L. To deal with this extra machinery, we combine the standard completeness argument for regular systems with a simple strategy that is used within completeness proofs for normal systems with the universal modality (see Blackburn – de Rijke –Venema 2001, ch. 7.1). But first, let us re-capitulate some standard terminology.[3]

Definition 2.9 (AM4-sets). A set Γ of formulas is *maximal AM4-consistent* ('an AM4-set') iff

- Γ is *consistent*, i.e. there is no $\{\alpha_1, \ldots, \alpha_n, \beta\} \subseteq \Gamma$ such that $\alpha_1 \wedge \ldots \wedge \alpha_n \supset \sim\beta$ is provable in AM4; and
- Γ is *maximal*, i.e. if $\alpha \notin \Gamma$, then $\Gamma \cup \{\alpha\}$ is not consistent. ■

Lemma 2.10. *Some well-known properties of maximal consistent sets:*

- *If* Γ *is an AM4-set,* $\Delta \subseteq \Gamma$ *and* α *is derivable form* Δ *in AM4, then* $\alpha \in \Gamma$*;*
- *If* Δ *is consistent then there is an AM4-set* Γ *such that* $\Delta \subseteq \Gamma$ *(Lindenbaum's Lemma);*

Proof. Standard (see Chellas 1980, ch. 2.6). □

[3] More details on maximal consistent sets and modal completeness proofs are provided by Blackburn et al. (2001, ch. 4), Chellas (1980, chs. 2.6-2.7, 4.5, 5.3) and Hughes – Cresswell (1996, ch. 6), who discuss normal systems. Chellas (1980, ch. 9) discusses completeness proofs for some non-normal systems.

Note that the above Lemma entails that if Γ is an AM4-set, then $\sim\alpha \in \Gamma$ iff $\alpha \notin \Gamma$ and $\alpha \wedge \beta \in \Gamma$ iff $\alpha, \beta \in \Gamma$.

Definition 2.11 (Pre-models). A *pre-model* is a tuple

$\mathcal{M}_0 = \langle \mathcal{W}_0, \mathcal{R}, \mathcal{N}_0, \mathcal{V}_0 \rangle$

where

- $\mathcal{W}_0$ is the set of all AM4-sets, and $|\alpha|_0 = \{\Gamma \in \mathcal{W}_0 \mid \alpha \in \Gamma\}$;
- $\mathcal{R}\Gamma\Delta$ iff $\{\alpha \mid L\alpha \in \Gamma\} \subseteq \Delta$, and $\mathcal{R}(\Gamma) = \{\Delta \mid \mathcal{R}\Gamma\Delta\}$;
- $\mathcal{N}_0(\Gamma) = \{|\alpha|_0 \mid V\alpha \in \Gamma\}$;
- $\mathcal{V}_0(p) = |p|_0$. ■

Lemma 2.12. *For all* $\Gamma \in \mathcal{W}_0, \mathcal{N}_0(\Gamma)$ *is closed under (binary) intersections.*

Proof. Assume that $X, Y \in \mathcal{N}_0(\Gamma)$. By the definition of $\mathcal{N}_0, X = |\alpha|_0$ and $Y = |\beta|_0$ for some $V\alpha, V\beta \in \Gamma$. By Lemma 2.10, $V(\alpha \wedge \beta) \in \Gamma$. Hence, $|\alpha \wedge \beta|_0 \in \mathcal{N}_0$. In other words, $|\alpha|_0 \cap |\beta|_0 \in \mathcal{N}_0(\Gamma)$. ■

Lemma 2.13. *If* $\Gamma \in \mathcal{W}_0$ *and* $L\alpha \notin \Gamma$*, then there is* $\Delta \in \mathcal{W}_0$ *such that*

- $\mathcal{R}\Gamma\Delta$ *and*
- $\sim\alpha \in \Delta$.

Proof. Standard (see Hughes – Cresswell 1996, 115-117). □

It is clear that, in pre-models, we can have some Γ, Δ, α such that $L\alpha \in \Gamma$, but $\alpha \notin \Delta$ (if not $\mathcal{R}\Gamma\Delta$). Hence, in the context of pre-models, L is not a universal modality. To fix this, we use a standard 'trick'.

Definition 2.14 (Canonical Λ-model). Let $\Lambda \in \mathcal{W}_0$. A canonical Λ-model is a tuple

$\mathcal{M}_\Lambda = \langle \mathcal{W}_\Lambda, \mathcal{N}_\Lambda, \mathcal{V}_\Lambda \rangle$

where

- $\mathcal{W}_\Lambda = \mathcal{R}(\Lambda)$ and $|\alpha|_\Lambda = |\alpha|_0 \cap \mathcal{W}_\Lambda$;
- $\mathcal{N}_\Lambda(\Gamma) = \{X \subseteq \mathcal{W}_\Lambda \mid X = X_0 \cap \mathcal{W}_\Lambda \text{ for some } X_0 \in \mathcal{N}_0(\Gamma)\}$;
- $\mathcal{V}_\Lambda(p) = |p|_\Lambda$. ■

Λ is seen as the 'centre' of the model, which universe is the set of AM4-sets reachable from the centre via $\mathcal{R}$. Crucially, every Λ-neighbourhood of any $\Gamma \in \mathcal{W}_0$ is a 'pre-neighbourhood' of Γ with every AM4-set *not* reachable from the centre 'bitten off'.

Now the goal is to show that, for every Λ, $\mathcal{M}_\Lambda$ is indeed a model.

Lemma 2.15. $|\alpha|_\Lambda \subseteq |\beta|_\Lambda$ *iff* $L(\alpha \supset \beta) \in \Lambda$.

Proof. We omit the simple argument establishing the right-to-left direction. To prove the converse, assume that $L(\alpha \supset \beta) \notin \Lambda$. By Lemma 2.13, there is $\Gamma \in \mathcal{R}(\Lambda)$ such that $\alpha \in \Gamma$ and $\beta \notin \Gamma$. Hence, $\Gamma \in |\alpha|_0 \cap \mathcal{W}_\Lambda$, but $\Gamma \notin |\beta|_0 \cap \mathcal{W}_\Lambda$. In other words, $|\alpha|_\Lambda \not\subseteq |\beta|_\Lambda$. □

Lemma 2.16. *If* $\Gamma \in \mathcal{W}_\Lambda$ *and* $L\alpha \in \Lambda$, *then* $L\alpha \in \Gamma$.

Proof. Follows from Lemmas 2.2 and 2.10. □

Lemma 2.17 (Frame Lemma). *For all* $\Lambda \in \mathcal{W}_0$ *and* $\Gamma \in \mathcal{W}_\Lambda$:

- (*c*) *If* $X, Y \in \mathcal{N}_\Lambda(\Gamma)$, *then* $X \cap Y \in \mathcal{N}_\Lambda(\Gamma)$;
- (*t*) *If* $X \in \mathcal{N}_\Lambda(\Gamma)$, *then* $\Gamma \in X$;
- (*iv*) *If* $X \in \mathcal{N}_\Lambda(\Gamma)$, *then* $\{\Delta \mid X \in \mathcal{N}_\Lambda(\Delta)\} \in \mathcal{N}_\Lambda(\Gamma)$.

Proof. (c) Assume that $X, Y \in \mathcal{N}_\Lambda(\Gamma)$. Then $X = |\alpha|_\Lambda$ and $Y = |\beta|_\Lambda$ for some $V\alpha, V\beta \in \Gamma$. By Lemma 2.12, $|\alpha|_0 \cap |\beta|_0 \in \mathcal{N}_0(\Gamma)$. Hence, $X \cap Y = |\alpha|_0 \cap |\beta|_0 \cap \mathcal{W}_\Lambda \in \mathcal{N}_\Lambda(\Gamma)$.

(t) Assume that $X \in \mathcal{N}_\Lambda(\Gamma)$. Then $X = |\alpha|_0 \cap \mathcal{W}_\Lambda$ for some $V\alpha \in \Gamma$. By Lemma 2.10 and axiom (B1), $\alpha \in \Gamma$, i.e. $\Gamma \in |\alpha|_0$. Consequently, $\Gamma \in X$.

(iv) Assume that $X \in \mathcal{N}_\Lambda(\Gamma)$. Then $X = |\alpha|_0 \cap \mathcal{W}_\Lambda$ for some $V\alpha \in \Gamma$. By Lemma 2.10 and axiom (B4), $VV\alpha \in \Gamma$. Now assume that $\{\Delta \mid X \in \mathcal{N}_\Lambda(\Delta)\} \notin \mathcal{N}_\Lambda(\Gamma)$. This means that $\{\Delta \mid X \in \mathcal{N}_\Lambda(\Delta)\} \neq |\beta|_0 \cap \mathcal{W}_\Lambda$ for no $V\beta \in \Gamma$. In particular, then, this holds for $VV\alpha$. In other words,

$$\{\Delta \mid X \in \mathcal{N}_\Lambda(\Delta)\} \neq |V\alpha|_0 \cap \mathcal{W}_\Lambda$$

Now there are two cases to check.

1. There is $\Delta \in \mathcal{W}_\Lambda$ such that $\Delta \in |V\alpha|_0 \cap \mathcal{W}_\Lambda$ but $|V\alpha|_0 \cap \mathcal{W}_\Lambda \notin \mathcal{N}_\Lambda(\Delta)$. The latter means that $|\alpha|_\Lambda \neq |\beta|_\Lambda$ for no $V\beta \in \Delta$. But $V\alpha \in \Delta$, so the assumption entails that $|\alpha|_\Lambda \neq |\alpha|_\Lambda$. Contradiction.

2. There is $\Delta \in \mathcal{W}_\Lambda$ such that $|V\alpha|_0 \cap \mathcal{W}_\Lambda \in \mathcal{N}_\Lambda(\Delta)$ but $\Delta \notin |V\alpha|_0 \cap \mathcal{W}_\Lambda$. In other words, $|\alpha|_\Lambda = |\beta|_\Lambda$ for some $V\beta \in \Delta$, but $V\alpha \notin \Delta$. The former entails, by Lemma 2.15, that $L(\alpha \supset \beta) \wedge L(\beta \supset \alpha) \in \Lambda$. By Lemma 2.16, $L(\beta \supset \alpha) \in \Delta$. But then, by Lemma 2.10 and axiom (C), $V\beta \supset V\alpha \in \Delta$. Consequently, $V\alpha \in \Delta$. Contradiction. □

Lemma 2.18 (Model Lemma). *For all* $\Lambda \in \mathcal{W}_0$ *and* $\Gamma \in \mathcal{W}_\Lambda, \alpha \in \Gamma$ *iff* $\mathcal{M}_\Lambda, \Gamma \vDash \alpha$.

Proof. We need to check that $\alpha \in \Gamma$ iff the truth-condition for α is satisfied with respect to Γ. The proof is by induction on the complexity of α. The base case $\alpha = p$ holds by definition. The cases of $\sim$ and $\wedge$ are easy (and standard) and we omit them. Only the 'modal' cases are checked explicitly.

We check that $L\alpha \in \Gamma$ iff $\mathcal{M}_\Lambda, \Delta \vDash \alpha$ for all $\Delta \in \mathcal{W}_\Lambda$. The right-hand side is equivalent to the claim that $\alpha \in \Delta$ for all $\Delta \in \mathcal{W}_\Lambda$ by the induction hypothesis. Now the left-to-right implication is an obvious consequence the definition of $\mathcal{W}_\Lambda$. The right-to-left implication follows from Lemma 2.13 and the definition of W_Λ.

Next, we check that $V\alpha \in \Gamma$ iff there is an $X \in \mathcal{N}_\Lambda(\Gamma)$ such that $X \subseteq |\alpha|_\Lambda$. If $V\alpha \in \Gamma$, then $|\alpha|_0 \in \mathcal{N}_0(\Gamma)$ and, hence, $|\alpha|_0 \cap \mathcal{W}_\Lambda \in \mathcal{N}_\Lambda(\Gamma)$. Conversely, if there is $X \in \mathcal{N}_\Lambda(\Gamma)$ such that $X \subseteq |\alpha|_\Lambda$, then $X = |\beta|_\Lambda$ for some $V\beta \in \Gamma$. $V\alpha \in \Gamma$ follows by Lemmas 2.15 and 2.16. □

The Frame and Model Lemmas ensure that every canonical Λ-model is a model and that membership in Γ is equivalent to truth in Γ. Completeness follows immediately.

Theorem 2.19 (Strong Completeness). Let Θ be any set of formulas. If Θ $\mathcal{F}$-entails α for every $\mathcal{F}$, then α is derivable from Θ in AM4.

Proof. Assume that α is not derivable from Θ. Then the set $\Theta \cup \{\sim \alpha\}$ is consistent. By Lindenbaum's Lemma, there is an AM4-set $\Lambda \supseteq \Theta \cup \{\sim\alpha\}$. Construct the Λ-canonical model $\mathcal{M}_\Lambda$. By Lemmas 2.17 and 2.18, there is a model $\mathcal{M}$ (namely $\mathcal{M}_\Lambda$) and a world w (namely Λ) such that $\mathcal{M}, w \vDash \beta$ for every $\beta \in \Theta$, but $\mathcal{M}, w \nvDash \alpha$. Hence, Θ does not $\mathcal{F}$-entail α for all $\mathcal{F}$. □

2.4. Cmorej's Results, Semantically

A direct consequence of the Completeness Theorem is that Cmorej's results may be established by using simple model-theoretic arguments.

Assume that (1) is not provable. Then, by the Completeness Theorem, there is a model $\mathcal{M}$ and a world w such that $M\,V(\alpha \wedge \sim V\alpha)$ is true in w. But then, by the truth-condition for L, $V\,(\alpha \wedge \sim V\alpha)$ is true in some u in the model $\mathcal{M}$. Soundness and (B2) imply that $V\alpha \wedge V\sim V\alpha$ holds in u and (B1) leads to the contradiction that $V\alpha \wedge \sim V\alpha$ holds in u.

The provability of (2) is a direct consequence of the provability of (1). If the schema (1) is valid then so is $\sim MV(\sim\alpha \wedge \sim V\sim\alpha)$ and, by Lemma 2.3, $\sim MV\sim(\alpha \vee V\sim\alpha)$ is valid as well.

Now assume that (3) is false in some $\mathcal{M},w$. Hence, $L\sim V\,\alpha \wedge M\sim(\alpha \equiv (\alpha \wedge \sim V\alpha))$ is true in w. This means that there is some u in $\mathcal{M}$ such that $\sim V\alpha \wedge \sim(\alpha \equiv (\alpha \wedge \sim V\,\alpha))$ holds in u. But this is impossible, since the latter formula is a substitution instance of a contradiction of classical propositional logic.

Finally, assume that (4) is false in $\mathcal{M},w$. Then $L\sim V\alpha \wedge M(\alpha \wedge MV\,\alpha)$ in w. By Lemma 2.2 and Soundness, $L\sim V\,\alpha \wedge \alpha \wedge M\,V\,\alpha$ in u. Contradiction. The nature of the latter two arguments suggests that the results concerning (3) and (4) are independent of *any* assumptions concerning V. We will return to this point in Section 4.

3. Unverifiability, absurdity, and unknowability

This section links Cmorej's results to two well-known philosophical problems, Moore's Paradox and the Knowability Paradox. Our sole aim is to point out some similarities between Cmorej's findings and the two paradoxes without going into philosophical detail.

Cmorej's main result is that

(5) $p \wedge \sim V\,p$,

as well as all its substitution instances, is provably unverifiable. (5) is similar in form to so-called (omissive) *Moorean sentences*, i.e. sentences of the form

(6) p and I do not believe that p,

with 'I believe that' replaced by 'It is verified that'. Moorean sentences and the air of absurdity surrounding them are at the heart of a famous problem, known as *Moore's Paradox*. Green and Williams explain that

> G.E. Moore observed that to say, 'I went to the pictures last Tuesday but I don't believe that I did' would be 'absurd' (1942, 53). Over half a century later, such sayings continue to perplex philosophers and other students of language, logic, and cognition. On the one hand, such sayings seem distinct from semantically odd Liar-type sayings such as 'What I'm now saying is not true'. Unlike Liar-type sentences, what Moore said might be true: One can readily imagine a situation in which Moore went to the pictures last Tuesday but does not believe that he did so. On the other hand, it does seem absurd to *assert* a proposition while, with no apparent change of mind, or aside to a different audience, going on to deny that one believes it. It seems no less absurd to *judge true* the following proposition: *p and I do not believe that p*. (Green – Williams 2007, 3; original emphasis)

(5) may itself be labelled as 'absurd to utter' or 'absurd to judge true'. Assume that I assert that p and that p is not verified at the same time. It seems, then, that my assertion implies that it lacks appropriate grounds: If the assertion is true, then one of the statements being asserted is unverified. But on what grounds is it asserted, then?

Cmorej's result concerning (1) can be construed as providing an explanation of the air of absurdity surrounding (5): (5) is unverifiable and, therefore, un-X-able for every X that requires verification.[4] This explanation is similar in spirit to Hintikka's (1962, 52-54) solution to Moore's Paradox, who argues that it is impossible for the speaker to believe (6).

Nevertheless, belief may be thought to be far too distant in nature from verification to ground any comparisons of Cmorej's (5) to the Moorean (6). Verification, it might be argued, is closer to (empirical) *knowledge*. Hence, it may seem more plausible to construe (5) along the lines of

(7) p and it is not known that p

On this account, Cmorej's result implies that propositions of the form (7) are *unknowable*. This observation is, of course, at the heart of another

4 In the sense that if some p is X-ed then p is verified.

famous problem, the *Knowability Paradox* due to Frederic Fitch and Alonzo Church (see Fitch 1963; Salerno 2009). Its gist is that the plausible assumption that every truth is *knowable* entails the ridiculous conclusion that every truth is *known*. For assume that every truth is knowable. Then, given the fact that (7) is unknowable, (7) is false. In other words 'If p, then it is known that p' is true. But p is arbitrary, so the claim holds for every p, i.e. every truth is known.

4. Independence results

This section is devoted to showing that the results concerning (1) and (2) are independent of the results concerning (3) and (4), and that the latter two are independent of any assumptions concerning the notion of verification. Consequently, the results concerning (1) and (2) are generalised, i.e. shown to hold for weaker notions of verification, and the results concerning (3) and (4) are shown to hold for every unary operator in place of V whatsoever.

The results are established as follows. Firstly, in section 4.1 we formulate AM1, a bi-modal logic for L and V that is rather weaker than AM4, but validates (1) and (2) without validating (3) or (4). Secondly, in section 4.2 we formulate another bi-modal logic AM0 with some very weak assumptions concerning L and no assumptions concerning V at all, and show that the logic validates (3) and (4) without validating (1) or (2). Section 4.3 provides some additional remarks. We note that both AM1 and AM0 will be formulated semantically, i.e. as sets of formulas valid in a class of frames. Axiom systems will be mentioned, but completeness will not be proved. The reason is that both completeness arguments are simple exercises extending the standard completeness proofs for 'classical' logics (see Chellas 1980).

4.1. (1) and (2) without (3) or (4)

AM1 will be defined as a set of formulas valid in a special class of bi-neighbourhood frames. Hence, we shall use neighbourhood models where both operators L and V are given truth-conditions in terms of neighbourhood functions. As a result, L in AM1 is a non-normal modality.

Definition 4.1 (AM1-Frames and Models). An AM1-frame is a triple

$\mathcal{F} = \langle \mathcal{W}, \mathcal{N}_L, \mathcal{N}_V \rangle$

where $\mathcal{W}$ is a non-empty set (interpreted as before) and both $\mathcal{N}_L, \mathcal{N}_V$ are functions from $\mathcal{W}$ to subsets of the power-set of $\mathcal{W}$. It is assumed that (for all w)

- (l) $\mathcal{W} \in \mathcal{N}_L(w)$;
- (m) For all Z and all $X \in \mathcal{N}_V(w)$, if $X \subseteq Z$, then for all $Y \in \mathcal{N}_V(w)$, there is $u \in Y$ and some $U \in \mathcal{N}_V(u)$ such that $U \subseteq Z$, for some U.

An AM1-model is an AM1-frame with a valuation, i.e. $\mathcal{M} = \langle \mathcal{F}, \mathcal{V} \rangle$. Truth-sets are defined as before, with the exception of

$|L\alpha|_{\mathcal{M}} = \{ w \mid |\alpha|_{\mathcal{M}} \in \mathcal{N}_L(w) \}$.

($V\alpha$ is dealt with as before, but in terms of $\mathcal{N}_V$.) Validity is defined as usual. AM1 is the set of formulas valid in every AM1-frame. ■

$\mathcal{N}_L(w)$, the set of L-neighbourhoods of w, is seen as the set of propositions necessary at w. It is assumed only that the 'maximal proposition' $\mathcal{W}$ is always necessary (l). The condition (m) might seem confusing, but its role is made clear by the proof of the following fact.

Fact 4.2. *If* $\alpha \in$ AM1, *then* $L\alpha \in$ AM1. *Moreover, every formula of the form*

$V{\sim}V\,\alpha \supset {\sim}V\alpha$

belongs to AM1.

Proof. Assume that $\alpha \in$ AM1 and take any $\mathcal{M}, w$. It follows that $|\alpha|_{\mathcal{M}} = \mathcal{W}$. Consequently, $|\alpha|_{\mathcal{M}} \in \mathcal{N}_L(w)$ and, hence, $L\alpha$ is true in w. By propositional logic, $V{\sim}V\alpha \supset {\sim}V\,\alpha$ is equivalent to $V\alpha \supset {\sim}\,V{\sim}V\,\alpha$. Now assume that $\mathcal{M}, w \vDash V\alpha$. We have to show that $\mathcal{M}, w \vDash {\sim}V{\sim}V\alpha$. Assume that this is not the case (indirect assumption). The first assumption entails that there is $X \in \mathcal{N}_V(w)$ such that $X \subseteq |\alpha|$. The indirect assumption entails that there is $Y \in \mathcal{N}_V(w)$ such that $Y \subseteq |{\sim}V\alpha|$. In other words, for all $u \in Y$ and all $U \in \mathcal{N}_V(u)$, $U \nsubseteq |\alpha|$. But this is precisely the negation of our condition (m). □

It is easy to show that the only non-tautologous axiom schema of AM4 that belongs to AM1 is (B2). This is done by constructing countermodels for all other axiom schemas. We give one example and leave the rest to the reader as an exercise.

Example 4.3. Let $\mathcal{W} = \{v, u\}$ and $|p| = \{u\}$. Moreover, let $\{\{v, u\}, \{u\}\}$ $(\emptyset)$ by the value of $\mathcal{N}_L(x)$ $(\mathcal{N}_V(x))$ for every $x \in \mathcal{W}$. It is easily checked that both (l) and (m) are satisfied. Moreover, Lp holds in v. However, p is false in v. The axiom schema (A1) fails as p is necessary but not true in some world of some model. ■

To facilitate comparison with AM4, we state (without proof) the following axiomatization result.

Proposition 4.4. *AM1 is soundly and completely axiomatized by the following axiom system. Every tautologous formula is an axiom and, moreover, every formula of the form*

(B1') $V{\sim}V\,\alpha \supset {\sim}V\alpha$

(B2) $V(\alpha \wedge \beta) \supset (V\alpha \wedge V\beta)$

is an axiom as well. The rules of inference are Modus Ponens, L-Necessitation and

(RE) If $\vdash \alpha \equiv \beta$, *then* $\vdash X\alpha \equiv X\beta$, *where* X *is* L *or* V.

Note that (B1') is a weak version of the axiom (B1), which is stating that every verified proposition is true. (B1') requires only that every verified proposition *of the form* $\sim V\alpha$ be true. The main observation is that this suffices to validate (1) and (2), while there are AM1-countermodels to both (3) and (4).

Proposition 4.5. *(1) and (2) are valid in AM1, but (3) and (4) are not.*

Proof. (1) Fact 4.2 and propositional logic entail that

$(V\alpha \wedge V{\sim}V\alpha) \supset {\sim}(V\alpha \wedge V{\sim}V\alpha)$

belongs to AM1. But (B2) is valid and, hence,

$V(\alpha \wedge {\sim}V\alpha) \supset {\sim}V\,(\alpha \wedge {\sim}V\alpha)$

is in AM1, which, by propositional logic, means that $\sim V(\alpha \wedge \sim V\alpha)$ belongs to AM1. By Fact 4.2 again, $L\sim V(\alpha \wedge \sim V\alpha)$ belongs to AM1.

(2) From the validity of (1) by propositional logic and repeated applications of (semantic counterparts of) the rule (RE).

(3) Our countermodel is as follows. $\mathcal{W} = \{v, u\}$ and $\mathcal{N}_V(x) = \{\{v\}\}$ for all $x \in \mathcal{W}$; $\mathcal{N}_L(v) = \{\emptyset, \mathcal{W}\}$ and $\mathcal{N}_L(u) = \{\mathcal{W}\}$; $|p| = \{v\}$. It is readily seen that this is indeed an AM1-model (the key to (m) is that $\mathcal{N}_V(x)$ is the same singleton for all $x \in \mathcal{W}$). Obviously, $|Vp| = \mathcal{W}, |\sim Vp| = \emptyset$ and $|\sim(p \wedge V\ p)| = \{u\}$. Consequently, $L\sim Vp$ holds in v (as $\emptyset \in \mathcal{N}_L(v)$), but $L\sim(p \wedge Vp)$ does not hold in v (as $\{u\} \notin \mathcal{N}_L(v)$). But, as is easily checked, $L\sim Vp \wedge \sim L\sim(p \wedge Vp)$ entails the negation of (3).

(4) The countermodel is just like the countermodel to (3) except for $|p| = \{u, v\}$. It is easily checked that, as before, $|\sim Vp| = \emptyset$ and, moreover, $|\sim L\sim Vp| = \{u\}$. Hence, $|\sim(p \wedge \sim L\sim Vp)| = \{v\}$. But this means that, as before, $L\sim Vp$ holds in v. However, as $\{v\} \notin \mathcal{N}_L(v)$, $L\sim(p \wedge M\ Vp)$ is false in v. Consequently, (4) is false in v.□

Proposition 4.5 generalises Cmorej's results concerning (1) and (2). It shows that the original results can be obtained by building on assumptions concerning the notions of verification and necessity that are far weaker that the ones originally used by Cmorej. The second upshot is that the results concerning (1) and (2) are independent of those concerning (3) and (4). In other words, one may construe 'verified' and 'necessary' in such a manner that $\alpha \wedge \sim V\alpha$ turns out to be 'unverifiable' (and $\alpha \vee F\alpha$ to be 'unsatisfiable'), but not every 'unverifiable' α is logically equivalent to $\alpha \wedge \sim V\alpha$ and not every 'unverifiable' α entails a proposition that says that α is 'unverifiable'.

4.2. (3) and (4) without (1) or (2)

The logic AM0 is defined similarly as AM1.

Definition 4.6 (AM0-Frames and Models). An AM0-frame is a couple

$\mathcal{F} = \langle \mathcal{W}, \mathcal{N} \rangle$

where all the components are as before, but only one condition is enforced:

- (iv) If $X \in \mathcal{N}(w)$, then $\{ v \mid X \in \mathcal{N}(v) \} \in \mathcal{N}(w)$.

An AM0-model $\mathcal{M} = \langle \mathcal{F}, \mathcal{V} \rangle$, as before. The truth-sets for Boolean formulas are defined as usual. Moreover:

- $|V\alpha|_{\mathcal{M}}$ is arbitrary;
- $|L\alpha|_{\mathcal{M}} = \{w \mid X \subseteq |\alpha|_{\mathcal{M}}$ for some $X \in \mathcal{N}(w)\}$.

AM0 is defined as the set of formulas valid in every AM0-frame. ■

In AM0, L takes the place of V and is given a truth-condition in terms of a neighbourhood function. It is the same truth-condition that was given to V in the semantics for AM4, but fewer restrictions ale placed on $\mathcal{N}$. The absence of any specific truth-condition for formulas of the form $V\alpha$ reflects the absence of any assumptions concerning the notion of verification. A formal consequence of this absence is that formulas of the form $V\alpha$ behave like propositional variables. Of course, substitution of equivalents then fails. α is necessary in w iff it 'follows from' some proposition in $\mathcal{N}_L(w)$, the set of 'core necessities' of w.

Fact 4.7. *The following schemas belong to AM0:*

- $MM\alpha \supset M\alpha$
- $M(\alpha \wedge \beta) \supset M\alpha$

Moreover, if $\alpha \supset \beta$ belongs to AM0, then so does $L\alpha \supset L\beta$.

Proof. The first validity is a consequence of (iv). Note that $MM\alpha \supset M\alpha$ belongs to AM0 if $L\alpha \supset LL\alpha$ does. It is routine to check that (iv) ensures that the latter in fact belongs to AM0. The second validity follows from the truth-conditions for $L\alpha$ and $L(\alpha \wedge \beta)$. The final claim is a standard consequence of the truth-condition for $L\alpha$ (see Chellas 1980). □

We skip the examples of AM0-models and the arguments that most AM4-axioms are not valid in AM0. To facilitate comparison with AM4, however, we state (without proof) the following axiomatization result.

Proposition 4.8. *AM0 is soundly and completely axiomatized by the following axiom system. Every tautologous formula is an axiom and, moreover, every formula of the form*

(A2') $L(\alpha \wedge \beta) \supset (L\,\alpha \wedge L\,\beta)$

(A4) $L\alpha \supset LL\alpha$

is an axiom as well. The rules of inference are Modus Ponens and

(REL) If $\vdash \alpha \equiv \beta, then \vdash\ L\alpha \equiv\ L\beta$.

The main observation is that AM0 validates (3) and (4), but not so for (1) and (2).

Proposition 4.9. *(3) and (4) are valid in AM0, but (1) and (2) are not.*

Proof. (3) is quite easy. Note (again) that

$\sim V\alpha \supset (\alpha \equiv\ (\alpha \wedge \sim V\alpha))$

is a tautologous formula. The rest follows by Fact 4.7.

(4) $MMV\alpha \supset\ MV\alpha$ and $M(\alpha \wedge\ MV\alpha) \supset\ MMV\alpha$ are valid by Fact 4.7. It follows by propositional logic that

$M(p \wedge\ MVp) \supset MV\alpha$

is valid in AM0. The rest follows by propositional logic and the definition of M.

(1) and (2) are very easy. Formulas of the form $V\alpha$ have arbitrary truth-sets. Hence, we can easily construct a model over $\mathcal{W} = \{v, u\}$ such that $|\sim V(p \wedge \sim Vp)| = \{v\}$ and $|\sim V\sim(p \vee V\sim p)| = \{u\}$, but $\mathcal{N}(v) = \{\{v, u\}\}$, for example. But then both $L\sim V(p \wedge \sim Vp)$ and $L\sim V\sim(p \vee\ V\sim p)$ are false in v. □

Proposition 4.9 shows that Cmorej's results concerning (3) and (4) are obtainable rather easily. In fact, they follow from two very weak assumptions concerning necessity and are independent of any specific interpretation of the operator 'V'.

4.3. Additional remarks

The results of the above two sections suggest that AM4 is not the weakest possible logic of necessity and verification for which Cmorej's results are derivable. Let us consider AM2, the combination of AM0 and AM1. We could discuss its semantics in terms of $\mathcal{N}_L$ and $\mathcal{N}_V$, but we only mention the corresponding axiom system. As usual, every tautologous

formula is an axiom and Modus Ponens is a rule of inference. The additional axiom schemata are (B1'), (B2), (A2') and (A4). Additional inference rules are (RE) and L-necessitation. It is clear that AM2 is weaker that AM4, but all of (1) – (4) are valid in AM4. Hence, Cmorej's original system is not the weakest one for which his main results hold.

Let us note that the converses of (3) and (4) are derivable in AM0.3, a system that results from AM0 by adding (A1) and (B1). (Again, providing a semantics for this system is easy.) Let us see why.

Firstly, if both (A1) and (B1) are valid, then so is

$$V\alpha \supset (\alpha \wedge MV\alpha)$$

But then, by Fact 4.7 (which obviously holds for AM0.3 as well),

$$MV\,\alpha \supset M(\alpha \wedge MV\alpha)$$

is valid. The validity of the converse of (4) follows by propositional logic and the definition of M. Secondly, let us assume that $MV\alpha$ holds in some world w for some α. Then $M(\alpha \wedge V\alpha)$ holds in w by (B1). By propositional reasoning and (REL), $M(\alpha \wedge \sim(\alpha \wedge \sim V\alpha))$. Consequently,

$$M\left((\alpha \wedge \sim(\alpha \wedge \sim V\,\alpha)) \vee ((\alpha \wedge \sim V\,\alpha) \wedge \sim\alpha)\right)$$

in w. But the latter means that $\sim L\,(\alpha \equiv (\alpha \wedge \sim V\alpha))$ in w.

Hence, a system in which all of (1) – (4) plus the converses of (3) and (4) hold is the combination of AM0.3 with AM2, which we can call AM3. (In an axiomatization of AM3, (B1') can be omitted in favour of (B1).) Again, it is rather clear that AM3 is weaker that AM4. This could be shown rigorously by model-theoretic arguments, but we shall not engage in this exercise here.

5. Conclusion

The present article has elaborated on Cmorej's (1990) interesting results concerning unverifiable and unfalsifiable empirical propositions in three ways. Firstly, we have provided simple model-theoretic arguments establishing the main results with respect to the logic AM4. This was made possible by our soundness and completeness results for AM4 using

a version of neighbourhood semantics. Secondly, we have pointed out some striking similarities of Cmorej's findings to aspects of two well-known philosophical problems, Moore's Paradox and the Knowability Paradox. Thirdly, we have generalised Cmorej's results and discussed logics weaker that AM4 in which some combinations of the results hold. It has been argued that, in fact, AM4 is not the weakest logic in which all of Cmorej's original results hold. Perhaps AM4 is to be preferred to such weaker logics on some other grounds, but we leave this issue open.

References

BLACKBURN, P. – DE RIJKE, M. – VENEMA, Y. (2001): *Modal Logic.* Cambridge: Cambridge University Press.

CHELLAS, B. (1980): *Modal Logic. An Introduction.* Cambridge: Cambridge University Press.

CMOREJ, P. (1988): Neverifikovateľné a nefalzifikovateľné empirické propozície. In: Svoboda, V. – Zapletal, I. (eds.): *Logica '87.* Prague: Institute of Philosophy and Sociology of the Czechoslovak Academy of Sciences, 81-99.

CMOREJ, P. (1990): Unverifiable and Unfalsifiable Empirical Propositions. In: *Logica et Methodologica* II. Bratislava: Faculty of Arts of the Comenius University in Bratislava, 145-166.

FITCH, F. B. (1963): A Logical Analysis of Some Value Concepts. *Journal of Symbolic Logic* 28, 135-142.

GREEN, M. – WILLIAMS, J.N. (eds.) (2007): *Moore's Paradox. New Essays on Belief, Rationality, and the First Person.* Oxford: Oxford University Press.

HINTIKKA, J. (1962): *Knowledge and Belief. An Introduction to the Logic of the Two Notions.* Ithaca: Cornell University Press.

HUGHES, G.E. – CRESSWELL, M.J. (1996): *A New Introduction to Modal Logic.* London: Routledge.

MOORE, G.E. (1942): A Reply to My Critics. In: Schilpp, P. (ed.): *The Philosophy of G.E. Moore.* La Salle: Open Court.

PAULY, M. (2002): A Modal Logic for Coalitional Power in Games. *Journal of Logic and Computation* 12, 149-166.

SALERNO, J. (ed.) (2009): *New Essays on the Knowability Paradox.* Oxford: Oxford University Press.

SEGERBERG, K. (1971): *An Essay in Classical Modal Logic.* Uppsala: Filosofiska Föreningen Och Filosofiska Institutionen Vid Uppsala Universitet.

VAN BENTHEM, J. – FERNÁNDEZ-DUQUE, D. – PACUIT, E. (2014): Evidence and Plausibility in Neighborhood Structures. *Annals of Pure and Applied Logic* 165, 106-133.

VAN BENTHEM, J. – PACUIT, E. (2011): Dynamic Logics of Evidence-Based Beliefs. *Studia Logica* 99, 61-92.

Organon F 22 (3) 2015: 378-394

The Minimalistic Definition of Conventions: One Step beyond Millikan's Approach

VOJTĚCH ZACHNÍK

Department of Philosophy. Faculty of Arts. University of Ostrava
Reální 5. 701 03 Ostrava. Czech Republic
zachnik.vojtech@gmail.com

RECEIVED: 16-04-2015 • ACCEPTED: 08-07-2015

ABSTRACT: The study proposes a new approach towards a social phenomenon called convention and submits a minimalistic definition of convention, which provides a promising basis for future analysis unburdened by contra-Lewisian objections. The definition itself, based on the insights of Ruth Millikan in the study *Language Conventions Made Simple*, represents a simple and efficient means of delimiting essential components of conventional behaviour (stripped of most of the controversial issues from previous debates on Lewis's notion) solely by means of the role of precedent and its ability to reproduce. Yet, it is argued that a few additional conditions are required for a valid and distinct notion of conventionality: namely, the inclusion of a coordination aspect and an extension of the concept of precedent. The final version of the definition, thereafter, meets intuitive requirements of conventionality (e.g., arbitrariness) and has the generality to embrace different types of conventions.

KEYWORDS: Convention – coordination – minimalistic definition – precedent – reproduction.

The domain of conventions certainly attracts a broad scientific interest – in addition to economists and sociologists, philosophers have also become engaged in this area, the most famous of whom being David Lewis,[1] whose

[1] "A regularity R in the behavior of members of a population P when they are agents in a recurrent situation S is a *convention* if and only if it is true that, and it is common knowledge in P that, in almost any instance of S among members of P,

pioneering work is characterized by a variety of considered components of conventionality and their profound analysis. Since then, however, research in this area has progressed significantly and there are a number of well-known objections refuting the – so called – "Lewisian" project. In this paper, I make a proposal for a definition of conventions in general and prove how the definition can provide a comprehensive explanation of convention without splitting the social phenomenon under consideration into separate subspecies.[2] Also, I briefly verify the immunity of the definition to contra-Lewisian objections. Unfortunately, there is a vast amount of literature on these objections and a careful examination and evaluation of each of them is beyond the scope of this study. I therefore confine myself only to an explanation of their major conclusions (together with references to the appropriate literature).

The main problems include the question of epistemic requirements for an agent. It is now evident that setting high standards backfires immediately, since the defined notion suffers from being too restrictive. A successful delimitation should abandon the attempt to introduce an overly rational agent into the account.[3] Second, there are compelling reasons not to con-

---

(1) almost everyone conforms to R;
(2) almost everyone expects almost everyone else to conform to R;
(3) almost everyone has approximately the same preferences regarding all possible combinations of actions;
(4) almost everyone prefers that any one more conform to R, on condition that almost everyone conform to R;
(5) almost everyone would prefer that anyone more conform to R' on condition that almost everyone conform to R';

where R' is some possible regularity in the behavior of members of P in S, such that almost no one in almost any instance of S among members of P could conform both to R' and to R" (Lewis 1969, 78). His later modification of the original account has been published in Lewis (1975/1983).

2 I do not intend to hold a pluralistic approach and fully admit an unbridgeable diversity of conventions. Although, it might seem intuitively flawed to seek common basis of all conventionality, in my opinion, the usual scientific practise proposes clearly defined terms on the grounds of ordinary concepts that are difficult to grasp. Therefore, in this paper, I follow the ongoing debate on the categorization of social reality and consequently propose my own contribution to it.

3 In this respect, the influential development of the evolutionary game theory shed light on new ways to solve many problems associated with conventions and newly re-

sider a set of expectations about others' actions (together with a conditional preference for conformity) as the only and exclusive source of motivation for following a particular conventional pattern. Sustainability of such behaviour is equally well explained without referring to the actual or potential mental state of individuals.[4] An adequate definition should, therefore, grant further impetus to a conforming action, or it should be more open to other unspecified inclinations leading to a proper result. Finally, it is necessary to take into consideration the fact that sequential decision-making[5] in conventional situations allows for the actual coexistence of several possible solutions (not only their logical possibility arising from a coordination problem with multiple equilibrium states). Accordingly, it seems redundant to require a uniform regular conformity with exactly one pattern of behaviour in a specific community[6] and it is necessary to define this social phenomenon in a way that permits the existence of conventions which currently cooccur with other alternatives without being in conflict.

Many efforts have been made to circumvent these objections. I do not intend to begin from point zero but rather to build on the basics already established. For my purposes – to conceptualize general conventionality – there is an eligible definition of the fundamental components of the researched phenomenon presented by Ruth Millikan in the paper *Language*

---

vealed that an origin and stability of social conventions can be explained without any need to accept a model of an ideally rational agent and solely on the basis of social dynamics and evolutionary rules. See Skyrms (1996), Young (1996) or Sugden (1998).

4 For example, Burge (1975) highlighted the importance of entirely irrational factors maintaining conventions, such as tradition and ignorance. See also Gilbert (2008) or Young (1996, 58).

5 More precisely, this holds for any dynamic game with perfect information.

6 This idea was submitted by Millikan (2005, chap. 1) as she argued that we normally regard some patterns of behaviour as conventional ones even if they occur in the information-transparent environment where individuals can immediately observe each other and thus modify their behaviour appropriately (by changing the pattern). Although situations like these are trivial coordination problems (or as she calls them "open coordinations"), it does not exclude them from classification under the notion of convention. Another relevant argument against the preference for uniform conformity was introduced by Miller (2001). He questioned an importance of non-action-determining preferences, among which he also ranked the above-mentioned preference.

Convention Made Simple.[7] Her project is particularly interesting because it successfully avoids (unintentionally) all the major critiques raised previously and, moreover, provides the austerity and simplicity of conditions beneficial with regard to the objective of finding the common core of conventions. These advantages imply broad fields of application for the delimited concept and further prospective refinements in specific contexts; consequently, the definition does not suffer from excessive restrictiveness. But on the other hand, it must be considered whether this virtue does not lead to the opposite extremes of radical openness and looseness, which would mean that the definition admits other cases of social behaviour quite apart from conventions. Such a result cannot be considered useful due to the lack of a distinctive force. Let us now look more closely at what she states, exactly:

> Natural conventionality is composed of two, quite simple, related characteristics. First, natural conventions consist of patterns that are 'reproduced' in a sense to be defined. Second, the fact that these patterns proliferate is due partly to weight of precedent, rather than due, for example, to their intrinsically superior capacity to perform certain functions. (Millikan 2005, 2)

Of course, this paragraph can hardly be regarded as a proper definition, yet it shapes intelligibly the principal features of conventionality and unquestionably differs from Lewis's analysis. As a result of this, it provides an entirely altered standpoint for further research in this field. Above all, it is evident that conventions (understood as patterns of behaviour) are specified by just two related conditions: reproducibility and proliferation based on precedent. In Millikan's view, nothing more is required.

Her instructions, therefore, could be reformulated (with a number of simplifications) in the following form:

> *Millikan's definition of conventions*:
>
> $\text{Convention}_m =_{df}$ a pattern of behaviour which is reproduced and expands by weight of precedent.

At first sight, the definition is merely an abridgement of the quoted passage; however, a closer look reveals that convention_m is different in sev-

---

7 The paper was reprinted in Millikan (2005).

eral aspects.[8] Namely, Millikan offers nothing more than an examination of *natural conventions*, which is only a subclass of conventions (an alternative to this are stipulated conventions).[9] Furthermore, she mainly highlights a causal effect of precedent in opposition to the intrinsically functional capacity of a particular pattern of behaviour. Regarding natural conventionality, I am convinced there are many problems with the "fission" strategy (a tendency to examine a subset of the general category without defining a parent concept). From a methodological point of view, it starts in the middle of a process without providing an adequate explanation of when it is widely permissible to talk about the convention. And that makes the strategy a double-edged sword: it recommends that we begin with a particular subclass of the phenomenon without identifying its universal class.[10] Of course, many other issues arise almost immediately in connection with whether and how a transition between these two types is made, how they differ from each other and from other forms of social interaction, etc. I think this problem could be avoided by simply supposing that everything she says can be considered in relation to $\text{convention}_{\text{m}}$ and I find a number of advantages in this: $\text{convention}_{\text{m}}$ establishes conditions that could satisfy recommendations learnt from post-Lewisian debate, and I believe there might even be a stronger and more general version ($\text{convention}_{\text{M}}$) which would provide a plan for setting out the basic components of conventionality. The aim of the study is also to explain how to get from the first draft ($\text{convention}_{\text{m}}$) to the complete definition ($\text{convention}_{\text{M}}$).

---

8 I should point out that this version shifts the meaning of the defined term in a way which Millikan would hardly have agreed with, yet the influence of the original text is still very noticeable and I feel obliged to call it "Millikan's definition".

9 This distinction is far from being unique, similar claims are very common. For example, Young (1996, 106) discerns conventions which are established by central authority and those established by gradual accretion of precedent.

10 On the one hand, Millikan gave up the attempt to submit a general definition by restricting her analysis on a partial group of conventions, specifically conventions of natural language. She was convinced that this great task (typical of the Lewisian project) remains unattainable, since it is vulnerable to a large number of counterexamples. On the other hand, as Bunzl – Kreuter (2003) pointed out, most of her examples have nothing to do with language, although she limited her account primarily to linguistic conventions.

Primarily, it must be confirmed that the resistance of the so-called Millikan definition to earlier objections is valid. Even a cursory inspection of convention$_m$ reveals that desirable epistemic standards are achieved: there is no demand to involve common knowledge in an explanation of the formation and maintenance of conventional behaviour.[11] Moreover, convention$_m$ is not based upon a restriction regarding what mental states individuals should have; it might sound radical, but whether or not an agent has beliefs concerning other agents' actions is not fundamental. Many philosophers have argued against epistemic conditions by claiming that it is not necessary to take these attributes into account and accordingly they have come up with many cases of conventional behaviour without any belief and common knowledge (e.g., dance moves, handshakes, and dress codes).[12] Naturally, others may counter with more complex and sophisticated examples which essentially rely on beliefs, and whose presence is quite vivid and conspicuous. I do not deny this, but I think their role should be reflected in a general theory of conventions. For example, a convention based on an explicit agreement indisputably gives rise to many beliefs between individuals; however, my point here is that a belief-state is neither a necessary nor sufficient condition of convention. Accordingly, I briefly comment on two epistemological objections and further two await an assessment. It is probably not surprising that convention$_m$ – strongly based on Millikan's view – is successful in relation to the doubtful assumption of regular conformity. After all, it was her aim to prove that conventions are not bound by this condition. And the definition does not contain anything that could, even indirectly, imply that people must conform to precisely one pattern of behaviour for a given type of situation. The simultaneous occurrence of a large number of patterns is in accordance with the definition, since it is not problematic to see that co-existence is permitted unless it is true that those

11 To what extent common knowledge is epistemically questionable depends essentially on particular details of the accepted conception. The recent interpretation made by Cubitt – Sugden (2003) suggests that Lewis had a relatively uncontroversial idea about how common knowledge works. Nevertheless, the question of whether this condition is truly necessary for a fully-fledged account of conventions is still a matter of debate, see Binmore (2008).

12 Arguments and examples in favor of the absence of belief-component are provided, for example, by Gilbert (2008) or Millikan (2005, 1-23).

patterns of behaviour are reproducible and proliferate due to precedent.[13] And finally, it seems to me that the current form of the definition leaves completely open the question of what the motives are for a decision to follow a conventional action. There could be heterogeneous sources for the formation of reasons, but it does not matter whether the major influence begins at the conscious or unconscious level, through explicit expectations or through ignorance. I assume, thus, convention$_m$ is not compromised by the objections that have appeared in post-Lewisian debate and it can serve as a new starting point in the search for the foundations of conventionality.

Nevertheless, this achievement has only minor effects, as it provides only initial confirmation that Millikan's definition is not threatened by earlier objections directed at a completely different conception. For the second stage, I have to assess the plausibility of the very definition, how precisely it determines the boundary line and how much responsive filter it provides to counterexamples. Without this, one could easily argue that convention$_m$ is an intentionally flexible notion allowing only a rejection of earlier difficulties and therefore lacking sufficient predictive power on its own. Such a problem would cause considerable difficulty, meaning the definition is too vague and general. This criticism has been raised already by philosophers Bunzl and Kreuter in Bunzl – Kreuter (2003). They oppose Millikan's proposal, referring to the fact that the definition is both too tolerant yet at the same time too restrictive. One of the major problems, in their opinion, lies in the second condition in which a causal capacity of precedent stands in opposition to intrinsic features of a pattern. As precedent has the most important role in the expansion of a pattern and since only its gradual accretion leads to a more stable convention, the definition obliges us to dismiss all social activities originating from the intrinsically superior capacity of patterns to perform certain functions (cf. Bunzl – Kreuter 2003, 420). In order to understand the aforementioned problem let me clarify what Millikan intended by this condition. Emphasising the role of precedent primarily preserves the intuition that conventions are especially *arbitrary* pat-

[13] In a fact, there might be cases in which the social dynamics gives rise to only one regularity in behaviour for the given situation of simultaneous decision-making whereas in a sequential structure of the same type of interaction, the existence of several regularities is very likely.

terns, i.e., patterns with an equivalent alternative, and their "pushing through" is determined by the predominance of precedent. Thus, the pattern that survives in competition with others is the one that spreads more widely among members of the community (as part of their repertoire of behavioural patterns). Many were worried that to give priority to a functional aspect of a pattern would result in the denial of arbitrariness, because then we would have no rationale to believe that there must be more than one pattern and that the convention may be otherwise. Another reason we should reject the objection raised by Bunzl and Kreuter lies in the fact that a convention understood as a pattern proliferating due to intrinsic features inevitably involves skills.[14] There is no doubt Millikan uses the second condition (proliferation due to precedent) as a protection against the possible objection that skills acquired by reproduction can be identified with conventions. She says:

> I learned from my mother, and she from hers, to open a stuck jar lid by first immersing it in hot water. Opening jars this way is not thereby 'conventional'. To be thought of as conventional, a reproduced pattern must be perceived as proliferated due, in important part, to weight of precedent, not to its intrinsically superior capacity to produce a desired result... (Millikan 2005, 7)

I believe the problem arises from the following false dilemma: either a convention is arbitrary and proliferates mainly due to precedent, or there is only a conventional pattern that somehow stands out and proliferates through its intrinsic properties, despite the fact that it cancels out arbitrariness. Yet obviously some patterns may be better with regard to a particular purpose and this is entirely consistent with their being conventional insofar as the plurality of patterns is maintained. These patterns perform a function equally well, since they have intrinsic properties of an equivalent quality. The idea ensures both the intuition of arbitrariness (i.e., the conjecture that conventions are a means for solving an equilibrium-selection problem) and the possibility of proliferation (among others) by intrinsically superior properties of conventional patterns.

---

[14] Millikan, as I will show below, does not refer to coordination problems as a basic structure of conventions, and therefore, her position is threatened by the inclusion of skills.

This implies that I managed to include the criticism of Bunzl and Kreuter into the account and secure the arbitrariness requirement, but unfortunately the defensive barrier separating the territories of conventions and skills collapsed. Clearly, there are many equivalent skills reproduced due to function and precedent, though they are not conventions (remember the jar lid opening). Is there any solution or treatment allowing convention$_m$ to avoid this unintended result? Yes, but the nature of this additional change significantly shifts the meaning of the definition far from Millikan's intention and closer to Lewis's original proposal. There is an assumption that shaped a rough idea of conventions at the beginning of the analysis (see Schelling 1960, part II) but was sidelined in convention$_m$ and also in Millikan's original view, despite the fact that it not only has a huge explanatory potential (to relate conventions to a particular type of social interactions) but also solves several issues (e.g., the questionable inclusion of skills). I am referring to the *coordination aspect of conventions*. No reference has been made so far to what type of social interactions conventions correspond to because Millikan thinks conventions (and correspondingly convention$_m$ captured the same idea) are not necessarily the results of coordination problems, since some instances of conventions disprove of this. Millikan (2005, 2) gives examples of swearing and expletives which, in her view, express emotion for private purposes (such as relieving pain, etc.) without involving any coordination at all. However, these instances of language phenomena can be explained as a parasitic form of coordinative language convention, which means that most language patterns are coordinative, and a few – non-coordinative – exceptions were derived from the majority.[15] I still think that this sample of cases proving the non-coordinative character of conventions is not sufficiently clear-cut to lead to any definite conclusion about the nature of conventionality. Moreover, the benefits of a coordination-feature clearly outweigh possible concerns and thus justify the inclusion of this concept into a theoretical framework. And the specification of interactions clarifies the area of social reality that is to be conceptualized under the notion. This results in a better understanding of rele-

15 This explanation is not sufficiently strong for a conclusive argument; however, it is supposed to show the existence of a few dubious examples that hide a coordinative function, and therefore pose no risk. Yet, I will not discuss this issue in greater detail, because I do not want to pay close attention to the specific case of linguistic conventions.

vant and irrelevant properties of given social phenomena; skills can now hardly be considered to be conventions (despite the fact that they satisfy conditions of convention$_m$), because they lack a coordination structure and do not occur in social interactions. Unlike conventions, skills are mainly useful for one-person optimalization problems not for social purposes, and this distinction is reflected in the inclusion of the coordination condition. The upshot is this: the definition needs to be tightened by the addition of the coordination aspect.[16] Yet, I do not claim that this move has completely removed all doubts. Nevertheless, at the very least, it is apparent that some objections have been addressed and the current position seems to be more robust than before.

Now, I will turn to the last objection to convention$_m$. If the definition explicitly identifies convention with a precedent-determined reproducible pattern of behaviour, then it also includes a number of observable behaviours in animal species. Is it indeed desirable to allow birds, or apes to "participate" in conventionality? Do we not have the feeling or intuition that humans alone can follow conventions? The minimal epistemic requirements implied by the definition do permit the inclusion of small children, who are unable to consciously reflect and justify their actions – which was not possible in Lewis's theoretical framework[17] – but bird song is also conventional.[18] In my opinion, there is nothing wrong with attributing conventional behaviour to other animals as long as the above-mentioned conditions are met. Besides, I agree with Binmore, who additionally uses this case as an illustration of the claim that convention is not conditioned by any sort of knowledge or doxastic state.[19] Many others, however, have less

16 From the perspective of a broader debate, there are a few technical issues concerning coordination. Some authors argue for a more subtle concept. Vanderschraaf (1995) proposes a correlated equilibrium instead of coordination. However, there are also those who are closer to the approach I have mentioned, like Bicchieri (2006, 29-42).

17 He says this explicitly with regard to language convention (see Lewis 1969, 51), which may sound less controversial, but this is a general corollary of his approach (see Lewis 1969, 75) and it causes major problems in other areas of social interaction.

18 Compare Millikan (2005, chap. 1) with Bunzl – Kreuter (2003) and Binmore (2008).

19 "Young birds learn to sing complicated arrangements of notes by listening to the songs of experienced birds. It matters a lot to them what song they sing, because the songs are used as a coordinating device in deciding who mates with whom. But the birds do not 'know' any of this" (Binmore 2008, 25).

sympathy for such an extension of conventionality. Even Millikan (2005, 7) rejects it, in the following terms:

> The songs of the various bird species are to a large degree arbitrary in relation to function, but they are not conventional, because they are not copied or reproduced in the sense defined above.

Nevertheless, I can see no evidence to suggest these songs are not conventions. Perhaps she believes they are reproduced solely by genes and not by "culture" (by a precedential imitation or learning; see Bunzl – Kreuter 2003, 421), in which case she should, nevertheless, accept bird song and many other animal behavioural patterns, if it turns out that their reproduction process largely involves imitation (or a similar form of precedential learning) rather than gene transfer. I would suppose there truly is an extensive region of such patterns (and its boundaries can be fully established by biologists). Therefore, I am willing to take a more inclusive stance in this matter and to accept cases of non-human conventional behaviour. At least it is evident that coordination problems are quite common in the animal kingdom and if their solutions are based on behavioural patterns reproduced by precedent then there is nothing that would justify their exclusion from conventionality, even though it might go against our intuition.

It has been argued so far, in the light of objections, that convention$_m$ can and should subsequently be updated by specifying the type of corresponding interactions, and it has also been pointed out that no harm follows from the extension of conventionality to other biological species. Those objections regarding excessive narrowness (a neglect of functionally proficient patterns) or, vice versa, looseness (skills) can from this moment be disregarded. Yet I am far from claiming that I have considered a wide range of crucial arguments; my aim is rather limited to proving that even a simple and efficient formulation of conditions may retain the potential for a general analysis of conventions, and ensure the key elements of this social phenomenon. In order to verify this, a final step remains to be taken: to explain how convention$_m$ (derived from Millikan's *natural conventionality*) may cover cases of *stipulated* convention, a completely different kind of conventionality. If the definition aims at the more ambitious project of seeking the notion of convention in general, it is necessary to prove that stipulated conventions meet the conditions. Compared with Lewis's study, which comprehensively explains how convention depends on a variety of states of

affairs (agreement, precedent) and how the states further determine our expectations and actions, Millikan's definition cannot compete in descriptiveness. Are we thus to declare the definition invalid as an explanation of the heterogeneous origins of conventional behaviour? I do not maintain this at all. Notwithstanding the fact that convention$_m$ is concerned exclusively with precedent and behaviour determined by it, there seems to be a possibility of a simple adjustment within the definition, allowing the integration of *an agreement* as a model of stipulated conventions. This procedure will secure the expansion of conventionality in the desired way. It is surprising, however, that the way to perform it has already been indirectly and unintentionally indicated by Millikan. When considering ways of reproduction she mentions a case in which a given convention is reproduced by verbal instruction. In her view, there might be a conventional regularity when "one person may tell another how a pattern goes. For example, Johnny's mother tells him that he is to put his letter in the mailbox and put up the flag..." (Millikan 2005, 4). Notice how the example satisfies conditions I have mentioned above (i.e., conditions of convention$_m$ and the condition of coordination); it is a pattern of behaviour reproduced by precedent (by means of verbal instructions) and clearly, it is a situation in which the sender and the mailman have to coordinate their patterns with regard to the implementation of postal services. I see no obstacle to acknowledging the fact that it is a convention. Moreover, if we generally declare – as Millikan did – that verbal instructions are one of many means of reproduction, then it also does not seem problematic to admit that these instructions can shape the precedent on their own. A required generalization of the definition, therefore, stands and falls with the acceptability of the assertion that precedent can be shaped by a set of verbal instructions, because in this manner it is possible to explain how the agreement is "transferable" to precedent.

How are we to understand the concept of precedent? Thus far, I have implicitly held Millikan's understanding of precedent as a model case from past experience whose performance is – in some essential aspects – binding for future applications.[20] Yet, if it is conceivable for a precedent to be re-

[20] However, Millikan (2005) does not explain it in detail. And perhaps the reason why so little has been said about the notion of precedent is that Lewis (1969, 36-37) consid-

produced by means of verbal instruction, it is, I believe, also reasonable to regard the instructions themselves as a precedent in an abstract form. Obviously, there are some troubling questions concerning this extension: What would the realization of the same conventional pattern look like in view of a different formation and how would it be covered by the definition? Take for instance the '*Casual Friday' convention*: it is permissible for a company employee or an office member to dress informally on Friday. Quite apart from the very specific details of the convention, let us focus exclusively on two different origins of it. First, it is logically and actually possible that this behaviour emerged due to accidental circumstances, simply by the fact that one day somebody dressed casually (e.g., Peter had taken his suit to the cleaners and had nothing to wear except jeans) and this behaviour was then imitated by some of his colleagues the following week (whether unconsciously or as a result of simple impulse: "Why don't I wear comfortable clothes on Friday as Peter did last week?"). Suppose then, the 'Casual-Friday' precedent is even more successful to the point that it has prevailed and replaced the original conventional pattern of formal dress; and this happens completely as a consequence of the gradual reproduction of a precedent. Therefore, an amended definition is perfectly suitable for an inclusion of this type of conventionality (*natural conventionality*). Second, an alternative scenario may be taken into account: one day, staff of a firm agrees on this convention and by the agreement a particular pattern is established. In this case, regardless of what others expected or preferred, a set of verbal instructions was submitted, determining precedent in an abstract form,[21] which consequently influenced the behaviour of a community. Hereafter, precedent is shared verbally (as noted by Millikan) or blindly imitated (as in the case of natural conventionality). Thus, neither is this type of conventionality excluded by the definition (of convention$_M$). Both of these examples prove the idea that the looser and wider notion of precedent is good for the generality of the definition.

---

ered it to be primitive once he identifies it with *salience*. Compare Sugden (1998) with Postema (2008).

21 By "abstract" I highlight the fact that no such (concrete) behaviour did occur in the intended environment and those instructions formed the precedent, so to speak, out of nothing. Also, I believe this distinction does not make the notion of precedent artificial, because the abstract precedent expresses only the fact of our minds and their capability to mentally model certain social interactions and their consequences.

Of course, this does not mean that none of the members of that community is able to form expectations of the others' behaviour, or reflect their own preferences and actions. All I say is that minimal sufficient and necessary conditions of convention in general are reproducibility and the role of precedent. Regarding stipulated convention, they fit perfectly into a modified version of precedent (and into the definition as well). Precedent in an abstract form has the same role as the instructions of a mother about how to send a letter, with the sole difference that in the first case, precedent is newly introduced and has been absent until now, whereas in the second, the mother passed on an already existing pattern. I deem this step in my work significant because it is a bridging element to the general definition, and for the sake of clarity, I present a brief structure of the argument as follows:

The Argument in favour of the Generalization of Convenition$_m$:

1. A precedent can be reproduced by means of verbal instructions.
2. If a set of verbal instructions can reproduce a precedent, it may also shape the precedent.
3. Therefore, a set of verbal instructions may shape a precedent.
4. An agreement is a set of verbal instructions.
5. Therefore, an agreement can shape a precedent.

The Premise 1 is closely based on the example mentioned by Millikan, in which she presents one of the options of reproduction. Another premise 2 admits an intuition that verbal instructions can not only reproduce – already established – regularity in behaviour, but that they are also able to stand themselves at its beginning. If that is right, I infer the conclusion 3 that indicates the possibility of a broader concept of precedent (which is specific or abstract). The statements 3 and 4, then, form premises of an ongoing argument whose conclusion 5 confirms the initial assumption about the role of agreement in relation to precedent.

As for criticism, the first series of objections can be anticipated regarding premise 2. Some may doubt the fact that these instructions actually give rise to precedent. It is evident, for instance, that an agreement made at the end of a meeting by saying: "Tomorrow here, again, then" is a useful reminder or an easy way of reproducing a functioning precedent (to meet at a certain place at a given time); however, its ability to create new precedent is weak and questionable. The answer to this objection is, in principle, triv-

ial, for premise 2 does not necessarily represent a claim that an identical set of instructions is sufficient both for reproduction and for the establishment of a precedent. It contains weaker condition, a requirement of a fundamental possibility to establish precedent. When I exemplarily put a postage stamp on the envelope and say "That's how it's done", and conversely when I explicitly give somebody instructions relating to the same conventional pattern, there is, undeniably, a difference in complexity and sophistication. Yet, it seems to me that as the first kind probably permits the second, a reproduction through instructions allows an establishment in the same way.[22] Another objection might be raised against premise 4 – what type of agreement do we have in mind? The kind of an implicit agreement which, at first glance, does not show any similarity to an accurate linguistic expression of the instructions? Admittedly, not every agreement in a broad sense[23] may fit into the conditions specified in premise 4. Nevertheless, for my purposes it is quite satisfying that the generality of the definition holds with respect to both kinds of natural and stipulated conventionality and that this expansion might be further refined (by the theory of conventions).[24] Thus, premise 4 ought to be taken modestly as a condition that if an agreement is made up of a set of verbal instructions, then it might be a source of precedent in abstract form. Although this conclusion seems lacking in descriptiveness and explanation, it satisfactorily meets the requirements of generality. Moreover, I believe the argument provides reasons to think that the new and modified version of Millikan's definition is widely applicable and can be expanded to cases of stipulated convention. This would secure beneficial prospects, to be precise, a broader notion of

---

22 As a matter of fact, the statement "That's how it's done" can be called into question, particularly whether it is truly a set of instructions. If it is, at most, a declaration that precedent was held, the premise 1 will consequently eliminate the possibility of this. All the same, it is not difficult to come up with a more sophisticated case of a similarly austere statement satisfying the premise, e.g., "Here it sticks." Therefore, the argument remains valid.

23 The agreement understood rather in terms of normative attitudes than based on its particular form.

24 The theory might provide a closer specification of the cases of conventional behaviour with normative attitudes (as an agreement in the broad sense) based on implicit precedents.

precedent adds explanatory potential, as it allows a larger framework for the categorization of social reality, especially conventions in all their variety.

After a series of adjustments to reinforce the definition derived from Millikan's original proposal, I summarize the definition as follows:

The Minimalistic Definition of Convention:

Convention$_M$ = $_{df}$ a pattern of behaviour *occurring as a result of coordination situations* which is reproduced and expands by weight of *precedent**.

The parts highlighted in italics indicate the changes that have resulted from the previous considerations and arguments, namely the inclusion of the coordination aspect (omitted in the convention$_m$) and extension of precedent as set out above. Regarding an overall assessment, there are several criteria whose fulfilment would be desirable. The most intuitive is, without doubt, *arbitrariness*, which is remarkably noticeable from the common oversimplification by which convention is routinely identified with arbitrary patterns.[25] The convention$_M$ meets this requirement in two ways: the very nature of coordinations demonstrates the existence of multiple-equilibria, and consequently different alternatives, and, furthermore, the essential role of precedent – with regard to a pattern-expansion – supports the fact that some patterns prevail due to accidental circumstances and would be replaced by a different (qualitatively equal) pattern if the social dynamics developed differently. Second, the definition assigns to convention an exact type of social interaction, specifically emphasising a coordination structure; and for this reason we have a clear idea of *the nature of social interactions* which conventions correspond to. In addition, it brings in its wake many other advantages, such as the elimination of undesirable cases (skills). The third point is *immunity against contra-Lewisian objections*. As mentioned above, it seems to me that all these objections have already been discarded in the case of convention$_m$, and no adjustment in convention$_M$ could restore them. Finally, I submitted an argument that secures *generality* (and provides an explanation of different types of conventional behaviour by means of the same conditions).

I conclude that the minimalistic definition of convention can withstand many critical reactions; offers a robust basis for any general theory of con-

25 Although it is undeniably a necessary condition, it is far from being sufficient. What is arbitrary, what could be otherwise, does not define a convention accurately.

ventions; and complies with other requirements. The achievement of these standards has visibly confirmed the presumption that the definition (of convention$_M$) provides an adequate account of the basic components of conventional behaviour.

References

BICCHIERI, C. (2006): *The Grammar of Society: The Nature and Dynamics of Social Norms*. New York: Cambridge University Press.

BINMORE, K. (2008): Do Conventions Need to Be Common Knowledge? *Topoi* 27, 17-27.

BUNZL, M. – KREUTER, R. (2003): Conventions Made too Simple? *Philosophy of the Social Sciences* 33, 417-426.

BURGE, T. (1975): On Knowledge and Convention. *Philosophical Review* 84, 249-255.

CUBITT, R. – SUGDEN, R. (2003): Common Knowledge, Salience and Convention: A Reconstruction of David Lewis' Game Theory. *Economics and Philosophy* 19, No. 2, 175-210.

GILBERT, M. (2008): Social Convention Revisited. *Topoi* 27, 5-16.

LEWIS, D. (1969): *Convention*. Cambridge: Harvard University Press.

LEWIS, D. (1975): Languages and Language. In: Lewis, D. (1983): *Philosophical Papers. Vol. 1*. Oxford: Oxford University Press.

MILLER, S. (2001): *Social Action: A Teleological Account*. Cambridge: Cambridge University Press.

MILLIKAN, R. (2005): *Language: A Biological Model*. Oxford: Clarendon Press.

POSTEMA, J. G. (2008): Salience Reasoning. *Topoi* 27, 41-55.

SCHELLING, T. (1960): *The Strategy of Conflict*. Cambridge: Harvard University Press.

SKYRMS, B. (1996): *Evolution of the Social Contract*. Cambridge: Cambridge University Press.

SUGDEN, R. (1998): The Role of Inductive Reasoning in the Evolution of Conventions. *Law and Philosophy* 17, 377-410.

VANDERSCHRAAF, P. (1995): Convention as Correlated Equilibrium. *Erkenntnis* 42, 65-87.

YOUNG, H. P. (1996): The Economics of Convention. *Journal of Economic Perspectives* 10, 105-122.

Organon F 22 (3) 2015: 395-424

Logika verzus jazykoveda o vetných spojkách[1]

FRANTIŠEK GAHÉR

Katedra logiky a metodológie vied. Filozofická fakulta Univerzity Komenského v Bratislave
Šafárikovo nám. 6. 814 99 Bratislava. Slovensko
fantisek.gaher@uniba.sk

ZASLANÝ: 22-12-2014 • AKCEPTOVANÝ: 04-06-2015

ABSTRACT: The author proposes bridging disparate explications of sentential connectives (operators). Logicians take sentential operators to denote truth functions and distinguish 16 such binary operators. On the other hand, linguists recognize much more sentential operators and, even for the same truth function, distinguish several connectives with different properties. For example, the conjunction "and" can in some uses have an adversative or conditional or restrictive feature, and sometimes has a temporal or causal meaning. The author favors logico-semantic tools over pragmatic ones and proposes distinguishing operators as truth-functions from complex constructions of truth-functions. The same truth function can be constructed in different ways, which enables one to explain the different properties of a single connective discussed by linguists. The author proposes two conditional connectives – namely one for the necessary conditional relations and another for the sufficient conditional relation – the negation and the logical conjunction as the basic operators. Using these simple connectives, he defines constructions of other sentential operators and demonstrates their various properties.

KEYWORDS: Linguistic – logic – necessary condition – semantics – sentential connectives – sufficient condition.

[1] Táto práca bola podporená grantom VEGA 1/0221/14 Evidencia v sémantike. Ďakujem kolegom P. Cmorejovi, M. Zouharovi, L. Bielikovi, M. Braxatorisovi a dvom anonymným recenzentom za početné a cenné pripomienky a návrhy k pracovnej verzii state.

1. Cieľ – krok logikov k lingvistom

V stati Gahér (2001) som upozornil na dvojkoľajnosť prístupu logikov a lingvistov k explikácii vetných spojok (všeobecnejšie *operátorov*[2]) i na to, že jazykovedci by mali záporovú časticu, použitú v niektorých významoch, povýšiť na vetnú spojku, resp. operátor, aby ich systém vetných spojok bol úplný. V tejto stati navrhujem, ako by mohli logici zmenšiť pomyselnú vzdialenosť týchto prístupov, vykročiť smerom k lingvistom a prísť zo sémantickým vysvetlením niektorých relevantných zistení jazykovedcov.

2. Logici o vetných spojkách

Logici považujú výrokové (vetné) spojky (v úlohe parataxy alebo nepravej hypotaxy, t. j. v paratatickom použití) zo sémantického hľadiska (z hľadiska významu) za pravdivostné funkcie, ktoré podľa princípu kompozicionality spoluurčujú výslednú pravdivostnú hodnotu zloženého výroku v závislosti od pravdivostných hodnôt podvýrokov. Ich sémantika je obmedzená len na túto denotačnú (referenčnú) rovinu. Logici nerozlišujú rozličné konštrukcie týchto funkcií. Logické dôsledky zo zložených výrokov nie sú na úrovni výrokovo-logického vyplývania ovplyvnené sémantickým obsahom jednoduchých viet.

Pri binárnych spojkách ide o 16 základných typov funkcií. Funkcie môžeme identifikovať usporiadanou *n-ticou* funkčných hodnôt pre jednotlivé kombinácie pravdivostných hodnôt podvýrokov ako argumentov funkcie – napr. spojku *a* (konjunktor) reprezentuje štvorica (1,0,0,0),[3] a celá konjun-

---

[2] Presnejšie by sme mali hovoriť o *operátoroch* (všeobecnejšie pomenovanie, ktoré sa používa v logike už od polovice 19. storočia, ale už aj v lingvistike, a to v podobnom význame), pretože si budeme všímať nielen to, čo jazykovedci nazývajú *vetnými spojkami*, ale aj *korelované dvojice* (*korelatíva*) a *spojkové spojenia* (spojkové výrazy). Bližšie o tom napr. Kesselová (2007, 354), Kesselová a spol. (2013, 9). Podľa Kesselovej operátory všeobecne sú „jazykové prostriedky so spájacou funkciou, vytvárajúce z jazykových jednotiek nižšieho rádu (vetné členy, vety) koherentné výpovedné celky“ Kesselová (2007, 354). Kvôli zachovaniu kontinuity tradície označovania budeme výrazmi *vetné spojky*, resp. *spojky* a ich derivátmi rozumieť často aj *korelované dvojice* a *spojkové spojenia* .

[3] Výrazy typu „(1,0,0,0)“, kde „1“ reprezentuje pravdivostnú hodnotu pravda a „0“ pravdivostnú hodnotu nepravda, budú skrátené spôsoby reprezentácie funkcie, kde jed-

kcia je pravdivá vtedy a len vtedy, keď oba podvýroky sú pravdivé. Niektoré spojky – zrejme tie, ktoré nie sú prakticky potrebné - nemajú svoje štandardizované základné zachytenie v prirodzenom jazyku (ide o funkcie charakterizované štvoricami funkčných hodnôt (1,1,1,1), (1,1,0,0), (1,0,1,0), (0,1,0,1), (0,0,1,1), (0,0,0,0)).[4]

Na druhej strane ostatných šesť pravdivostných funkcií má zvyčajne viacero odlišných vyjadrení, pričom niektoré z nich nie sú len odlišným pomenovaním s totožným priebehom funkčných hodnôt, ale ich netotožná skladba môže signalizovať určitý – sémanticky nezanedbateľný – rozdiel. Paradigmatickým príkladom môže byť spojka *ak-tak* a spojka *len vtedy, keď*, ktoré vyjadrujú tú istú pravdivostnú funkciu (1,0,1,1), ale často plnia odlišné úlohy. Prvá sa v empirickej oblasti používa na vyjadrenie dostatočnej podmienky (zachytenej za slovkom *ak*) uskutočnenia nejakého deja (zachyteného za slovkom *tak*), zatiaľ čo druhá na vyjadrenie nutnej podmienky (zachytenej za spojením *len vtedy, keď*) uskutočnenia nejakého iného deja (zachyteného prvou zložkou súvetia). Tento významový rozdiel je sémanticky nezanedbateľný.

Logici[5] by mali zdôvodniť odpoveď na otázku: Prečo sa sémantická úloha vetných (výrokových) spojok redukuje na ich extenziu (plochú funkciu) v každom kontexte,[6] zatiaľ čo pre iné typy (plnovýznamových) výrazov jazyka existuje celá škála kontextov (propozičné a pojmové postoje), kde je sémanticky prioritná ich štruktúrovanosť? Aký na to existuje napríklad dôvod v Transparentnej intenzionálnej logike (TIL) (Tichý 1988), kde na zachytenie štruktúrovaného významu (hyperintenzie) sú vytvorené komfortné nástroje?[7] Zrejme je to otázka tradície, ktorá v modernej logike vychádza

---

notlivé členy predstavujú len funkčné hodnoty, pričom argumenty sú zamlčané, a platí, že prvý člen v postupnosti je funkčnou hodnotou pre dvojicu argumentov (1,1), druhý pre dvojicu argumentov (1,0), tretí pre dvojicu argumentov (0,1) a štvrtý pre dvojicu argumentov (0,0).

[4] Bližšie o tom pozri napríklad Cmorej (2001, 55 nasl.), Gahér (2013, 59 nasl.).

[5] Samozrejme, rátam medzi nich i seba.

[6] Toto zjednodušenie si logici vo všeobecnosti uvedomujú a upozorňujú na fakt, že spojky vyjadrujú ešte iné významy, ktoré však – zdá sa – nie sú dôležité pre logické vyplývanie; pozri napríklad Cmorej (2001, 57 nasl.), Svoboda – Peregrin (2009, 36 nasl.).

[7] V TIL-ke sa samozrejme ponúkajú definície jedných spojok pomocou iných (napr. Raclavský 2012, 252), ale ich odlišnosť nie je v tomto smere využitá.

z Fregeho prác. Jeho základnou motiváciou však nebolo budovať logické základy prirodzeného jazyka a jeho sémantickej analýzy, ale predovšetkým základy matematiky. Je známe, že voči zaužívanému spôsobu používania prirodzeného jazyka mal Frege značné výhrady v otázke jeho neprotirečivosti. Hlavné je však to, že matematické pravdy sú atemporálne a objektívne v zmysle epistemickej neutrálnosti, zatiaľ čo významy mnohých vetných spojok, ktoré sa dnes používajú v prirodzenom jazyku, sa vyvinuli[8] či mohli vyvinúť zo špecifických časových, kauzálnych, podmienkových, dôsledkových, vysvetľovacích a pod. spojok a môžu byť v niektorých použitiach citlivé napr. na časové vlastnosti spájaných viet – napr. na smerovanie času diania (Žigo 2013, Sokolová – Žigo 2014) alebo na zmenu subjektívneho poznania (Kesselová 2013, 14 nasl.). Veď podmienka (príčina) uskutočnenia nejakého deja v realite má podľa zakorenených predstáv o príčinnosti časovo predchádzať tomuto deju a poznanie človeka sa s časom môže meniť.

3. Jazykovedci o vetných spojkách

Podľa jazykovedcov sa vetné spojky podieľajú svojím významom na význame zloženého výrazu (súvetia) výraznejším spôsobom ako len spoluurčením pravdivostnej hodnoty výsledného súvetia. Podľa slovenskej morfológie významom spojky je špecifický spôsob spájania: „Významom čiže obsahom každej spojky je istý syntagmatický vzťah.“[9] Spojky môžu byť jednovýznamové alebo viacvýznamové. „Takmer všetky špecifické spojky sú jednovýznamové čiže jednofunkčné. Také sú napríklad spojky *ale, ba, nielen – ale aj*, *jednako*, *kdežto,* arch., *len čo, hoci, pretože*“ (MSJ, 679). Všeobecné spojky sú viacvýznamové – napr. *a, alebo, keď*.

Jazykoveda rozlišuje dva základné typy, druhy spojok podľa ich primárneho použitia – priraďovacie (parataktické) spojky a podraďovacie (hypotaktické) spojky, ale „sekundárne možno parataxou vyjadrovať aj podraďovanie

[8] Vývinom významov spojok nemyslíme vývoj samých abstraktných entít (tie sú mimo času), ale postupnú zmenu kódovania: Napr. ten istý spojkový výraz mal pôvodne aj časovo zjavne citlivý význam, ktorý bol neskôr nahradený významom s nízkou alebo žiadnou citlivosťou na časové parametre korelátov.

[9] Dvonč a kol. (1966, 679); kapitolu *Spojky* spracoval Jozef Ružička. Túto prácu budeme ďalej uvádzať pod skratkou „MSJ“.

a hypotaxou priraďovanie. Tak sa rozlišuje pravá parataxa a pravá hypotaxa od nepravej parataxy a nepravej hypotaxy“ (MSJ, 679). Niektoré spojky nie sú homogénne – sú hybridné, pretože môžu mať aj priraďovací, aj podraďovací význam, pričom jeden z nich je primárny, druhý sekundárny.

Celkovo je parataktických vetných spojok viac, ako uvažujú logici – ich presný počet však podľa Čermáka (2008, 316) nie je nijako ustálený. Keď k nim ešte prirátame sekundárne použitie hypotaktických spojok na vyjadrenie parataxy a zohľadníme rôzne špecifické významy vetných operátorov,[10] tak celkový počet odlišných špecifických významov takýchto spojok je oveľa väčší ako oných 16 pravdivostných funkcií. To je značný rozpor.

Stanovenie povahy príčinnosti zrejme už prekračuje kompetencie logiky i lingvistiky, ale nejaká kooperácia medzi vyjadrením podmienok či príčin deja a vyjadrením smerovania času môže pôsobiť na celkovú úlohu vetných operátorov bez zmeny princípu kompozicionality. Vyvstávajú pred nami otázky: Kde sú v tejto oblasti presne hranice logiky, čo ešte majú vetné spojky „zvažovať“ z logického hľadiska okrem pravdivostných hodnôt a rozličnej pozície podmienky? V čom majú ísť logici bližšie k lingvistom, aby sa dvojkoľajnosť výkladu spojok (operátorov) zmenšila tak, aby si dostatočne rozumeli?

4. Kritiky a návrhy

4.1. Kritika „formalistického“ výkladu spojok z pohľadu filozofov prirodzeného jazyka

Iný pohľad na spojky zaujali filozofi jazyka. Strawson končí svoj slávny článok *O referovaní* zásadnou myšlienkou:

> Presnú logiku výrazov prirodzeného jazyka neposkytujú ani aristotelovské, ani russellovské pravidlá, pretože prirodzený jazyk nemá presnú logiku. (Strawson 1950, 344; 1992, 146)

K tejto kritike formalizmu sa Strawson vrátil (1952, 57) a rozviedol ju. Spochybnil, že by výrazy prirodzeného jazyka, ako sú „ak“, „a“, „nie“, „každý“,

[10] Napr. pri vedľajších časových vetách utvorených pomocou spojok *keď, prv ako, odkedy, kým* Žigo konštatuje: „Pri rovnakej gramatickej prezentácii predikátov je celkový význam súvetia z časového hľadiska podmienený sémantikou spojok“ (Žigo 2010, 208).

„nejaký“, „alebo“, ktoré figurovali ako logické konštanty v úsudkových schémach logikov, boli vhodnými kandidátmi na túto úlohu, pretože tak, ako sú používané, im chýba stabilita a jednoduchosť významu, ktorá je potrebná pre úlohu logických konštánt.

Neale (1992, 5) zmodifikoval a dotvoril jeho zdôvodnenie asi takto: Ak hovorca H tvrdí prehovorom vetu formy *p alebo q*, tak hovorcu H budeme zvyčajne chápať tak, že nemal faktuálny, pravdivostno-funkcionálny podklad pre jednotlivé zložky vety. To znamená, že nevedel, ktorá zo zložiek – p, či q – je pravdivá. Filozof prirodzeného jazyka na základe tohto zistenia môže uzatvoriť, že prehovor formy *p alebo q*, pre ktorý nie sú splnené uvedené podmienky, je zneužitím jazyka. Časťou významu *p alebo q* je teda to, že takýto prehovor je použitý korektne len vtedy, keď hovorca nevie, ktoré z p, q je pravdivé. Ak táto podmienka nie je splnená, prehovor nemožno chápať tak, že vyjadruje pravdu. Preto filozof prirodzeného jazyka môže uzatvoriť, že by bolo vážnou chybou predpokladať, že význam slovenského slova „alebo“ je daný sémantikou logického operátora „∨“, ako je definovaný pravdivostnou tabuľkou. Naopak, sémantika spojky „alebo“ je určená použitím – skutočnou jazykovou praxou, ktorá sa nezhoduje s tabuľkovou analýzou.

4.2. Tri čítania logických spojok a nevyslovená zložka významu

Grice (1989, 22) prišiel s názorom, podľa ktorého aj formalisti, aj „informalisti“ (ako Strawson) sa mýlia v spoločnom predpoklade, že formálne prostriedky a ich náprotivky v prirodzenom jazyku sa z hľadiska významu rozchádzajú. Každá z týchto strán podľa Grica (1989, 24) venuje neadekvátnu pozornosť povahe a dôležitosti podmienok, ktoré ovládajú konverzáciu.

Podľa Grica, pokiaľ sa formalisti sústreďujú na formulovanie všeobecných vzorov logického vyplývania, formálne prostriedky sú nepopierateľne výhodnejšie ako ich náprotivky v prirodzenom jazyku a umožňujú prehľadne budovať logické systémy.

Na druhej strane *informalisti* podľa Grica (1989, 23) môžu argumentovať, že prirodzený jazyka plní aj iné dôležité úlohy mimo vedeckého skúmania a jeho úspešný používateľ nemusí poznať analýzu významu každého výrazu prirodzeného jazyka. Navyše veľa úsudkov a argumentov explicitne nevyužíva formálne logické prostriedky a napriek tomu sú mnohé z nich nepochybne správne. Preto by sme mohli hovoriť o logike prirodzených náprotivkov formálnych prostriedkov, pričom pravidlá pre formálne prostried-

ky nemusia platiť pre ich prirodzené náprotivky. Na vysvetlenie fungovania logiky prirodzeného jazyka Grice zaviedol pojem *implikatúry* ako nevyslovenej súčasti významu prehovoru, pre ktorý je rozhodujúca intencionalita hovorcu.

Neale (1992, 23) pokračoval v Gricových úvahách, navrhuje rozvinúť jeho návrhy a na príklade vetnej spojky „a" zvýrazňuje niektoré dôležité metodologické úvahy. Hoci nepopiera, že pri spájaní viet plní „a" rovnakú úlohu ako logická spojka „&", predsa v niektorých vetách plní inú úlohu, ako napríklad:

(1) Anna sa zosobášila s Jánom a Anne sa narodili dvojčatá.
(2) Grice sa mračil a študent sa začal triasť.

O niekom, kto vysloví (1), si podľa Neala budeme (typicky) myslieť, že za nevyslovenú súčasť významu prehovoru (jeho *implikatúru*) považoval to, že Anna sa zosobášila s Jánom prv, než sa Anne narodili dvojčatá.

Podobne budeme považovať za typický fakt, že podľa hovorcu, ktorý vysloví (2), je nevyslovenou súčasťou významu prehovoru to, že Gricovo mračenie nejako prispelo k študentovej triaške. Preto by sme mohli prísť k záveru, že „a" sa nie vždy chápe ako „&", pretože existujú prinajmenej tri druhy jeho čítania: *pravdivostno-funkcionálne, časové* a *kauzálne*. Riešenie tejto ambiguity či vysvetlenie prinajmenej troch odlišných významov spojky „a" sa tak situuje do pragmatiky, pričom dôležitá časť významu nie je pre adresáta ani vyslovená, ani vyjadrená, ale je súčasťou komunikačného zámeru hovorcu, ktorý nie je priamo prístupný adresátovi – ten ho musí odhaliť v jeho použití výrazov.

4.3. Môže logická sémantika prispieť k vysvetleniu rôznych vlastností toho istého vetného operátora?

Pohľad logikov na spojky len ako na pravdivostné funkcie sa zdá byť napriek jeho efektívnosti pri budovaní logických systémov naozaj príliš reštriktívny, zjednodušujúci a prinajmenšom neústretový k tomu, ako explikujú vetné operátory lingvisti. Veď na základe tabuľkovej definície spojky „a" ani nevieme, či je zlučovacia alebo odporovacia. Vlastne ani nemôžeme povedať, aká v tomto zmysle je, pretože sama pravdivostná funkcia nemá vlastnosti typu odporovania či zlučovania.

4.4. Kritika kontextualizmu

Kontextualizmus je predmetom kritiky z rôznych pozícií. Jednou z nich je sémantický minimalizmus.[11] Sémantický minimalizmus rešpektuje tézu, podľa ktorej obsah, ktorý sa vyjadrí použitím vety vo vzťahu k nejakému kontextu, môže byť bohatší ako doslovný význam vety, identifikovaný výlučne na základe sémantiky jej jednoduchých podvýrazov a spôsobu ich syntaktického usporiadania do zloženého výrazu. Rozhodujúcou otázkou je to, akým procesom sa nevypovedané zložky v doslovnom význame vety dostanú do skutočne vyjadreného obsahu. Kontextualizmus odpovie, že ide o pragmatický proces, čerpajúci z kontextu, ktorý dopĺňa vyjadrený obsah o nevypovedané zložky bez ohľadu na to, či sú alebo nie sú vynútené syntaktickou stavbou vety.

Sémantický minimalizmus prišiel s riešením, podľa ktorého musíme zohľadniť syntaktické zložky, ktoré hoci sú zamlčané, predsa sa dajú identifikovať na úrovni „skrytej" logickej formy. Ak ide o zamlčané voľné premenné, doplnenie vyjadreného obsahu sa deje *saturáciou*, teda priradením hodnôt „vopred pripraveným syntaktickým jednotkám vyskytujúcim sa v nejakej vrstve syntaktickej stavby vety" (Zouhar 2013, 31). Podľa sémantických minimalistov kontext môže vplývať najmenej dvomi spôsobmi.

Prvý významný vplyv kontextu je sémantický či pravdivostný (Zouhar 2011, 253) a niektorí ho nazývajú slabým pragmatickým vplyvom (Stanley 2007, 140). Pomocou neho sa doslovný význam vety doplní na komplexný vyjadrený význam jednak dodaním referentov pre obsiahnuté indexické a deiktické výrazy a zámená, a jednak priradením hodnôt pre výrazy, ktoré sú identifikovateľné až na úrovni skrytej logickej formy (napr. ohodnotením implicitných premenných) (Stanley 2007, 79). Podľa konzervatívnej demarkácie sémantiky a pragmatiky (sémantika – skúmanie významu, ktorý je nezávislý od mimojazykového kontextu, pragmatika – skúmanie aj významu, ktorý je závislý od mimojazykového kontextu) je tento vplyv už na úrovni pragmatiky. Minimálni sémantici či indexikalisti ho považujú za sémanticky relevantný, pretože až takéto doplnenie významu vety nám umožňuje identifikovať pravdivostné podmienky.

Druhý vplyv kontextu – silný pragmatický vplyv – už nie je relevantný sémanticky, ale nanajvýš pragmaticky (Zouhar 2011, 255). Ide o vplyv, ktorý

[11] Napr. Cappelen – Lepore (2005), Stanley (2007), u nás Zouhar (2011), ktorý je zástancom verzie sémantického minimalizmu – tzv. minimálneho indexikalizmu.

nie je vyvolaný žiadnymi stabilnými či variabilnými zložkami vyjadrených či zamlčaných lexikálnych jednotiek, ale len intenciou hovorcu.

V úvahách na pôde sémantického minimalizmu som však nenašiel analýzu vetných operátorov – vysvetlenie ich rôznych vyjadrených významov oproti doslovným významom, ktoré sú však také samozrejmé pre lingvistiku. Vetné spojky jednoducho nezaradili medzi kontextovo citlivé výrazy.

My vychádzame z hypotézy, že významy vetných operátorov, ktoré logici explikujú ako extenzionálne pravdivostné funkcie, interagujú v nejakej minimálnej miere s časovými, kauzálnymi či epistemickými vektormi, ktoré sú vyjadrené bez toho, aby sa explicitne použili samostatné časové, kauzálne či epistemické operátory (Gahér 2012, 25). Na to, aby sme mali vhodný priestor na prípadné konzistentné vysvetlenie tejto interakcie a na jej základe vysvetlili prisudzovanie rôznych vlastností významom spojok, ktoré sa však z hľadiska vyjadrenia pravdivostných podmienok neodlišujú, navrhujeme najprv rozšíriť skúmanie operátorov o skúmanie samej skladby typických formulácií týchto operátorov.

Základnou výstužou, na ktorú chceme naviazať naše vysvetlenie rôznych vlastností tabuľkovo zhodných vetných operátorov, je ich odlišná štruktúra typického vyjadrenia, ktorá môže byť inšpiratívna aj pre zdôvodnené zachytenie rôznych logických konštrukcií extenzionálne zhodných spojok. Inak povedané, predpokladáme, že logicko-sémantický prístup napríklad v podobe hyperintenzionálnej sémantiky (Tichý 1988; Duží – Jespersen – Materna 2010) má v tomto ohľade ešte nevyužitý explikačný potenciál a môže prispieť k požadovanému vysvetleniu rôznych vlastností vetných operátorov a ich odlišnej úlohe v skladbe súvetí.

5. Konštrukcie vetných operátorov

5.1. Základné vetné operátory

Ak rozšírime skúmanie spojok o sémantickú rovinu ich rôznej skladby pomocou vybraných základných spojok, tak by sa nám mohla otvoriť cesta k vysvetleniu rozmanitých vlastností vetných operátorov, o ktorých hovorí lingvistika.

Napríklad, ak by sme vzali ako základné logické spojky:

1. záporovú spojku ***nie je pravda, že*** **(*nie, ne-*)**

2. podmienkovú spojku ***keď*** **(*vtedy, keď; ak*),**

tak by sme vedeli definovať všetky logické vetné spojky, ako to už urobil Frege (1879). Mimochodom to, že si vybral podmienkovú spojku ako základnú, bolo nepochybne veľmi prezieravé, pretože keby si namiesto nej vybral zlučovaciu spojku *a,* resp. (ne)vylučovaciu spojku *alebo,* mohol dosiahnuť definovateľnosť všetkých výrokových spojok, ale nevedel by zachytiť prirodzenú (mnohorakú) povahu podmienkového spojenia. Určite by nestačilo konštatovať, že ide o tú spojku, pre ktorú sú všetky kombinácie výrokov A, B pravdivé okrem prípadu, v ktorom A je pravdivé a B je nepravdivé.

Určitú podobnosť nášho návrhu môžeme vidieť v intuicionistickom vysvetlení usudzovania, nazývanom Brouwerova-Heytingova-Kolmogorova explanácia. Podľa nej logická štruktúra propozície vyjadruje návod, ako dokázať túto propozíciu: Konjunkcia A, B je dokázaná dokázaním A a B separátne; disjunkcia A, B je dokázaná dokázaním jednej z propozícií A a B; implikácia A, B je dokázaná demonštráciou, ako previesť dôkaz A na nejaký dôkaz B atď. Zavedenie pravidiel tzv. prirodzenej dedukcie je blízke tomuto vysvetleniu (Plato 2009, 683).

Na rozdiel od Fregeho nám nejde primárne o definovanie spojok ako pravdivostných funkcií, ale aj o rôzne konštrukcie týchto funkcií, rôzne spôsoby ich identifikácie tak, ako vo všeobecnosti môžeme identifikovať tú istú funkciu rôznymi predpismi. To znamená, že najprv treba určiť typy základných konštrukcií.[12] Samozrejme pôjde o premenné pre atomárne vety (α, β, γ) a o jednoduché konštrukcie základných spojok. Ak by to boli podľa Fregeho vzoru negátor ($\neg$) a implikátor ($\rightarrow$),[13] tak by sme mali problém, lebo by sme nevedeli rozlíšiť rozličné spôsoby identifikácie tej istej pravdivostnej funkcie (1,0,1,1), a teda nevedeli by sme rozlíšiť význam spojky *ak-tak* od významu spojky *-len vtedy, keď-*.

O tomto rozdiele však predpokladáme, že je dôležitý. Preto navrhujeme miesto symbolu „→“, ktorý zvyčajne označuje priamo podmienkový operá-

[12] Nasledujúci text obsahuje len poloformálne explikácie a schémy, ktoré nebudú spĺňať prísne kritériá výstavby formálneho systému, pretože chceme, aby bola zvýraznená základná hypotéza. Výraz *konštrukcia* budeme používať v určitom intuitívnom význame, nepopierame však, že tento význam je inšpirovaný pojmom konštrukcie objektov rôznych typov, ako je rozpracovaný v TIL-ke.

[13] Frege používal dvojdimenzionálnu symboliku a základnou spojkou okrem negátora bola obrátená implikácia.

tor, používať dva základné symboly „∪→“, „→∪“, ktoré označujú tú istú podmienkovú pravdivostnú funkciu (1,0,1,1), ale pomáhajú vyjadriť jej odlišné formulácie, ktoré by mali naznačiť aj dôvod ich odlišných sémantických čŕt.

Ak predpokladáme lineárny ľavo-pravý zápis, priehlbina symbolizuje pozíciu podmienky v zloženom súvetí v jazyku tak, ako to robia v súvetiach výrazy *ak,* resp. *keď*, ktoré sú zložkami podmienkových operátorov. Symbol „∪→“ použijeme na označenie podmienkového spojenia, v ktorom podmienka je prvou zložkou a podmienené druhou zložkou súvetia. Symbol „→∪“ použijeme na označenie podmienkového spojenia, v ktorom podmienka je druhou zložkou a podmienené je prvou zložkou súvetia. Obidva symboly „∪→“ a „→∪“ teda identifikujú tú istú pravdivostnú funkciu, a pritom zachytávajú zásadný rozdiel, ktorý je medzi *dostatočnou* a *nutnou podmienkou*.[14] Smer podmieňovania je vždy orientovaný od pozície podmienky, čo sa môže pri zachytení konštrukcie spojky -*len vtedy, keď*- zdať protirečivé, pretože šípka je orientovaná proti smeru podmieňovania. Takéto označenie je vecou konvencie a niektoré symboly pre implikáciu (šípka, podkova a pod.) mohli (mali) navodzovať čosi ako smerovanie relácie medzi významami viet, hoci sama označená pravdivostná funkcia nič také nevykazuje. Samozrejme, môžeme si zvoliť označenie bez naznačovania akéhokoľvek smerovania. My sme postupovali konzervatívne a šípku sme tam ponechali aj preto, aby sa zachovala čiastočná kontinuita označovania pravdivostných funkcií.

Vzťahy príčinnosti, dôvodu, vysvetlenia a pod. považujeme za mimologické vzťahy a ich intuitívne pochopenie predpokladá, že vieme rozlíšiť ich nezameniteľné koreláty: príčinu od účinku, dôvod od dôsledku, vysvetlenie od vysvetľovaného. Ak robíme napríklad kauzálnu predikciu, tak si nesmieme zamieňať príčinu s účinkom. V podmienkovej vete, ktorú na to použijeme, musíme opis príčiny situovať do pozície podmienky, nie podmie-

[14] Pojmy dostatočnej a nutnej podmienky sa niekedy považujú za symetrické v tom zmysle, že „P je nutná podmienka pre Q vtt Q je dostatočná podmienka pre P“ (Oddie – Tichý 1980, 227). V stati Gahér (2012) sme sa pokúsili demonštrovať, že takáto symetria nie je vo všeobecnosti platná – napr. neplatí pre empirickú oblasť. Napríklad v súvetí „Ak sa maslo zohrieva, tak sa topí“ pri jeho opisnom použití vyjadrujeme dostatočnú podmienku topenia masla, ale topenie masla nie je nutnou podmienkou jeho zohrievania.

neného, inak to nebude sémanticky korektná predikcia.[15] Okrem zhodného jadra významu ako pravdivostných funkcií majú spojky *ak-tak* a *len vtedy, keď* odlišnú sémantickú črtu, signalizovanú odlišnou pozíciou podmienky, na ktorú je naviazaná aj prípadná časová orientácia a cieľ použitia (predikcia, konštatácia, vysvetľovanie). V empirickej oblasti musí podmienka vždy časovo predchádzať podmienené. Na vyjadrenie konštrukcie spojky, ktorá (konštrukcia) je v logickom priestore ako abstraktná entita bezčasová, sa môžeme pozerať ako na itinerár nadväznosti fáz diania či procesu poznávania významu súvetia v čase.

5.2. Spojka ak-tak *– vyjadrenie dostatočnej podmienky*

V prirodzenom jazyku používame hypotaktickú spojku (presnejšie *korelovanú dvojicu*) *ak-tak* aj na vyjadrenie parataktického spojenia (ide o tzv. nepravú hypotaxu). Jej dubletou je slovo *pokiaľ* v úlohe uvádzania podmienkovej vety (MSJ, 733). Opisne vyjadrené: Schémou $\alpha \cup\!\!\rightarrow \beta$ označujeme zloženú štruktúru súvetí, kde význam prvej vety je vo vzťahu **dostatočnej podmienky** k významu druhej vety. Platí všeobecný princíp, podľa ktorého podmienkové súvetia máme „čítať“ vždy v smere vektora podmienky, t. j. od podmienky k podmienenému (Gahér 2012a, 25). Preto aj súvetie utvorené pomocou spojky *ak-tak* máme tak „čítať“ aj z hľadiska určenia pravdivostných podmienok: Pravdivosť prvej (podmieňujúcej) vety v prípade pravdivosti celého súvetia zabezpečuje pravdivosť druhej (podmienenej) vety.

Príklad súvetia, v ktorom je uvedená dostatočná podmienka:

(3) Ak niekto nájde stratenú vec, tak (ten) je povinný ju vydať vlastníkovi.[16]

5.3. Spojka len vtedy, keď *– vyjadrenie nutnej podmienky*

Na identifikáciu tej istej pravdivostnej funkcie ako pomocou spojenia *ak-tak* používame v prirodzenom jazyku aj spojenie *len vtedy, keď* (*iba ak;*

[15] Vetu „Ak prší, ulice budú mokré“ môžeme považovať za predikciu, ale vetu „Ak sú ulice mokré, tak bude pršať“ nemôžeme považovať ani za predikciu, ani za vysvetlenie. Vyjadrenie vysvetlenie si vyžaduje zmenu časovej orientácie: „Ak sú ulice mokré, tak pršalo“

[16] Parafráza ustanovenia § 135 ods. 1 zákona č. 40/1964 Zb. Občiansky zákonník.

len ak), pričom ide o vyjadrenie odlišného pojmu – pojmu **nutnej podmienky**. Schémou $\alpha \rightarrow\cup \beta$ označujeme zloženú štruktúru súvetí, v ktorej význam druhej vety je vo vzťahu **nutnej podmienky** k významu prvej vety. Pozícia podmienky je zamenená voči poradiu v štruktúre dostatočnej podmienky a smer podmienkového vzťahu je obrátený. Aj takéto súvetie z hľadiska pravdivostných podmienok máme „čítať" v smere od podmienky: Pravdivosť **druhej** (podmieňujúcej) vety v prípade pravdivosti celého súvetia zabezpečuje pravdivosť **prvej** (podmienenej) vety. Skrátene, pravdivosť podmieňujúcej vety **obmedzuje**, **ale nezabezpečuje** pravdivosť podmienenej vety.

Príklad:

(4) Rastlina rastie len vtedy, keď má dostatok vlahy.

Na okraj poznamenávame, že parataktická spojka *iba* má obmedzujúcu funkciu – „vyjadruje odpor obmedzením" (MSJ, 721). Jazyková prax (pozri slovniky.korpus.sk/?w=iba) dnes už stiera rozdiely medzi týmto významom spojky *iba* a významom spojky *len* ako tiež vyjadrujúcej „odporovací vzťah obmedzením platnosti predchádzajúcej vety" (slovniky.korpus.sk/?w=len). Demonštrujú to aj prekladateľské slovníky do angličtiny, nemčiny, francúzštiny, ruštiny. Vzhľadom na štandardné vyjadrovanie vzťahu logickej rovnocennosti (ekvivalentnosti) medzi výrokmi, utvorenými pomocou spojenia *vtedy a len vtedy, keď*, kde slovo *len* je integrálnou zložkou spojkového spojenia, preferujeme práve spojku *len*.

V stati *Len* (Gahér 2012b) sme navrhli všeobecnú definíciu jednotného významu tejto spojky pre veľký rozsah jej použitia. Keďže táto stať nadväzovala na problematiku rozlíšenia dostatočnej a nutnej podmienky, nemohli sme tieto pojmy použiť ako explikát pre význam slova *len*. Výsledná definícia významu slova *len* ako štruktúrovaného operátora bola schémou formulovanou v predikátovej logike druhého rádu. Ak by sme použili túto schému, tak by sme problematiku vetných spojok v tejto fáze skúmania nielen príliš skomplikovali, ale dostali by sme sa do bludného kruhu. Keďže v inej stati (Gahér 2012a) sme podali akceptovateľnú explikáciu pojmov dostatočná a nutná podmienka, tu ich použijeme ako základné explikačné nástroje podobne, ako to urobil Lepore (2003, 89). K spojke *keď (ak)*, ktorá signalizuje dostatočnú podmienku, teda pridáme aj spojkový výraz *len vtedy, keď*, ktorý bude signalizovať nutnú podmienku.

Pri spojke *keď* jazykovedci oprávnene nástoja na tom, že pri hypotaktickom (podraďovacom) použití „má všeobecný časový význam" (MSJ, 723),

ktorý sa stráca pri jej nepravom hypotaktickom použití ako podmienkovej spojky – v úlohe parataktického spojenia. Logici s ňou pracujú predovšetkým ako s podmienkovou spojkou bez časového významu, čo pri aplikácii na bezčasové analytické súvetia (v oblasti logiky, matematiky, právd na základe sémantického významu a pod.) nevedie k žiadnym diskrepanciám. Použitý gramatický čas sa v týchto prípadoch nazýva *gnómický*[17] – významy takýchto súvetí sa považujú za mimočasové alebo časovo všeobecné fakty, preto sa zdá signalizácia času *nadbytočná*. Pri aplikácii logiky na vety o empirickom dianí toto „odčasovanie" podmienkovej spojky začína škrípať. V nadväzujúcej stati by sme chceli vysvetliť, prečo je to tak, i to, ako podmienková spojka kooperuje so smerovaním času fyzikálneho diania a so smerovaním gramatického času. Pri empirických vetách spojka *keď* akoby znovu dostala časový význam. Toto na prvý pohľad podivné strácanie a znovuzískavanie časového významu sa týka aj spojky *ak-tak* a spojkové výrazu *len vtedy, keď*.

5.4. Popieranie – operátor nie je pravda, že

Aj s popieraním sú spojené mnohé problémy. Diskusie o správnom popieraní, negovaní prebiehali už v antike medzi stoikmi, megarikmi a Aristotelovými žiakmi (bližšie o tom Gahér 2000). To, čo lingvisti nazývajú vytyčovacou hodnotiacou časticou *nie,* slúži na vyjadrenie prostého záporu vo vete alebo znásobuje zápor, ktorý je vo vete už inak vyjadrený (MSJ, 782). Na podobné ciele slúži aj morféma *nie* v záporových tvaroch slovesa *byť* (MSJ, 471). Spojenie *nie je pravda, že* je operátorom, ktorý logici nazývajú *negátor*. Predpona *ne* pri záporových tvaroch slovies plní rovnakú funkciu. Negátor (¬) budeme považovať za základnú singulárnu vetnú spojku. Jej význam môžeme opísať (v metajazyku) aj ako **opak je pravdou.**

5.5. Rôzne významy jednej a tej istej spojky

Hoci sa výklad vetných spojok zvyčajne začína paradigmatickou zlučovacou spojkou *a*, predsa má táto spojka podľa lingvistov veľa odlišných významov, ktoré nie sú zlučovacie – stupňovací, odporovací, vylučovací, dôsledkový a prípustkový význam. Ako to ide dohromady – na jednej strane máme čiro zlučovací význam spojky, na druhej strane máme veľa od zlučo-

[17] Žigo (2010, 185): „... deje vyjadrené gnómickým prézentom, prítomným časom v jeho sekundárnej funkcii majú všeobecnú časovú platnosť, vzhľadom na súradnicu času sú symetrické."

vacieho vzťahu odlišných významov použitia tej istej spojky? Ako môžeme zo základného zlučovacieho významu prísť k takým odlišným významom spojky *a*? Môže k tomu vysvetleniu prispieť aj sémantika, alebo ide výlučne o záležitosť pragmalingvistickú?

Odsúvanie sémantických problémov do kompetencie pragmatiky môžeme prirovnať k práci *lenivého sémantika,* ako hovorievala Hajičová. Toto odsúvanie sémantických problémov, ktoré kritizuje aj Kripke (2011, 342), sa často opiera o tzv. *chybu pragmatického odpadkového koša*, ktorú môžeme nazvať maximou „lenivého sémantika":

(LenivýSem) „… *robme sémantiku takú jednoduchú, ako je len možné, všetko ostatné je pragmatika, ale o tom nemáme premýšľať*" (Grice 1989, 4).

My sa budeme riadiť inou maximou – maximou „usilovného sémantika":

(UsilovnýSem) „*Robme sémantiku tak, aby sme vysvetlili všetko to, čo je možné vysvetliť na jej základe bez použitia pragmatiky, aj keď táto teória nebude taká jednoduchá, ako by sme očakávali.*"

Výsledky takto budovanej sémantiky môžu byť inšpiratívne nielen pri preverovaní filozofických problémov a jemnejšom chápaní jazyka, ale aj pri presnejšej demarkácii sémantickej (systémovo-lingvistickej) a pragmatickej stránky významu výrazov, a teda aj pre empirických lingvistov.

5.6. Jednoduché a zložené konštrukcie spojok

Naše vysvetlenie rozličných významov tej istej spojky sa opiera aj o to, že v skutočnosti ide o dve odlišné úrovne konštruovania[18] spojok. Na bazálnej úrovni ide o primitívne, základné významy spojok, ako sú zlučovacia spojka ***a***, záporová spojka ***nie*** a dve podmienkové spojky ***ak-tak*** a ***len vtedy, keď*** *(iba ak)*. Tieto významy spojok nie sú navzájom definovateľné a zastupiteľné a zodpovedajú im jednoduché konštrukcie.[19] Z týchto spojok spolu s ve-

[18] Výraz „konštrukcia" nebudeme používať v technickom zmysle.

[19] Hoci pomocou konjunktora a negátora sú definovateľné ostatné extenzionálne logické spojky, nestačí to na odlíšenie „nadstavbových" sémantických čŕt spojky *ak-tak* od spojky *len vtedy, keď*.

tami vieme skonštruovať všetky ostatné sémantické významy spojok (zložené súvetia) s ich odlišnými sémantickými vlastnosťami. Práve rozličné zložené typické (priezračné) formulácie nám umožnia spolu so zohľadnením vektorov času, kauzality a relevancie[20] vysvetľovať rôzne sémantické správanie tej istej spojky.

5.7. Spojka a *v rýdzo zlučovacom význame*

Predpokladáme, že je len jeden jednoduchý znak pre základné spojky. Nech symbol ∧ identifikuje akoby priamo (primitívnym zmyslom) pravdivostnú funkciu určenú štvoricou funkčných hodnôt (1,0,0,0). Tento význam bude mať spojka *a* použitá v čiro zlučovacom význame bez „nabalenia sa" o nejaký dodatočný význam. V prirodzenom jazyku majú tento význam aj niektoré použitia spojok *i*, *aj*, *...aj...aj*.

Opisne môžeme ich význam identifikovať ako **oba (výroky) sú pravdivé**, pričom význam takto zloženého konjunktívneho súvetia nemá žiadnu skrytú či zamlčanú časť – napríklad vyjadrenie odporovania.

Príklad:

(5) Sokrates sa prechádza a rozpráva.

Takto použitá spojka *a* je komutatívna – poradie podvýrokov nie je dôležité.

6. Konštrukcie významov zložených vetných operátorov

6.1. Podmienkové operátory

Pomocou pojmov **popieranie, dostatočné podmieňovanie**, **nutné podmieňovanie a zlučovanie** môžeme definovať všetky ostatné typy konštruovania zložených propozícií – v jazyku lingvistiky všetky ostatné typy spojenia viet do priraďovacích súvetí.

Na prvý pohľad sa to môže zdať ako úzka báza na vysvetlenie všetkých takých vlastností spojok, ako je zlučovanie, odporovanie, vylučovanie, obmedzovanie, pripúšťanie a pod., o ktorých podrobne hovoria jazykovedci.

---

[20] Vysvetlenie pojmov vektor času, kauzality a relevancie pozri v Gahér (2012a).

V prvom kroku navrhneme definičné konštrukcie významov ostatných (v prirodzenom jazyku používaných) binárnych spojok. Zatiaľ nebudeme využívať spojku *a* použitú v rýdzo zlučovacom význame. Máme osem základných logických možností kombinovania záporu (singulárnej spojky), aplikovaného na jednoduché vety zastúpené premennými a dvoch podmienkových spojení (dostatočného a nutného podmieňovania) medzi jednoduchými vetami (niektoré z nich nemusia byť významom žiadneho zavedeného jednoduchého výrazu v prirodzenom jazyku):

$\alpha \cup\!\!\rightarrow \beta$
$\alpha \cup\!\!\rightarrow \neg\beta$
$\neg\alpha \cup\!\!\rightarrow \beta$
$\neg\alpha \cup\!\!\rightarrow \neg\beta$
$\alpha \rightarrow\!\!\cup \beta$
$\alpha \rightarrow\!\!\cup \neg\beta$
$\neg\alpha \rightarrow\!\!\cup \beta$
$\neg\alpha \rightarrow\!\!\cup \neg\beta$

Prvú a piatu konštrukciu bez záporov sme vzali ako základne konštrukcie rýdzej dostatočnej, resp. číro nutnej podmienky.

6.1.1. Vylučovacia podmienka

Vzťah[21] medzi významami viet v pozíciách α, β v pravdivom súvetí tvaru $\alpha \cup\!\!\rightarrow \neg\beta$ budeme opisovať ako α **je vylučovacia podmienka pre** β.

Definícia 1: α je **vylučovacia** podmienka pre β vtt[22] pravdivosť α v prípade pravdivosti výroku ($\alpha \cup\!\!\rightarrow \neg\beta$) vylučuje pravdivosť β.

Príklad:

(6) Ak Richard príde včas do práce, tak nedostane pokarhanie.

[21] Prísne vzaté, spojky nevyjadrujú tradične chápané vzťahy ako relácie medzi indivíduami.

[22] Skratku „vtt“ budeme používať v definíciách pre spojenie „vtedy a len vtedy, keď“.

6.1.2. Negatívna dostatočná podmienka

Vzťah medzi významami viet v pozíciách α, β v pravdivom súvetí tvaru ¬α ∪→ β budeme opisovať ako α **je negatívna dostatočná podmienka pre** β.

Definícia 2: α je **negatívna dostatočná** podmienka pre β vtt nepravdivosť α v prípade pravdivosti výroku (¬α ∪→ β) zabezpečuje pravdivosť β.

Príklad:

(7) Ak šofér neuhne doprava, tak narazí do protiidúceho auta.

Schéme tohto operátora zodpovedajú v prirodzenom jazyku spojky ***inak, ináč*,** ktoré majú podľa jazykovedcov vysvetľovací význam (MSJ, 722).

(7*) Šofér uhne doprava, inak narazí do protiidúceho auta.

6.1.3. Negatívna vylučovacia podmienka

Vzťah medzi významami viet v pozíciách α, β v pravdivom súvetí tvaru ¬α ∪→ ¬β budeme opisovať ako α **je negatívna vylučovacia podmienka pre** β.

Definícia 3: α je **negatívna vylučovacia** podmienka pre β vtt nepravdivosť α v prípade pravdivosti výroku (¬α ∪→ ¬β) vylučuje pravdivosť β.

Príklad:

(8) Ak Ján nespraví skúšku, tak nepostúpi do ďalšieho ročníka štúdia.

6.1.4. Negatívna nutná (negatívna obmedzujúca) podmienka

Vzťah medzi významami viet v pozíciách α, β v pravdivom súvetí tvaru α →∪ ¬β budeme opisovať ako α **má ako nutnú podmienku nepravdivosť** β alebo v aktívnom mode ako β **je negatívna nutná podmienka pre** α.

Definícia 4: β je **negatívna nutná** podmienka pre α vtt nepravdivosť β v prípade pravdivosti výroku (α →∪ ¬β) obmedzuje (podmieňuje, ale nezabezpečuje) pravdivosť α.

Príklad:

(9) Daniel vystúpi na Mont Everest len vtedy, keď ho nezastihne monzún.

6.1.5. Nutná (obmedzujúca) podmienka opaku

Vzťah medzi významami viet v pozíciách α, β v pravdivom súvetí tvaru ¬α →∪ β budeme opisovať ako **negácia α má ako nutnú podmienku** β alebo v aktívnom mode ako β **je nutnou podmienkou pre negáciu** α.

Definícia 5: β je **nutnou** podmienkou pre negáciu α vtt β v prípade pravdivosti výroku (¬α →∪ β) obmedzuje (podmieňuje, ale nezabezpečuje) nepravdivosť α.

Príklad:

(10) Slovensko nepostúpi zo skupiny majstrovstiev sveta v hokeji len vtedy, keď prehrá posledné dva zápasy.

6.1.6. Negatívna nutná (obmedzujúca) podmienka opaku

Vzťah medzi významami viet v pozíciách α, β v pravdivom súvetí tvaru ¬α →∪ ¬β budeme opisovať ako **negácia α má ako nutnú podmienku nepravdivosť** β alebo v aktívnom mode ako β **je negatívna nutná podmienka pre negáciu** α.

Definícia 6: β je **negatívna nutná** podmienka pre negáciu α vtt nepravdivosť β v prípade pravdivosti výroku (¬α →∪ ¬β) obmedzuje (podmieňuje, ale nezabezpečuje) nepravdivosť α.

Príklad:

(11) Slovensko nepostúpi zo skupiny na majstrovstvách sveta v hokeji len vtedy, keď nezíska z posledných dvoch zápasov ani jeden bod.

Poznámka: Vzťah negatívnej nutnej podmienky opaku je vlastne vzťahom dostatočnej podmienky v obrátenom smere bez zložiek popierania: ¬α →∪ ¬β **je rovnocenné s** β ∪→ α. Veta v pozícii pôvodnej podmienky sa zachovala, zo zápornej vety sa však zmenila na kladnú vetu.

Príklad:

(12) Ak Slovensko získa z posledných dvoch zápasov aspoň jeden bod, tak postúpi zo skupiny na majstrovstvách sveta v hokeji.

6.2. Vylučovacia spojka alebo*, resp.* buď-alebo

V prirodzenom jazyku je veľmi frekventovaná spojka *alebo*. Lingvisti ju považujú primárne za spojku s vylučujúcim významom (MSJ, 710). V systémoch logiky bol historicky naozaj primárny jej vylučovací význam (Gahér 2006, 147), ktorý je dnes výraznejšie vyjadrovaný dvojčlennými spojkami *buď–alebo, buď–buď, alebo–alebo* (MSJ, 711). Touto spojkou je identifikovaná pravdivostná funkcia (0,1,1,0). Jej význam môžeme opísať aj spojením **práve jeden z dvoch (výrokov) je pravdivý**.

Vylučovací význam tejto spojky môžeme opísať nasledovne: Pravdivosť jedného vylučuje pravdivosť druhého (klad jedného vylučuje klad druhého), a nepravdivosť jedného vylučuje nepravdivosť druhého (zápor jedného vylučuje zápor druhého), čo môžeme zachytiť konštrukciou z doteraz uvedených spojok (a rýdzo zlučovacej konjunkcie):

Definícia 7: (buď α alebo β) $=_{df}$ (((α ∪→ ¬β) ∧ (β ∪→ ¬α)) ∧ ((¬α ∪→ β) ∧ (¬β ∪→ α)))

Je zrejmé, že hoci sa s definičnou skratkou v kontextovej definícii operátora (v definiende) ľahšie manipuluje, predsa jej presný význam je zobrazený konštrukciou definiensa. Táto definícia môže byť vhodným podkladom na vysvetlenie postoja, ktorý podľa analógie Gricovho návrhu je zaviazaný mať hovorca súvetia s touto logickou štruktúrou: Nevie o pravdivosti ani o nepravdivosti niektorého z elementárnych výrokov (netvrdí sa tam konjunkcia jedného elementárneho výroku s nejakým, hoc zloženým výrokom), ale vie, že sa navzájom vylučujú.

Fakt, že je táto definícia ťažkopádna, nahráva tomu, aby sa táto spojka považovala za základnú, nedefinovanú spojku – ako to urobil Chrysippos (Gahér, 2006, 147). V takom prípade je zložité opakované odporovanie si členov (vzájomné vylučovanie) súvetia integrované v dvojčlennom operátore.

Príklad:

(13) Buď prišiel Ján neskoro na zápas, alebo nastúpil v základnej zostave.

V rovnakom vylučovacom význame sa používa aj parataktická spojka *či-či* (MSJ, 716).

6.3. Zlučovacia spojka alebo

V logike, matematike, programovaní a príbuzných disciplínach sa dnes spojka *alebo* používa prevažne v nevylučovacom význame. Ňou identifikovaná pravdivostná funkcia je určená štvoricou (1,1,1,0), ktorú môžeme metajazykovo opísať spojením **aspoň jeden z dvoch (výrokov) je pravdivý**.

Nevylučovacia črta významu tejto spojky je odhaliteľná konštrukciou z doteraz uvedených spojok: Nepravdivosť jedného výroku je dostatočnou podmienkou pre pravdivosť druhého výroku, čo môžeme opísať tak, že najmenej jeden z nich je pravdivý, a preto nie je medzi nimi vzťah vylučovania.

Definícia 8: $(\alpha \text{ alebo } \beta) =_{df} (\neg\alpha \cup\!\!\rightarrow \beta) \wedge (\neg\beta \cup\!\!\rightarrow \alpha)$.

Táto definícia môže byť vhodným podkladom na vysvetlenie postoja, ktorý podľa Grica je zaviazaný mať hovorca súvetia s touto logickou štruktúrou: Nevie o pravdivosti niektorej z elementárnych viet, ale vie, že aspoň jedna je pravdivá.

Príklad:

(14) Natália sa venuje koňom alebo rozpráva.

Niekedy nevylučovací význam spojky *alebo* je „prebitý“ vylučovacím významom prísudkov spájaných viet, čo spôsobuje zdanie, že aj v tomto prípade ide o vylučovaciu spojku. Napríklad v súvetí:

(15) Kopytníky sú párnokopytné alebo nepárnokopytné.[23]

Podobne, ak by sme v prípade súvetia

(16) Natália obriadi koňa alebo pôjde so psami na prechádzku,

predpokladali, že Natália má čas len na jednu z opísaných činností, tak by sme ju interpretovali ako vylučujúcu disjunkciu, hoci takou nie je, a vyluču-

[23] Predpokladáme bisekciu súboru kopytníkov, t. j. pojmy párnokopytník a nepárnokopytník sú pre univerzum kopytníkov komplementárne.

júci význam celého súvetia je podmienený dodatočným predpokladom. Problém rozlíšenia číro vylučovacieho a nevylučovacieho významu disjunktívneho spojenia je typický nielen pre slovenčinu, ale aj pre mnohé indoeurópske jazyky (trpí ním napr. aj angličtina).

Podľa lingvistov má parataktická spojka *či-či* aj zlučovací význam (MSJ, 716) a parataktický spojovací výraz *jednak- (a) jednak* má tiež zlučovací význam (MSJ, 722).

6.4. Opakovaná spojka ani-ani

Pomocou rýdzo zlučovacej spojky *a* a dvakrát použitého záporu *nie je pravda, že* môžeme konštruovať spojku, ktorá identifikuje pravdivostnú funkciu (0,0,0,1). Opisne môžeme tento vzťah identifikovať ako **žiadny (výrok) nie je pravdivý.**

Definícia 9: Nie je pravda, že α; a nie je pravda, že β $=_{df} \neg\alpha \wedge \neg\beta$.

V slovenčine tejto konštrukcii zodpovedá aj súvetie utvorené opakovaním spojky ***ani***, ktorá signalizuje už v sebe integrovaný zápor, pričom opakovaný zápor vo funkcii zdôraznenia je syntakticky vyjadrený pred slovesami jednotlivých vetných zložiek. Možno aj preto vzniká váhanie, či ide o spojku so zlučovacím, resp. rozlučovacím významom (MSJ, 711). V jazykoch, kde nie je dovolený druhý, zdôrazňovací zápor, sa záporová predpona pri slovesách nevyskytuje.

Príklad:

(17) Ani Slovan sa nestal majstrom Slovenska, ani Trnava nevyhrala Slovenský pohár.

Pravdivostnú funkciu (0,0,0,1) môžeme konštruovať aj odlišným spôsobom. Napríklad:

Definícia 10: (ani α ani β) $=_{df} \neg(\neg\alpha \cup\!\!\rightarrow \beta) \wedge \neg(\neg\, \beta \cup\!\!\rightarrow \alpha)$.

Túto odlišnú konštrukciu by sme použili vtedy, keď by sme nemali otestované jednotlivé podvýroky, ale by sme disponovali informáciou, že nepravdivosť jedného z výrokov negarantuje pravdivosť druhého.

6.5. Spojka a *vo významoch nie rýdzo zlučovacích*

6.5.1. Prípustkový význam spojky a

Okrem predpokladu primitívneho významu spojky *a* ako rýdzo zlučovacej spojky bez akýchkoľvek momentov odporovania, vylučovania a pod., budeme práve na základe dvoch podmienkových vzťahov a záporu vysvetľovať dodatočné významy spojky *a*, ktoré majú odporovací, vylučovací či prípustkový moment. Podobne sa dá postupovať aj v prípadoch ostatných vetných spojok.

Uvedieme také rozličné konštrukcie pravdivostnej funkcie (1,0,0,0), ktoré nám umožnia vysvetliť rozdiely v povahe spájacieho vzťahu pomocou spojky *a*. V tejto fáze skúmania nebudeme reflektovať rôzne druhy podmieňovania (pripúšťanie, odporovanie či vylučovanie) určené výlučne alebo primárne samými významami spájaných viet.

Pravdivostnú funkciu (1,0,0,0 – konjunktor) môžeme konštruovať pomocou záporu a dostatočného a nutného podmieňovania viacerými spôsobmi.

Formulácie súvetí tvaru $\neg(\alpha \cup\!\!\rightarrow \neg\beta)$ sú jednou z možných vyjadrení pravdivostnej funkcie konjunkcie (1,0,0,0). Ide o vyjadrenie **popretia vylučovacieho vzťahu** medzi prvým výrokom a negáciou druhého výroku. Tento vzťah môžeme vyjadriť aj ako α **pripúšťa** β.

Príklad:

(18) Nie je pravda, že ak kapitán Scott vedel o nástrahách cesty na Južný pól, tak sa na ňu nevydal.

V prirodzenom jazyku tejto formulácii zodpovedajú spojky ***hoci, hoc***, resp. korelovaná dvojica ***hoci-predsa***, ktoré sa považujú za základné prípustkové spojky (MSJ, 719). Sme presvedčení, že tejto konštrukcii v skutočnosti zodpovedajú aj spojky ***aj keď, i keď***, ktoré podľa jazykovedy vyjadrujú krajnú podmienku (MSJ, 705) a spojky ***a jednako, ale jednako***, ktoré podľa lingvistov majú odporovací význam s odtienkom prípustky (MSJ, 700). Spojka *ale* má základný odporovací význam – často zbadateľný až v predikátovo-logickej štruktúre.

(18*) Hoci kapitán Scott vedel o nástrahách cesty na Južný pól, predsa sa na ňu vydal.

6.5.2. Spojka a *vo význame neobmedzovania*

Konštrukciou súvetia typu $\neg(\alpha \rightarrow\cup \neg\beta)$ tiež konštruujeme funkciu (1,0,0,0), ale ide o **popretie vzťahu obmedzovania** – popretie toho, že nepravdivosť druhej vety obmedzuje pravdivosť prvej vety. Tento vzťah môžeme opísať aj ako pravdivosť druhej vety **neobmedzuje** pravdivosť prvej vety. V prirodzenom jazyku tejto konštrukcii zodpovedá zložené spojenie *Nie je pravda, že ... len vtedy, keď nie ....*

Príklad:

(19) Nie je pravda, že ujmu vzniknutú ubytovateľovi predčasným zrušením ubytovania je ubytovaný povinný nahradiť len vtedy, keď nemohol ubytovateľ ujme zabrániť.[24]

Podľa tejto formulácie fakt, že ubytovateľ nemohol ujme zabrániť, neobmedzuje fakt, že ubytovaný je povinný nahradiť ujmu vzniknutú ubytovateľovi predčasným zrušením ubytovania. V našej právnej úprave platí opak – je medzi nimi vzťah obmedzovania.

Reglementáciou (19) v tvare s číro zlučovacou spojkou *a* vznikne:

(19*) Ubytovaný je povinný nahradiť ujmu vzniknutú ubytovateľovi predčasným zrušením ubytovania a ubytovateľ mohol ujme zabrániť.

V (19*) sa však stratí napríklad identifikácia pôvodne obmedzujúcej podmienky. Preto pôvodná formulácia na rozdiel od novej formulácie neskrýva žiadnu časť komplexného významu.

6.5.3. Spojka a *vo význame popretia nepripúšťania*

Z hľadiska pravdivostných podmienok konjunktívne spojenie súvetí môžeme dosiahnuť aj formuláciou súvetia v tvare $\neg(\neg\alpha \cup\rightarrow \beta)$ – ide o **popretie vzťahu negatívnej vylučovacej podmienky**. Povedané inými slovami, takto utvoreným súvetím popierame, že nepravdivosť prvej vety vylučuje pravdivosť druhej vety, čo môžeme vyjadriť opisne aj ako **nie je pravda**, že nepravdivosť prvej vety **nepripúšťa** pravdivosť druhej vety.

Príklad:

(20) Nie je pravda, že ak neprší, tak ulice sú suché.

---

[24] Negácia parafrázy § 759 ods. 1 zákona č. 40/1964 Zb. Občiansky zákonník.

6.5.4. Spojka a *vo význame popretia obmedzovania*

V súvetí tvaru $\neg(\neg\alpha \rightarrow\cup \beta)$ ide o **popretie vzťahu obmedzovania opaku**, čo môžeme vyjadriť aj takto: Pravdivosť druhej vety neobmedzuje nepravdivosť prvej vety.

Príklad:

(21) Nie je pravda, že ulice nie sú suché len vtedy, keď prší.

Pripomíname, že výsledný priebeh hodnôt pravdivostnej funkcie, vyjadrenej uvedenými typmi konštrukcií súvetí, je ten istý – (1,0,0,0), to znamená, že výsledkom je vzťah zlučovania, ktorý je však dosiahnutý popieraním či už vylučovania, alebo obmedzovania.

Pochopiteľne, spojka *a* použitá v niektorom z uvedených významov, ktoré sú odlišné od rýdzo zlučovacieho významu, nie je vo všeobecnosti komutatívna – inak povedané, súvetia utvorené pomocou nej sú citlivé na poradie podviet.

6.6. Spojka (vtedy), keď

Spojku *vtedy, keď* môžeme jednoducho definovať pomocou spojky *ak-tak*:

Definícia 11: α vtedy, keď $\beta =_{df} \beta \cup\rightarrow \alpha$.

Rozdiel oproti spojke *ak-tak* je len v poradí spájaných zložiek – ako za slovom *ak* nasleduje veta v úlohe podmienky (v pozícii antecedenta implikácie), tak aj za slovom *keď* vždy nasleduje podveta v úlohe podmienky (v pozícii druhej zložky v obrátenej implikácii). Na to, aby sa zachoval vetosled, definujeme nový symbol ($\leftarrow\cup$) pre obrátenú dostatočnú podmienku:

$\alpha \leftarrow\cup \beta =_{df} \beta \cup\rightarrow \alpha$.

6.7. Spojky ibaže[25]*,* ledaže*, resp.* vtedy, keď nie

Podľa lingvistov slovo *ibaže* je dubleta spojky *iba*, ktorá „vyjadruje odpor s obmedzením: z platnosti prvého člena syntagmy by vyplývala neplatnosť druhého člena, no aj ten platí“ (MSJ, 721). V stati *Len* (Gahér 2012b) sme vychádzali z toho, že dnes je dubletou spojky *iba* aj spojka *len* (angl. *only*)

[25] V anglickom jazyku jej zodpovedá slovo *unless*.

a navrhli sme explikáciu jej všeobecného významu nielen ako parataktickej spojky, ale aj jej významu v použití v spojení s nevetnými členmi.[26] Naše vysvetlenie významu slova *len* je v súlade s tými črtami, ktoré identifikovala lingvistika – *obmedzuje dačo na istý prípad, vydeľuje jeden prípad z väčšieho množstva*. Podľa našej rekonštrukcie významu spojky *len* (*iba*) ide o zložený operátor, ktorého konštrukcia zahŕňa všeobecnú kvantifikáciu a zápor (dvakrát), pričom tento prístup súvisel s cieľom vyjasniť si pojmy dostatočnej a nutnej podmienky. V prístupe, ktorý sme zvolili tu, je stratégia opačná: Pojmy dostatočnej a nutnej podmienky považujeme za základné (nedefinované) a na význame pojmu nutnej podmienky vyjadrenej spojením *len vtedy, keď* sa podieľa význam slova *len* nevyčleneným spôsobom.[27] Preto nemôžeme v tomto systéme predložiť jeho sémantickú analýzu.

Na druhej strane nám vyjadrenia dostatočnej a nutnej podmienky poskytujú komfortné nástroje na definovanie toho významu spojok *ibaže*, *ledaže*,[28] ktorý je identický s významom anglického *unless*. Spojku *ibaže* môžeme jednoducho definovať pomocou spojky (*vtedy), keď* a záporu *nie*:

Definícia 12: α ibaže $\beta =_{df} \alpha \leftarrow\cup \neg\beta$

Druhá zložka súvetia utvorená pomocou spojky *ibaže* je negatívnou dostatočnou podmienkou pre prvú zložku. Opisne to môžeme vyjadriť aj spojením *s výnimkou, že*.

Príklad:

(22) „Živoriť bude aj potom, ibaže uhol tvrdšej robote“ (http://slovniky.korpus.sk/?w=ibaže).

V prirodzenom jazyku už toto použitie spojky *ibaže* ustupuje[29] a nájdeme ho „zakonzervované“ v niektorých ustanoveniach právnych noriem, napríklad:

---

[26] S výnimkou napríklad vyjadrovania priania spojením *len ak*, *kiežby* (ang. *if only*).

[27] Ako sme už uviedli, podobne postupoval Lepore (2003, 89).

[28] V slovenčine sa spojky *ledaže, leda* považujú za zriedkavé (MSJ, 726), naproti tomu sú v češtine vo význame anglického *unless* používané často – počet ich výskytov v českom národnom korpuse je 698, resp. 1698 – pozri Křen a kol. (2010).

[29] Dnes prevažuje jej použitie v jednoduchom odporovacom význame, aký má spojka *avšak*, o ktorej sa už pred takmer šesťdesiatimi rokmi hovorilo, že je zastaraná (MSJ, 712). Napriek tomu sa stále používa.

(23) Uschovávateľ zodpovedá za škodu na uloženom listinnom cennom papieri, ibaže ju nemohol odvrátiť pri vynaložení odbornej starostlivosti. (§39, ods. 2, 566/2001 Z.z. Zákon o cenných papieroch)

Spojke *ibaže* zodpovedá aj spojenie *pokiaľ nie*, *ak nie*:

(22*) „Živoriť bude aj potom, pokiaľ neuhol tvrdšej robote“

Práve zápor, ktorý pri preklade spojky *unless* musí byť pripojený k slovesu, sa niekedy stráca zo zreteľa a namiesto správneho prekladu sa uvádza chybné spojenie *iba ak*. Táto chyba je, žiaľ, zakorenená aj v prekladových slovníkoch (Gahér 2003, 81).

6.8. Spojenie vtedy a len vtedy, keď

V logike, matematike, informatike a v príbuzných disciplínach sa v úlohe priraďovacej vetnej spojky používa spojenie *vtedy a len vtedy, keď*, ktoré identifikuje pravdivostnú funkciu (1,0,0,1). Jej dubletou je spojka *práve vtedy, keď*. Jazykovedci ich medzi vetnými spojkami neuvádzajú ani v súčasnosti[30] a identifikujú slovo *práve* ako vytyčovaciu zdôrazňovaciu časticu (MSJ, 786). Ich primárny význam pri použití v uvedených disciplínach je nečasový a môžeme ho definovať ako číre zlučovanie (konjunkciu) dostatočnej a nutnej podmienky, pričom v pozícii oboch podmienok je druhá podveta.

Definícia 13: α *vtedy a len vtedy, keď* $\beta =_{df} ((\alpha \leftarrow\cup \beta) \wedge (\alpha \rightarrow\cup \beta))$.

Príklad:

(24) Číslo je deliteľné šiestimi *vtedy a len vtedy, keď* je deliteľné dvomi a tromi.

Táto zložená spojka sa nazýva *ekvivalentor* a používa sa napríklad na vyjadrenie definičnej rovnosti, keď definovaný výraz vieme definovať len v zapojení do vetnej konštrukcie, takže obe strany definície sú vetami.

Často sa považuje za samozrejmé, že ekvivalentor vyjadruje symetrické spojenie, a preto by sme – zdá sa – nemali mať dôvod vyznačovať pozíciu

[30] Napr. v MSJ nie sú v zozname spojok a nenájdeme ich ani v slovenských slovníkoch – pozri http://slovniky.juls.savba.sk/.

podmienky, resp. podmienok. Tak tento operátor funguje v deduktívnych systémoch. To však – zdá sa – nie je v súlade tým, že pozícia podmienky je len na jednej strane spojenia. Táto otázka čaká na ďalšie vysvetlenie.

7. Záver

Pomocou rôznych vybraných typických formulácií či konštrukcií významu vetných operátorov, ktoré však vedú k tej istej pravdivostnej funkcii, sme sa pokúsili vysvetliť rôzne vlastnosti vetných operátorov, o ktorých hovoria jazykovedci. V nadväzujúcej stati by sme chceli tento základ využiť na ďalšie vysvetlenie rôznych „čítaní" vetných operátorov, o ktorých hovoria nielen jazykovedci, ale aj iní skúmatelia a používatelia jazyka. Chceli by sme podrobnejšie skúmať najmä to, ako interagujú vetné operátory jednak s vektorom diania, ktoré je opísané súvetím, jednak so smerovaním gramatického času zachyteného v súvetí, a napokon, ako toto všetko je koordinované s epistemickým cieľom použitia významu súvetia (predikciou, vysvetlením a pod.).

Literatúra

CAPPELEN, H. – LEPORE, E. (2005): *Insensitive Semantics. A Defense of Semantic Minimalism and Speech Act Pluralism*. Oxford: Blackwell Publishing.

CMOREJ, P. (2001): *Úvod do logickej syntaxe a sémantiky*. Bratislava: Iris.

ČERMÁK, F. (2008): Syntagmatika, kombinace a kumulace konjunkcí. In: Štícha, F. – Fried, M. (eds.): *Grammar and Corpus. Gramatika a korpus.* Praha: Academia.

DUŽÍ, M. – JESPERSEN, B. – MATERNA, P. (2010): *Procedural Semantics for Hyperintenzional Logic: Foundations and Applications of Transparent Intensional Logic*. Springer.

DVONČ, L. – HORÁK, G. – MIKO, F. – MISTRÍK, J. – ORAVEC, J. – RUŽIČKA, J. – URBANČOK, M. (1966): *Morfológia slovenského jazyka*. Bratislava: Vydavateľstvo SAV.

FREGE, G. (1879): *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle.

GAHÉR, F. (2000): Negation and Presuposition. In: *The Logica Yearbook* 2000. Praha: Filosofia, 133-150.

GAHÉR, F. (2001): O vetných spojkách: logika verzus lingvistika. *Jazykovedný časopis* 52, č. 1, 23-37.

GAHÉR, F. (2003): *Logika pre každého*. 4. dopl. vyd. Bratislava: Iris.

GAHÉR, F. (2006): *Stoická sémantika a logika z pohľadu intenzionálnej logiky*. 2. dopl. vyd. Bratislava: Univerzita Komenského.

Gahér, F. (2011): Sú pojmy dostatočná a nutná podmienka pre empirickú oblasť symetrické? *Organon F* 18, č. 3, 331-350.
Gahér, F. (2012a): Revízia definícií pojmov dostatočná a nutná podmienka. *Organon F* 19, č. 1, 16-37.
Gahér, F. (2012b): Len. *Organon F* 19, č. 3, 306-324.
Grice, P. (1989): S*tudies in the Ways of Words*. Cambridge (Mass.) – London: Harvard University Press.
Kesselová, J. (2007): Komunikačný potenciál operátorov na rozširovanie a rozvíjanie výpovede. In: Orgoňová, O. (ed.): *Jazyk a komunikácia v súvislostiach II*. Bratislava: Univerzita Komenského.
Kesselová, J. – Karpinský, P. – Imrichová, M. – Podhájecká, K. (2013): *Spojky a spájacie prostriedky v slovenčine*. Bratislava: Veda.
Kripke, S. (2011): Unrestrikted Exportation and Some Morals for the Philosophy of Language. In: Kripke, S.: *Philosophical Troubles*. Collected Papers, Vol. 1. Oxford: Oxford University Press, 322-350.
Křen, M. – Bartoň, T. – Cvrček, V. – Hnátková, M. – Jelínek, T. – Kocek, J. – Novotná, R. – Petkevič, V. – Procházka, P. – Schmiedtová, V. – Skoumalová, H. (2010): *SYN2010: žánrově vyvážený korpus psané češtiny*. Praha: Ústav Českého národního korpusu FF UK. Dostupný z: http://www.korpus.cz.
Lepore, E. (2003): *Meaning and Argument. An Introduction to Logic Through Language*. Oxford: Blackwell.
McCawley, J. D. (1993): *Everything that Linguists have Always Wanted to Know about Logic but were Ashamed to Ask*. Chicago: University of Chicago Press.
Neale, S. (1992): Paul Grice and the Philosophy of Language. *Linguistic and Philosophy* 15, 509-559.
Oddie, G. – Tichý, P. (1980): The Logic of Ability, Fredom and Responsibility. *Studia Logica* 41, 227-248.
Plato, J. (2009): Gentzen's Logic, In: Gabbay, Dov M. – Woods, J.: *Handbook of the History of Logic. Vol. 5. Logic from Russell to Church*. Elsevier.
Raclavský, J. (2012): Je Tichého logika logikou? (O vztahu logické analýzy a dedukce). *Filosofický časopis* 60, č. 2, 245-254.
Sokolová, M. – Žigo, P. (2014): *Verbálne kategórie aspekt a tempus v slovenčine*. Bratislava: Veda.
Svoboda, V. – Peregrin, J. (2009): *Od jazyka k logice. Filozofický úvod do moderní logiky*. Praha: Academia.
Stanley, J. (2007): *Language in Context. Selected Essays*. Oxford: Clarendon Press.
Strawson, P. F. (1950): On Referring. *Mind* 59, No. 325, 322-344.
Strawson, P. F. (1952): *Introduction to Logical Theory*. London: Methuen.
Strawson, P. F. (1992): O referencii. In: Oravcová, M. (ed.): *Filozofia prirodzeného jazyka*. Bratislava: Archa, 126-147.
Tichý, P. (1988): *The Foundations of Frege´s Logic*. Berlin – New York: de Gruyter.
Zouhar, M. (2011): *Význam v kontexte*. Bratislava: aleph.

ZOUHAR, M. (2013): Je možné brániť kontextualizmus? Odpoveď Lukášovi Likavčanovi, *Pro-Fil* 14, č. 1, 30-39. Dostupné online: http://www.phil.muni.cz/journals/index.php/profil/article/view/506.

ŽIGO, P. (2010): Kategória času. In: Dolník, J. (ed.) (2010): *Morfologické aspekty súčasnej slovenčiny*. Bratislava: Veda, 167-217.

Organon F 22 (3) 2015: 425-432

Mark Risjord: *Philosophy of Social Science: A Contemporary Introduction* Routledge, London, 2014, 288 pages

The past few centuries have witnessed an extraordinary boost in our capacity to explore our world and to amass knowledge of it. We have established an apparently efficient method of exploration in which we seek for deterministic causal laws that govern all happenings in our world, so that they make it possible for us to predict future happenings. And although sometimes, such as when researching living organisms, the laws appear less clear, nevertheless everything seems to be ultimately underlain by the wonderfully simple laws of physics.

This procedure, however, falters when our aim is to explore our own, human communities. Here, no deterministic laws appear to be in view; and skeptics would say that the only thing we have managed to acquire is the *illusion* that we have something as knowledge. Why do human communities so stubbornly resist our efforts to extract underlying laws from them?

One answer which would seem to be forthcoming, is that our communities are simply too complex for us to get a hold on them. After all, it is not only the communities that are so impenetrable for us, but also other extremely complex systems. Consider, for example, the weather: we still seem unable to arrive at very reliable forecasts, and our excuse is thought to be the fact that the weather is so multidimensionally complex. Maybe in the future we will develop methods of mastering more complex systems and then we will have the ability to predict weather more reliably. And maybe, by the same token, one day we will be also able to determine the laws of human communities.

But it is far from clear that it is only an issue of complexity differentiating the exploration of nature from that of human communities. One traditional view is that what makes a more substantial difference is that what is in play are two utterly different kinds of *understanding* – and that understanding human communities is not a matter of determining causal laws. Thus, philosophers from Dilthey to Gadamer speak about searching for *sense* – about *hermeneutics*. However, the concepts of hermeneutics generally remain somewhat esoteric, which usually prevents them from offering much guidance to social *scientists*.

On the one hand, then, we have philosophers who claim that understanding human communities must, at base, be akin to understanding nature, and

assume that if as yet we have not acquired any reliable laws, then it is only because our social sciences are still immature. (The hope is often that this will change once we master the interconnections of human neurology and human behavior.) On the other hand, we also have philosophers who think that it is futile to try to understand human communities in the way we have come to understand nature; and that trying to look for deterministic laws governing social events is nonsensical.

All this has spawned the situation whereby the foundations of social sciences are surrounded by sufficient philosophical problems to warrant a wholly specific philosophical discipline: the *philosophy of social sciences*. While what is usually understood under the traditional heading *philosophy of science* is the philosophy and methodology of *natural* sciences, this new discipline concentrates on the philosophical and methodological problems of specifically social sciences and humanities.

Risjord's introduction into this new philosophical discipline is a very well written book, surveying the multifarious specifica of investigating human communities and humans as its members. This should be particularly appreciated in view of the fact that the discipline is still very much in its infancy and has, as yet, no standardly accepted structuring of its specific topics. (Indeed, as far as I can see, there is no general agreement on what it is to comprise.)

The first topic Risjord discusses in his book is the question of objectivity in social sciences. The point is that while in natural sciences there is usually no problem in assuming the standpoint of a detached observer, this is less easy in social sciences; and here there are voices that one of the features distinguishing social sciences (and humanities) is that they are not able to clearly separate facts from values. If this is the case, then social sciences, it would seem, cannot be objective in the same sense in which natural sciences strive for objectivity – we cannot just tell stories about what there is, without slipping into talking about what there *should be*.

Risjord, it must be said, evades expressing a clear view on this matter. In the first part of the chapter he sees the situation in the eyes of those who would want social sciences to come as close to natural sciences as possible and discusses some obstacles to this; while in the second part of the chapter he switches to the view of those who think that striving for the natural scientific kind of objectivity in social sciences is futile – that achieving it is both impossible and pointless. Personally I would like to hear more about the confrontation of these two views.

In the next two chapters Risjord addresses the quarrel of naturalism (i.e. the conviction that social sciences do not differ substantially from natural

ones) vs. interpretivism, and then especially the tenets of interpretivism. Exponents of this view maintain that explaining a human society is not a matter of finding causes of what happens in the society, but rather of finding reasons for why the members of the society do what they do. And as reasons are beliefs, which in turn consist of concepts, this amounts to finding the beliefs of the people and understanding the concepts with which they operate. Thus, the main business of interpretativism is to *make sense* of the community under study, especially by aligning *their* concepts with *our* ones – in order to be able to understand *their* reasons in terms of *our* ones and thus to understand why they do what they do.

Next Risjord turns his attention to the questions of agents and agency. This, in one view, is closely connected with interpretivism, for it seems that there is a crucial difference between studying us humans and studying anything else in our world – humans can be seen not only as organisms behaving in certain ways, but also as actors carrying out actions, and to understand the latter aspect requires us to *make sense* of them. Also it may lead us to the game-theoretical models of human intercourse, which have become so popular in some social sciences.

The following chapter discusses the possibilities of reducing the social to something simpler, typically to the individual. Of course, such a reduction might render the specifically social sciences superfluous; but as the book clearly shows, there are so many specific problems related to the social level that even if someone believes that this level can be reduced, "in principle", to some underlying levels, the problems of the specifically social level would still remain relevant and unresolved.

Then Risjord gets to what I would take as the most distinctive feature of the social – i.e. norms and rules. (In my personal view, the whole level of the social may be seen as grounded in our human capacity to assume normative attitudes.) Risjord pays special attention to the discussion between the so called *normativists* (those who believe that normativity is a *sui generis* phenomenon that must be explained as such) and anti-normativists (who want to reduce normativity to non-normative phenomena).

The next chapter is devoted to collective intensionality and related phenomena which seem to be emergent only on the collective level. Here, too, he pays attention to game-theoretic models.

In the final two chapters of the book, Risjord moves on to discuss problems related to causality within social sciences. It is clear that although it might not be possible to assimilate the whole of social sciences to the search for causal laws as in natural sciences, it would be preposterous to conclude that looking

for causal connections is not also of importance for a social scientist. Risjord points out that social scientists often profit from building causal models; and he further discusses the general question of the existence of causal laws governing human societies.

Many of the problems discussed in Risjord's book are quite complex; hence it is not the kind of introductory book that is easily accessible to complete outsiders. However, for people with some grounding in philosophy and ways to account for human societies, the book constitutes a nice compendium of philosophical problems specific to the social sciences, to the understanding of human communities and to the understanding of us humans as members of such communities.

Jaroslav Peregrin
jarda@peregrin.cz

Barry Dainton – Howard Robinson (eds.): *The Bloomsbury Companion to Analytic Philosophy*
Bloomsbury Publishing Plc., London, 2014, 672 pages

"In this Companion we provide a guide to analytic philosophy's past, present, and future; we also attempt to specify what – if anything – is genuinly distinctive about it" (p. xi).

With these words of editors Barry Dainton (University of Liverpool) and Howard Robinson (Central European University) starts *The Bloomsbury Companion to Analytic Philosophy*. After passing a short Introduction and Preface, the book continues with three parts dedicated to the past, present and the future of the analytic philosophy.

Part I: History, Methods, and Problems. The main topic of the first part is the history, or we can say the stories of the most well-known figures from the analytic philosophy. This part is written by the editors Dainton and Robinson and begins with a description of changing opinions about the world in the middle of the 19th century (the story of Francis Bradley and his holistic view of the world). After this short introduction, there appear George Edward Moore, Bertrand Russell and Gottlob Frege. The author of these chapters, Barry Dainton, focuses on the famous problems and possible solutions to them (proposed by each of the authors) connected with the beginning of the analytic philoso-

phy. The problems include, for example, the problem of the nature of some entities (numbers), the class paradox, reference, propositions, etc.

The story of the history of analytic philosophy continues with description of another famous figures and problems from the past. To mention just a few, the chapter concerns Vienna Circle, Ludwig Wittgenstein (the *Tractatus* phase and, of course, later Wittgenstein, each of these described in separate chapters), Willard van Orman Quine, Donald Davidson, Saul Kripke, Hilary Putnam and so on. We can see the lain of the emerging problems and attempts to solve them – the problem of meaning, method, descriptions, truth conditions, etc.

In general, we can say that the first part provides a detailed and complex outline of the most fundamental ideas and problems discussed in the beginning of the analytic philosophy; many of them, however, remain still topical in the present discussions.

Part II: Current Research and Issues. As the title suggests, the second part deals with the actual themes and problems connected to the cotemporary analytic philosophy. The crucial attribute of this chapter is that each part of it is written by different authors, specialists in a particular field. So we can read the contribution about the philosophy of mathematics and logic (Mary Lang), the philosophy and language (Barry C. Smith), meaning and normativity (Richard Gaskin), the philosophy of science (James Ladyman), metaphysics (E.J. Lowe), knowledge (Bryan Frances and Allan Hazlett), causation (Helen Beebee) and many others.

Since it would be very difficult to discuss all of these parts I will rather deal with two of them in some detail – the one concerning the philosophy of language and the one concerning the philosophy of science.

Let discuss the philosophy of language chapter first, written by Barry C. Smith, professor of philosophy and director of the Institute of Philosophy at the School of Advanced Study, University of London.

At the very outset of his contribution, Smith introduces this field and sketches the main topics. He puts it as follows:

> Philosophers of language in the analytic tradition have mainly focused their attention on two central concerns: the ability of language to express and communicate our thoughts; and the relation of language to reality. Broadly speaking, both issues bear on the language's representational powers: its ability to encode thought and portray aspects of reality. (p. 201)

According to Smith it is also very "important to distinguish language from communication. Communication can be nonverbal" (p. 202). The crucial issues include the meaning of words, identification of words with noises or written tags and, of course, the way of connecting words to the sentences. Various approaches to these problems are illustrated by the works of Willard van Orman Quine, who "think[s] of language as sets of well-formed word-strings, where words are conventional sound-meaning pairs and, well-formedness is licensed by the grammar rules for the language in question" (p. 204); and Noam Chomsky who criticised Quine's conception because, according to him, sentences are not only word strings and their structure is not linear. This debate between Quine and Chomsky serves as an introduction to the more interesting topic, namely the problem of demarcation between semantics and pragmatics.

Smith introduces two main approaches to this problem. On the one hand, we have Paul Grice and his theory stressing that "people always mean what is said by uttering the sentence – that is, it is always part of what they assert – but they may mean or assert more besides: (8) I'm tired ... For Grice, I have asserted that I am tired. That is, what is said by uttering (8), and what is meant over and above what is said depends on a conversational implicature that the hearer must infer from what was said and the background information" (p. 215). On the other hand, we have Francois Recanati and his attitude that "what is said is not identical to what is determined by the meaning of words in a sentences and its linguistic form, but rather is something pragmatically determined by what is said by the utterance in context" (p. 221). Concerning this distinction, Smith writes:

> The distinction between semantics and pragmatics is the subject of a large controversy, and it should be said at the outset that perhaps there is no fully satisfactory unifying account of the relations between uttering and meaning, nor any single theory that can take care of all examples. ... Extremes include uttering without meaning anything at all (babble), and meaning without uttering; for example, when you did not utter a sound you may have meant something by your silence. (p. 214)

The second part of this chapter deals with one of the most respectable branches of philosophy these days: the philosophy of science (written by James Ladyman, the University of Bristol).

As in the previous case, the author firstly focuses on the historical introduction and then discusses four important parts of the philosophy of science:

scientific methodology, the metaphysics of science, the epistemology of science and the philosophy of science. Importantly or not, Ladyman does not touch the philosophy of social sciences.

In the first part of this entry, Ladyman discusses the different approaches to scientific method. On one hand, there is, historically speaking, Bacon's inductive method. On the other hand, we have Popper and falsificationism. These different approaches to the scientific method are nailed down by Kuhn and his rejection of the (one) typical scientific method. As the last example of different approaches to the discussion about scientific method, the author mentions Nancy Cartwright who "argues that science is an overlapping patchwork of models without the kind of hierarchy supposed by reductionists and physicalists, and without even consistency between different parts of the whole" (p. 262).

Beside the problem of induction, this section aims at related problems, such as underdetermination of the theory by data, the problem of probability and its relation to the actual events in the world or the problem of theoretical terms and their reference to the objects in the world.

The second bigger part of this chapter (*The Metaphysics in Science*) discusses mainly the problem of laws. It is stated that "[r]ecent discussions of laws of nature has focused on the metaphysics of laws. What is a law of nature and do they differ from generalizations that happen to be universally true, like no gold sphere is bigger than earth?" (p. 265). The problems of laws are also discussed from another point of view; it can be said that some alternatives are provided: "new approaches" to the same questions. For example, we can mention David Lewis' approach: "Laws are the theorems and axioms of deductive systems that achieve the best combination of simplicity and strength" (p. 265). Unfortunately there is a problem with clarification of the term "the best combination of simplicity and strength".

Another approach is that of Bas van Fraassen and Nancy Cartwright who are "sceptical about the traditional views of the importance of fundamental laws in the analysis of scientific theories. Cartwright and John Dupre are well-known critics of the idea that science is unified. Cartwright argues that the relationship between theories, models, and reality undermines what she calls 'fundamentalism' about laws. ... Such a view is often defended in philosophy of biology where it frequently seems as if a single scientific term such a species or gene is understood differently in different parts of the science, and for which pluralism about meaning and reference is proposed" (pp. 266-267).

The part entitled *The Epistemology of Science* is mainly dedicated to the debate between the realists and antirealists. The core of this part is focused on

the arguments in favour or against realistic approach. The author describes the famous argument against realism which is dealing with the success of science and the reference of the terms.

Part III: New Directions in Analytic Philosophy. Last section of this book deals with the actual challenges in analytic philosophy. According to the authors (Dainton, Robinson), the situation in analytic philosophy is not so bad as many people might think. There are still many different problems and challenges that need to be resolved and that can be considered as interesting in many different areas such as mathematics, logic, cognitive sciences etc.

In the end, we would like to mention a few sentences as closing remarks. We can say that this book gives a really good introduction to the history and main problems of the analytic philosophy. The fact that each field is described and analysed by a specialist makes this book really helpful and clarifying for students of philosophy. This book also includes a chronology of key events from the history of analytic philosophy and the dictionary of the important terms and concepts.

Ivana Klimová
klimova.ivana@gmail.com

Organon F 22 (3) 2015: 433-434

PhiLang 2015 – The Fourth International Conference on Philosophy of Language and Linguistics (University of Łódź, Poland, 14-16 May 2015)

PhilLang – the International Conference on Philosophy of Language and Linguistic – belongs among those few conferences which focus on overlaps between philosophy and linguistics, including overlaps between philosophy and philosophy of mind, linguistics, literary theory, semantic, and pragmatic theories and metaphilosophy. *PhiLang 2015*, the conference's fourth continuant, was (again) organized in Łódź and, as usually, brought together philosophers, logicians and linguists from all around the world.

The conference commenced with two (out of five) plenary lectures, namely Manuel-García Carpintero's (Universitat de Barcelona) and Joanna Odrowąż-Sypniewska's (University of Warsaw) entitled *Predicativism and the Presuppositional View of Proper Names* and *Context, Vagueness, and Reference*, respectively. The conference then split into three parallel sessions dedicated primarily to questions on the border between philosophy, language and linguistics. To list just a few (namely those I had a chance to attend), Gabrielle Mras (University of Vienna) discussed *The Sense of Frege's Reference*, André Bazzoni (University of California, Berkeley) proposed *The Cluster-Occurrence Theory of Proper Names*, Matthew Cameron (University of St-Andrews) discussed *Speaker's Intentions and the Formal Representation of Context* and Mark Pinder (University of Reading) posed a question: *Are Folk Intuitions Relevant to Arguments from Reference?* One of the afternoon sessions started with *Re-reading Kripke's Normativity Argument* co-authored by Krzysztof Posłajko and Jacek Wawer (Jagiellonian University Cracow) and continued with Dan Zeman's (University of the Basque Country) *Relativism and the Multi-Perspectivality of Predicates of Personal Taste* and Luis Fernández Moreno's (Universidad Complutense de Madrid), asking: *Is the Semantics of Natural Kind Terms Extendable to Artificial Terms?*.

The third plenary lecture delivered by Richard Gaskin (University of Liverpool) and entitled *Reference and Linguistic Idealism* started the second day of the conference. Again, after the lecture and a vivid discussion the conference split into three parallel sessions featuring, for example, Peter Ridley's (King's College London) *Who's Mum*, Nathan Duckett's (University of Manchester) *Allegedly Isn't an Epistemic Modal*, Natalia Karczewska's (University of Warsaw)

Challenges to Metalinguistic Negotiation as Disagreement or Ashley Atkins's (Simon Fraser University) *Modality as a Window into Cognition*. The fourth plenary lecture given by Wolfram Hinzen (ICREA/Universitat de Barcelona) concerned *The Grammar of Essential Indexicality* and was followed by *Indicative Conditionals, Probabilistic Relevance and Discourse Structure* by Arno Goebel (University of Konstanz), *Does Fiction Make Sense? Understanding Fiction* by Crister Nyberg (University of Helsinki) and Lukáš Bielik's (Slovak Academy of Sciences) *Thought Experiments in Semantics: An (Apparent) Puzzle*.

The last day started with the last plenary lecture, namely Marián Zouhar's (Slovak Academy of Sciences) *Against Descriptivism: On an Essential Difference between Proper Names and Definite Descriptions* followed by the last three parallel sections. In them, Heimir Geirsson (Iowa State University) considered *Empty Names and Error Theory*, Luca Sbordone (University of Cambridge) accounted for *Vagueness, Contingency and Assessment-Sensitivity*, Martin Vacek (Slovak Academy of Sciences) discussed *Alien Properties and Impossible Worlds* and Halina Święczkowska, together with Beata Piecychna (both from the University of Białystok), provided some *Reflections on Some of the Issues of Rationalist Philosophy of Mind and Philosophy of Language. Remarks on the Margins of a Philosophical Discourse Concerning Speech by Gerauld de Cordemoy*.

In words of the *PhiLang 2015* organizer, Piotr Stalmaszczyk, any conference is as good as its participants are. And although I have not listed every speaker that presented a paper, it is more that clear that the conference has again proven that it belongs among the most influential philosophical conferences in Europe. It has also showed that philosophy has a lot to say on the issues concerning language and linguistic and does thus contribute to the actual as well as traditional debates. All this being said, it is not a surprise that organizers plan yet another *PhiLang – PhiLang 2017*.

Martin Vacek
martinvacekphilosophy@gmail.com