Contents

ARTICLES

DISCUSSIONS

BOOK REVIEWS

REPORTS

Organon F 24 (1) 2017: 2-9

The Role of Disjunction in Some Alleged Non-Monotonic Inferences[1]

MIGUEL LÓPEZ-ASTORGA

ABSTRACT: Lukowski has argued that, if it is the case that there are actual non-monotonic inferences, they are very hard to find. In this paper, a representative kind of inference that is often considered to be non-monotonic is addressed. Likewise, certain arguments provided by Lukowski to demonstrate that that type of inference is not really non-monotonic are reviewed too. Finally, I propose an explanation of why, despite the fact that the arguments given by him seem to be convincing, it is usually thought that those inferences are not monotonic. In this way, I also try to account for the role that disjunction has in this issue and argue in favor of the idea that we can continue to suppose that the human mind does not ignore the essential requirements of classical logic.

KEYWORDS: Disjunction – inference – logic – monotonicity – non-monotonicity.

1. Introduction

Each theory claiming that the human inferential activity is logical must face the problem of the non-monotonic inferences. The motive of that is that classical logic is monotonic and it appears that, to solve the difficulties of those inferences, it is necessary either assuming the thesis that human

[1] Received: 2 September 2016 / Accepted: 15 December 2016

✉ Miguel López-Astorga
Institute of Humanistic Studies "Juan Ignacio Molina", University of Talca
Av. Lircay s/n, Talca, Chile
e-mail: milopez@utalca.cl

reasoning is not logical or looking for a logic other than the classical one, which can be thought to be non-monotonic. This is considered to be a fact and, from perspectives holding that the human mind does not work resorting to logical forms, it is often said that this problem is crucial for the theories stating that reasoning is about formal rules (see, e.g., Johnson-Laird, Khemlani, & Goodwin 2015, 201-202).

However, Lukowski's (2013) paper allows thinking about another possibility. According to him, while it cannot be maintained for sure that there are not real non-monotonic inferences, it is actually difficult to find an instance of inference that is so without a doubt. The truth is that his arguments seem to be absolutely convincing. Therefore, a justified question in this regard can be: if it is so hard to find non-monotonic inferences, why is the contrary a generally assumed idea?

To answer that question is the main aim of this paper. To do that, I will review an emblematic case of allegedly non-monotonic inferences and some arguments given by Lukowski (2013) in order to prove that they are not really non-monotonic. That kind is the one of the inferences related to "a reasoning increasing preciseness" (Lukowski, 2013, 67). Then I will show that the cognitive science literature provides results that enable us to understand why, in spite of his arguments, people tend to think that such inferences do be non-monotonic. Likewise, I will describe the relevance that disjunction and the logical rule of disjunction introduction (from now on, DI) have in this way and offer some commentaries supporting both Lukowski's (2013) theses and the idea that classical logic can be important in the human mind.

2. Monotonicity versus non-monotonicity

As it is well known, classical logic is, as mentioned, monotonic. As remembered by Lukowski (2013, 63-65), this basically means that, in that logic, if it is correct that $\{A\} \vdash \{B\}$, then it is also correct that $\{A \cup \Gamma\} \vdash \{B\}$. Thus, non-monotonicity is just the opposite. It refers to the situation in which, while that $\{A\} \vdash \{B\}$ is correct, that $\{A \cup \Gamma\} \nvdash \{B\}$ is so as well.

Clearly, as mentioned, this is a problem for any approach arguing that human reasoning is related to classical logic, since, if that idea were right,

all of our inferences would have to be monotonic. However, the point is that the non-monotonic inferences seem to be frequent. Lukowski (2013) reviews several examples in this regard, but one of them can be representative enough. That example is, as said, the one referring to the reasoning increasing preciseness.

Following Lukowski (2013, 65-67), the reasoning increasing preciseness is very usual in medical diagnosis contexts. Generally, in those contexts, the tests provide a number of results $\{\alpha_1, \ldots, \alpha_n\}$ that are linked to a number of possible conditions $\{\beta_1, \ldots, \beta_n\}$. Thus, if we based on his essential ideas without considering necessarily the order in which he presents his arguments, it can be said that the link can be understood as a deduction relationship, and that, if we assume these definitions:

$$A = \{\alpha_1 \wedge \ldots \wedge \alpha_n\}$$
$$B = \{\beta_1 \vee \ldots \vee \beta_n\}$$

it can also be stated that $\{A\} \vdash \{B\}$.

Nevertheless, the physicians often continue to carry out tests and hence obtain more results. Thus, $\{A\}$ can be transformed into a set $\{A'\}$ to which this identity corresponds:

$$A' = \{\alpha_1 \wedge \ldots \wedge \alpha_n \wedge \alpha_{n+1}\}$$

In this situation, it is absolutely possible that A' provides further information and that the new data lead the physicians to a new set of conditions $\{B'\}$ that can be defined as follows:

$$\{B'\} = \{\beta_2 \vee \ldots \vee \beta_n\}$$

As it can be noted, what has happened is that the new datum $\{\alpha_{n+1}\}$ has removed a possible condition $\{\beta_1\}$, and this can be considered to be a clear example of non-monotonic reasoning. The motive is obvious: while $\{A\} \vdash \{B\}$, $\{A \cup \alpha_{n+1}\} \nvdash \{B\}$. $\{A \cup \alpha_{n+1}\}$ is identical to A', and what can be deduced from it is not $\{B\}$, but $\{B'\}$, that is, $\{A'\} \nvdash \{B\}$, but $\{A'\} \vdash \{B'\}$.

In Lukowski's view, this is a really important process in medical contexts because it gives more preciseness progressively. Nonetheless, it is not clear that it describes a non-monotonic inference. His argument is straightforward, too (cf. Lukowski 2013, 67):

If $\{A'\} \vdash \{B'\}$, then $\{A'\} \vdash \{B\}$ too, as, in classical propositional calculus, $\{B'\} \vdash \{B\}$. In other terms, if $\{\alpha_1 \wedge \ldots \wedge \alpha_n \wedge \alpha_{n+1}\} \vdash \{\beta_2 \vee \ldots \vee \beta_n\}$, then $\{\alpha_1 \wedge \ldots \wedge \alpha_n \wedge \alpha_{n+1}\} \vdash \{\beta_1 \vee \ldots \vee \beta_n\}$ too, as, in classical propositional calculus, $\{\beta_2 \vee \ldots \vee \beta_n\} \vdash \{\beta_1 \vee \ldots \vee \beta_n\}$.

Therefore, the question is: if all of this is so evident, why this kind of inference is often considered to be non-monotonic? In my view, the key is in DI. I try to explain this in the next section with the help of the results about this last rule that are to be found in the cognitive science literature.

3. The problems of DI in human reasoning

Indeed, DI is the rule that is necessary to derive $\{\beta_1 \vee \ldots \vee \beta_n\}$ from $\{\beta_2 \vee \ldots \vee \beta_n\}$. As it is well known, DI is a rule that can be assumed as basic in a logic based on Gentzen's (1935) natural deduction calculus and formally expressed in this way:

$$\frac{p}{(\text{Ergo})\ p \vee q}$$

However, it is a controversial rule as well. The literature informs that people do not always tend to use it in a natural way. In fact, most of the time most of the people do not apply this rule (see, e.g., Orenes & Johnson-Laird 2012), and several current psychological theories about reasoning have explanations for this phenomenon. The case of the mental logic theory (e.g., Braine & O'Brien 1998a; O'Brien 2009; 2014; O'Brien & Li 2013; O'Brien & Manfrinati 2010) is especially relevant here, since it is one of the theories that continue to claim that human reasoning is based on logic nowadays. Nevertheless, this theory is empirical and does not consider all of the formal rules of classical logic to be essential schemata of the human thought, but only the rules of this last logic that are clearly used by people. This is important because, given that, as mentioned, the results reported in the literature show that individuals do not usually apply DI, the mental logic theory does not accept it as a basic rule, which means that it cannot be expected that people habitually use it.

Of course, if we assume the mental logic theory, this is an explanation of why the kind of inferences considered in the previous section is generally thought to be non-monotonic. If, to be aware that those inferences are actually monotonic, it is necessary to apply DI and people tend not to accept that rule, most of the individuals may not note the real logical nature of them and consider them to be inferences in which, when a new premise is added, what can be drawn is not exactly the same.

True, this option solves a problem. Nonetheless, it raises another one. If people do not often apply DI and that is a very important rule in classical propositional calculus, it is doubtful that the human mind works in accordance with that calculus. But solving this second problem is relatively easy. On the one hand, the fact that people generally use logic in their inferences does not mean that all of the inferences that can be made have the same difficulty level. Obviously, it can be assumed that some inferences and rules are harder than others. In addition, the mental logic theory also has the necessary machinery to respond to an objection such as this one. As explained, the theory proposes that there are a number of schemata that are not difficult and that, in all probability, people use whenever they have the opportunity. However, it is also possible to speak about sophisticated individuals that are able to make more complex inferences (see, e.g., Braine & O'Brien 1998b, 223). So, the possibility exists that certain individuals, who, for any reason, make logical inferences more easily than other people, use DI without difficulties. Thus, it can be said that the fact that we reason resorting to logical rules does not mean that all of us do that in the same way.

On the other hand, the proponents of the mental logic theory also claim that affirming the existence of a logic in the human mind does not necessarily imply stating that the only factor that plays an important role in the human intellectual activity is that logic (O'Brien 1998, 36-37). Thus, this very theory proposes, in the same way, that pragmatics is essential in reasoning too (Braine & O'Brien 1998d, 46ff) and that the mental logic is not absolutely incompatible with non-logical processes in the human mind (O'Brien 1998, 38). So, based on arguments of this kind, it can be said that ideas such as that the abductive inferences (that is, a kind of inference that is not admitted by classical logic) are used in medical diagnosis contexts as well (e.g., Pukancová & Homola 2015) are not a problem for the argumentation above either, since it can be thought that,

while it is obvious that the human cognitive architecture includes certain clearly logical schemata, it enables to resort, in some cases, to other mechanisms to obtain conclusions too. From this perspective, the difficulties related to the fact that people do not always apply any specific logic rule in a particular circumstance become relative, as the reasons for that fact can be many.

4. Conclusions

Lukowski (2013) also reviews other cases of alleged non-monotonic inferences. But, as far as I understand his arguments, his main idea is that most of them refer to monotonic inferences in which the conclusion does not change really due to addition of another premise, but because either a premise is changed by another one, which transforms the inference in other different inference, or the initial inference is not correct and the second one shows that. An example of this last case given by him is that of the "Tweety the ostrich" (Lukowski 2013, 60-70). It presents the situation in which, in principle, given that it is said that "Tweety is a bird" – {a}, it is drawn from it that "Tweety can fly" – {b}, since it is assumed that $\{a\} \vdash \{b\}$ (cf. Lukowski 2013, 69) Nevertheless, a problem can appear if, after that, it is stated that "Tweety is an ostrich" – {c}, as we would have to accept that, while $\{a\} \vdash \{b\}$, $\{a, c\} \nvdash \{b\}$ (cf. Lukowski, 2013, 69). The reason is evident: although Tweety is a bird, it is an ostrich too, and, as it is known, ostriches cannot fly. However, the Lukowski's explanation of why this is not a real non-monotonic inference is also clear. The key is just that a mistake has been made: "the error of generality" (Lukowski 2013, 70). Thus, as I interpret Lukowski's arguments, what really happens here is that it is not true that $\{a\} \vdash \{b\}$, since cases of $\{a \wedge \neg b\}$ are possible. In fact, the scenario in which Tweety is an ostrich is one in which we have $\{a \wedge \neg b \wedge c\}$. So, the inference is not actually non-monotonic, because the suitable deduction relationships are not $\{a\} \vdash \{b\}$ and $\{a, c\} \nvdash \{b\}$, but $\{a\} \nvdash \{b\}$ and $\{a, c\} \nvdash \{b\}$. Therefore, it seems that, from his point of view, accounting for the non-monotonic reasoning is really accounting for a type of very exceptional reasoning that is not common (or, if preferred, that is difficult to find).

The consequence of this is evident: the idea that a logical system based to a greater or lesser extent on frameworks similar to that of Gentzen

(1935) continues to be valid. However, this does not imply that there are not certain challenges to face. This paper has shown which two of those challenges can be. On the one hand, the cognitive theories holding that the human mind follows logical schemata must clarify when and under what circumstances the particular rules of inference are used, and when and under what circumstances they are not. We already know that DI is hard but not why. Likewise, it would be worth being absolutely sure about the difficulty of other schemata that have not yet been extensively studied and the reasons of it. On the other hand, it is also necessary to explain what being a sophisticated individual exactly means, the characteristics that are needed to be so, and, maybe, the variables that can have an influence on the fact that people become sophisticated as well. Likewise, given that not only logic takes action in our thought, it would be also desirable to clarify what the other factors or types of processes are actually and the particular circumstances under which those factors or processes can be used.

If works such as those indicated above are reviewed, there is no doubt that the mental logic theory, although, as said, it does not accept all of the schemata valid in classical logic, has made a significant progress in regard to the first challenge. It proposes even a reasoning program indicating the order in which the main schemata are usually applied (see Braine & O'Brien 1998c, 82-83, Table 6.2). Of course, further research is needed in this way, but it can be stated that there are already important conclusions obtained. In connection with the concept of logical sophistication, as far as I know, the situation is not the same. So the research on it is, to some extent, more urgent. Finally, although it is true that there are many studies about the non-logical machinery that the human mind can have (some examples have been cited in this paper), perhaps it would be interesting to continue to explore the exact situations in which logic has to be left and only other types of inferences can be made.

References

BRAINE, M. D. S. & O'BRIEN, D. P. (eds.) (1998a): *Mental Logic*. Mahwah: Lawrence Erlbaum Associates, Inc., Publishers.

BRAINE, M. D. S. & O'BRIEN, D. P. (1998b): A Theory of if: A Lexical Entry, Reasoning Program, and Pragmatic Principles. In: Braine, M. D. S. & O'Brien, D. P. (eds.): *Mental Logic*. Mahwah: Lawrence Erlbaum Associates, Inc., Publishers, 199-244.

BRAINE, M. D. S. & O'BRIEN, D. P. (1998c): The Theory of Mental-Propositional Logic: Description and Illustration. In: Braine, M. D. S. & O'Brien, D. P. (eds.): *Mental Logic*. Mahwah: Lawrence Erlbaum Associates, Inc., Publishers, 79-89.

BRAINE, M. D. S. & O'BRIEN, D. P. (1998d): How to Investigate Mental Logic and the Syntax of Thought. In: Braine, M. D. S. & O'Brien, D. P. (eds.): *Mental Logic*. Mahwah: Lawrence Erlbaum Associates, Inc., Publishers, 45-61.

GENTZEN, G. (1935): Untersuchungen über das logische Schließen I. *Mathematische Zeitschrift* 39, 176-210.

JOHNSON-LAIRD, P. N., KHEMLANI, S. & GOODWIN, G. P. (2015): Logic, Probability, and Human Reasoning. *Trends in Cognitive Sciences* 19, No. 4, 201-214.

LUKOWSKI, P. (2013): Is Human Reasoning Really Non-Monotonic? *Logic and Logical Philosophy* 22, 63-73.

O'BRIEN, D. P. (1998): Mental Logic and Irrationality: We Can Put a Man on the Moon So Why Can´t We Solve Those Logical Reasoning Problems? In: Braine, M. D. S. & O'Brien, D. P. (eds.): *Mental Logic*. Mahwah: Lawrence Erlbaum Associates, Inc., Publishers, 23-43.

O'BRIEN, D. P. (2009): Human Reasoning Includes a Mental Logic. *Behavioral and Brain Sciences* 32, 96-97.

O'BRIEN, D. P. (2014): Conditionals and Disjunctions in Mental-Logic Theory: A Response to Liu and Chou (2012) and to López-Astorga (2013). *Universum* 29, No. 2, 221-235.

O'BRIEN, D. P. & LI, S. (2013): Mental Logic Theory: A Paradigmatic Case of Empirical Research on the Language of Thought and Inferential Role Semantics. *Journal of Foreign Languages* 36, No. 6, 27-41.

O'BRIEN, D. P. & MANFRINATI, A. (2010): The Mental Logic Theory of Conditional Proposition. In: Oaksford, M. & Chater, N. (eds.): *Cognition and Conditionals: Probability and Logic in Human Thinking*. Oxford: Oxford University Press, 39-54.

ORENES, I. & JOHNSON-LAIRD, P. N. (2012): Logic, Models, and Paradoxical Inferences. *Mind & Language* 27, No. 4, 357-377.

PUKANCOVÁ, J. & HOMOLA, M. (2015): Abductive Reasoning with Description Logics: Use Case in Medical Diagnosis. In: *Proceedings of the 28th International Workshop on Description Logics.* Athens.

Organon F 24 (1) 2017: 10-36

A Valid Rule of β-conversion for the Logic of Partial Functions[1]

MARIE DUŽÍ – MILOŠ KOSTEREC

ABSTRACT: The goal of this paper is to examine the conditions of validity for the rule of β-conversion in TIL, which is a hyperintensional, typed λ-calculus of partial functions. The rule of β-reduction is a fundamental computational rule of the λ-calculi and functional programming languages. However, it is a well-known fact that the specification of this rule is ambiguous (see, e.g., Plotkin 1975 or Chang & Felleisen 2012). There are two procedurally non-equivalent ways of executing the rule, namely β-conversion '*by name*' and β-conversion '*by value*'. In the λ-calculi conversion by name is usually applied, though it is known that such a conversion is not unconditionally valid when partial functions are involved. If a procedure that is typed to produce an argument value is improper by failing to produce one, conversion by name cannot be validly applied. On the other hand, conversion by value *is* valid even in the case of improperness. Moreover, we show that in a typed λ-calculus the specification of λ-closure is also not unambiguous. There is an interpretation of this specification under which β-reduction by name is not valid even when the argument procedure does *not* fail to produce a value. As a result, we present a universally valid rule of β-reduction *by value*.

[1] Received: 3 October 2016 / Accepted: 19 December 2016

✉ Marie Duží
VSB-Technical University of Ostrava, Department of Computer Science FEI
17. listopadu 15, 708 33 Ostrava, Czech Republic
e-mail: marie.duzi@vsb.cz

✉ Miloš Kosterec
Department of Logic and Methodology of Sciences, Comenius University in Bratislava
Gondova 2, 814 99 Bratislava, Slovak Republic
e-mail: milos.kosterec@gmail.com

KEYWORDS: β-reduction by name – β-reduction by value – λ-Closure – λ_τ-Closure – substitution – validity.

0. Introduction

The goal of this paper is to sort out the conditions of validity of β-conversion in a hyperintensional, partial, typed λ-calculus. Since Transparent Intensional Logic (TIL) is such a system, we will examine these conditions in TIL as a sample theory.[2] The terms of TIL are interpreted procedurally, which is to say that they denote *procedures* (roughly, Church's functions-in-intension) producing set-theoretical functions/mappings (Church's functions-in-extension) rather than the mappings themselves. This is in good harmony with the original interpretation of the terms of the lambda calculus, which was indeed procedural. For instance, Barendregt says:

> [I]n this interpretation the notion of a function is taken to be intensional, i.e., as an algorithm. (Barendregt 1997, 184)

We would rather say, "… is taken to be *hyperintensional*, i.e., as a procedure", because the term 'intensional' is currently reserved for mappings from possible worlds (if not among proof-theoretic semanticists, then at least among model-theoretic semanticists).

Thus λ-Closure, $[\lambda x_1 \ldots x_n\, X]$, transforms into the very procedure of producing a function by abstracting over the values of the variables $x_1, \ldots, x_n$. Similarly, Composition, $[X\, X_1 \ldots X_n]$, transforms into the very procedure of applying a function produced by the procedure X to the tuple-argument (if any) produced by the procedures $X_1, \ldots, X_n$. The procedural semantics of TIL makes it possible to explicitly deal with those features that are otherwise hidden if dealing only with the products of the procedures, i.e. functions-in-extension. These features concern in particular the operations in a hyperintensional context where the very procedure denoted by a term is being operated on, and such features show up also when dealing with β-conversion.

---

[2] For details on TIL see, in particular, Tichý (1988) and Duží et al. (2010).

The rule of β-reduction is a fundamental computational rule of the λ-calculi and functional programming languages. In the λ-calculi the rule is usually specified thus:

$$(\lambda x\, M)\, N \vdash M\,[x := N]$$

where M is a procedure with a formal parameter x, and M calls another procedure N to supply the actual argument value. Hence by '$M\,[x := N]$' is meant the collision-less substitution of N for all the occurrences of the variable x in the calling procedure M. However, Plotkin in (1975) pointed out that this specification is ambiguous. There are two procedurally or operationally non-equivalent ways of executing the rule, namely β-reduction '*by name*' and β-reduction '*by value*'. From the operational point of view, these two ways differ in the way the argument value is being passed for the formal parameter x. If by name, then the procedure denoted by the term N is executed *after* its substitution for all the occurrences of the variable x in the calling-procedure body M (after appropriate renaming of λ-bound variables to prevent collision). If by value, then the procedure N is executed *first*, and only if N does not fail to produce an argument *value* is this value substituted for all the occurrences of x in the body M. Plotkin (1975) put forward a programming language and a formal calculus for each calling mechanism and then showed how each determines the other. As a result, he proved that the two mechanisms are *not* operationally equivalent. Moreover, in Duží (2013 and 2014) it has been *logically* proved that these two ways of executing the conversion are not only operationally but also *denotationally non-equivalent* whenever *partial functions* are involved.

By *validity* of the β-conversion rule we mean the following. The rule is valid if and only if both the terms on the right-hand and the left-hand side of the rule denote procedures that are *strictly equivalent* in the sense that under any valuation v the two procedures produce the same function/mapping or are both v-improper, that is, fail to produce anything.[3]

In Duží & Jespersen (2015) it has been proved that β-reduction by name is not valid if partial functions are involved and the procedure denoted by

[3] As an extreme case, the produced function/mapping might be nullary, i.e. an atomic object. The produced object can be also a lower-order procedure.

the term *N* fails to produce an argument value.[4] However, there is an interpretation of λ-Closure, namely λ_τ-Closure, under which β-reduction by name is not valid even if the procedure *N* does produce an argument value.

The novel contributions of this paper are as follows. We define a variant of β-conversion by value and prove its validity regardless of whether particular constituents are improper and regardless of whether we deal with λ- or λ_τ-Closure. However, we also prove that in a special case of λ_τ-Closure this rule is not applicable. Moreover, this paper provides a systematic study of the applicability of β-conversion in a *hyperintensional* lambda calculus of partial functions, which to the best of our knowledge has not been presented until now, though similar work has been undertaken since the early 1970s, but merely for simple-typed or untyped λ-calculi. Moreover, the call-by-name strategy *cannot* be applied in a hyperintensional context, i.e., in hyperintensional λ-calculi such as TIL. The reason is that in such a context the formal parameter *x* is contained within a displayed (as opposed to executed) procedure that figures here only as an object to operate on, which makes the substitution *logically unfeasible*. Our *substitution method* based around the functions *Sub* and *Tr* is similar to Chang & Felleisen (2012)'s call-by-need reduction by value. However, their work is couched in an *un*typed λ-calculus.

The rest of the paper is organized as follows. Section 1 presents the fundamentals of TIL, especially the technical apparatus needed to deal with the rules of β-conversion. In Section 2 we introduce three variants of β-conversion and examine their validity; they are *β_n-conversion by name*, *β_v-conversion by value* and *restricted β_r-conversion by name*. In Section 3 we examine Tichý's λ_τ-Closure and show that there is an interpretation of his definition under which neither of the conversions by name is valid and β_v-conversion by value is not applicable. Thus, we recommend using β-reduction by value and λ-Closure only. Section 4 contains some concluding remarks.

---

[4] There are two other defects connected with this way of executing the rule, i.e. with calling by name, that are also demonstrated by Duží in (2013 and 2014), to wit, a loss of analytic information and non-effectiveness.

1. TIL in brief

In this section, we briefly recapitulate the technical fundamentals of TIL necessary for dealing with β-conversions. The terms of the TIL language denote abstract procedures that produce set-theoretical mappings (functions-in-extension) or lower-order procedures. These procedures are rigorously defined as TIL *constructions.* Being procedural objects, constructions can be executed in order to operate on input objects (of a lower-order type) and produce the object (if any) they are typed to produce, while non-procedural objects, i.e. non-constructions, cannot be executed. There are two atomic constructions that present input objects to be operated on. They are *Trivialization* and *Variables.* The operational sense of Trivialization is similar to that of constants in formal languages. A Trivialization presents an object X without the mediation of any other procedures. Using the terminology of programming languages, the Trivialization of X, '^0X' in symbols, is just a *pointer* to X. Variables produce objects dependently on valuations; they v-construct. We adopt an objectual variant of the Tarskian conception of variables. To each type (see Definition 2) are assigned countably many variables that range over this particular type. Objects of each type can be arranged into infinitely many sequences. The valuation v selects one such sequence of objects of the respective type, and the first variable v-constructs the first object of the sequence, the second variable v-constructs the second object of the sequence, and so on. Thus the execution of a Trivialization or a variable never fails to produce an object. However, the execution of some of the molecular constructions can fail to present an object of the type they are typed to produce. When this happens, we say that the constructions are *v-improper*. There are two kinds of improperness. Either a construction is compounded in a type-theoretically incoherent ('nonsensical') way, or it is an application of a function to an argument at which the function is not defined.

Thus, we define:

Definition 1 (*construction*)

(i) *Variables* x, y, … are *constructions* that construct objects (elements if their respective ranges) dependently on a valuation v; they v-construct.

(ii) Where X is an object whatsoever (even a construction), ${}^0X$ is the *construction Trivialization* that constructs X without any change.

(iii) Let X, $Y_1,\ldots,Y_n$ be arbitrary constructions. Then the *Composition* $[X\ Y_1\ldots Y_n]$ is the following *construction*. For any v, the Composition $[X\ Y_1\ldots Y_n]$ is *v-improper* if one or more of the constructions X, $Y_1,\ldots,Y_n$ are v-improper, or if X does not v-construct a function that is defined at the n-tuple of objects v-constructed by $Y_1,\ldots,Y_n$. If X does v-construct a v-proper function, then $[X\ Y_1\ldots Y_n]$ v-constructs the value of this function at the n-tuple.

(iv) (λ-) *Closure* $[\lambda x_1\ldots x_m\ Y]$ is the following *construction*. Let x_1, x_2, …, x_m be pair-wise distinct variables and Y a construction. Then $[\lambda x_1\ldots x_m\ Y]$ *v-constructs* the function f that takes any members B_1, …, B_m of the respective ranges of the variables x_1, …, x_m into the object (if any) that is $v(B_1/x_1,\ldots,B_m/x_m)$-constructed by Y, where $v(B_1/x_1,\ldots,B_m/x_m)$ is like v except for assigning B_1 to x_1, …, B_m to x_m.

(v) Where X is an object whatsoever, ${}^1X$ is the *construction Single Execution* that v-constructs what X v-constructs. Thus if X is a v-improper construction or not a construction as all, ${}^1X$ is v-improper.

(vi) Where X is an object whatsoever, ${}^2X$ is the *construction Double Execution*. If X is not itself a construction, or if X does not v-construct a construction, or if X v-constructs a v-improper construction, then ${}^2X$ is v-improper. Otherwise ${}^2X$ v-constructs what is v-constructed by the construction v-constructed by X.

(vii) Nothing is a *construction*, unless it so follows from (i) through (vi). □

Note that the (λ-) Closure $[\lambda x_1\ldots x_m\ Y]$ is not v-improper for any valuation v, as it always v-constructs a function. Even if the constituent Y is v-improper for every valuation v, the Closure is not v-improper. Yet in such a case the resulting function is a bizarre object; it is a degenerate function that is undefined at all arguments.

With constructions of constructions, constructions of functions, functions, and functional values in our stratified ontology, we need to keep track of the traffic between multiple logical strata. The *ramified type hierarchy* does just that. The type of first-order objects includes all objects that are not constructions. Therefore, it includes not only the standard objects of

individuals, truth-values, sets, etc., but also functions defined on possible worlds (i.e., the intensions germane to possible-world semantics). The type of second-order objects includes constructions of first-order objects and functions that have such constructions in their domain or range. The type of third-order objects includes constructions of first- and second-order objects and functions that have such constructions in their domain or range. And so on, ad infinitum.

Definition 2 (*ramified hierarchy of types*)
Let *B* be a *base*, where a base is a collection of pair-wise disjoint, non-empty sets. Then:

T_1 (*types of order 1*).
(i) Every member of *B* is an elementary *type of order 1 over B.*
(ii) Let α, β_1, …, β_m ($m > 0$) be types of order 1 over *B*. Then the collection ($\alpha\ \beta_1 \dots \beta_m$) of all *m*-ary partial mappings from $\beta_1 \times \dots \times \beta_m$ into α is a functional *type of order 1 over B.*
(iii) Nothing is a *type of order 1 over B* unless it so follows from (i) and (ii).

C_n (*constructions of order n*)
(i) Let *x* be a variable ranging over a type of order *n*. Then *x* is a *construction of order n over B.*
(ii) Let *X* be a member of a type of order *n*. Then ^0X, ^1X, ^2X are *constructions of order n over B.*
(iii) Let *X,* X_1, …, X_m ($m > 0$) be constructions of order *n* over *B*. Then $[X\ X_1 \dots\ X_m]$ is a *construction of order n over B.*
(iv) Let x_1, …, x_m, *X* ($m > 0$) be constructions of order *n* over *B*. Then $[\lambda x_1 \dots x_m\ X]$ is a *construction of order n over B*.
(v) Nothing is a *construction of order n over B* unless it so follows from C_n (i)-(iv).

T_{n+1} (*types of order n + 1*)
Let $*_n$ be the collection of all constructions of order *n* over *B*. Then
(i) $*_n$ and every type of order *n* are *types of order n + 1.*
(ii) If $m > 0$ and α, β_1, …, β_m are types of order $n + 1$ over *B*, then ($\alpha\ \beta_1 \dots \beta_m$) (see T_1 ii)) is a *type of order n + 1 over B.*
(iii) Nothing is a *type of order n + 1 over B* unless it so follows from (i) and (ii). □

For the purposes of natural-language analysis, we are usually assuming the following base of ground types:

o: the set of truth-values {**T**, **F**};
ι: the set of individuals (the universe of discourse);
τ: the set of real numbers (doubling as discrete times);
ω: the set of logically possible worlds (the logical space).

We model sets and relations by their characteristic functions. Thus, for instance, (οι) is the type of a set of individuals, while (οιι) is the type of a relation-in-extension between individuals. Empirical expressions denote *empirical conditions* that may or may not be satisfied at the world/time pair selected as points of evaluation. We model these empirical conditions as possible-world-semantic *intensions.* Intensions are entities of type (βω): mappings from possible worlds to an arbitrary type β. The type β is frequently the type of the *chronology* of α-objects, i.e., a mapping of type (ατ). Thus α-intensions are frequently functions of type ((ατ)ω), abbreviated as '$\alpha_{\tau\omega}$'. *Extensional entities* are entities of a type α where $\alpha \neq (\beta\omega)$ for any type β. Where *w* ranges over ω and *t* over τ, the following logical form essentially characterizes the logical syntax of empirical language:

$\lambda w \lambda t$ [*…w….t…*].

Examples of frequently used intensions are: *propositions* of type $o_{\tau\omega}$, *properties of individuals* of type $(o\iota)_{\tau\omega}$, *binary relations-in-intension between individuals* of type $(o\iota\iota)_{\tau\omega}$, *individual offices* (or *roles*) of type $\iota_{\tau\omega}$.

Logical objects like *truth-functions* and *quantifiers* are extensional: ∧ (conjunction), ∨ (disjunction) and ⊃ (implication) are of type (οοο), and ¬ (negation) of type (οο). The *quantifiers* $\forall^\alpha$, $\exists^\alpha$ are type-theoretically polymorphic total functions of type (o(oα)), for an arbitrary type α, defined as follows. The *universal quantifier* $\forall^\alpha$ is a function that associates a class *A* of α-elements with **T** if *A* contains all elements of the type α, otherwise with **F**. The *existential quantifier* $\exists^\alpha$ is a function that associates a class *A* of α-elements with **T** if *A* is a non-empty class, otherwise with **F**. Below all type indications will be provided outside the formulae in order not to clutter the notation. Moreover, the outermost brackets of Closures will be omitted whenever no confusion can arise. Furthermore,

'X/α' means that an object X is (a member) of type α. '$X \rightarrow_v \alpha$' means that X is typed to v-construct an object of type α, regardless of whether X in fact constructs anything. We write '$X \rightarrow \alpha$' if what is v-constructed does not depend on a valuation v. Throughout, it holds that the variables $w \rightarrow_v \omega$ and $t \rightarrow_v \tau$. If $C \rightarrow_v \alpha_{\tau\omega}$ then the frequently used Composition $[[C\ w]\ t]$, which is the intensional descent (a.k.a. extensionalization) of the α-intension v-constructed by C, will be encoded as 'C_{wt}'.

In order to work with a hyperintensional context, in which a *construction* is operated on, we need two special functions, *Sub* and *Tr*. The polymorphic function *Sub* of type $(*_n*_n*_n*_n)$ operates on constructions as follows. When applied to constructions C_1, C_2, C_3, *Sub* returns as its value the construction D that is the result of the correct (i.e. collision-less) substitution of C_1 for C_2 in C_3. For instance, the result of the Composition $[^0Sub\ ^{00}John\ ^0him\ ^0[^0Wife\_of_{wt}\ him]]$ is the Composition $[^0Wife\_of_{wt}\ ^0John]$. The logical operation of substitution is treated as a theoretical primitive.

The likewise polymorphic function *Tr* returns as its value the Trivialization of its argument. Thus the result of $[^0Tr\ ^0John]$ is ^0John. If what is wanted as output is the Trivialization of the Trivialization of John, the corresponding Composition is $[^0Tr\ ^{00}John]$. When x ranges over ι, the Composition $[^0Tr\ x]$ *v(John/x)*-constructs ^0John. Note one essential difference between the function *Tr* and the construction Trivialization. Whereas the variable x is *free* in $[^0Tr\ x]$, the Trivialization ^0x *binds* the variable x by constructing just x independently of valuation.

2. β-conversion

In the lambda calculi, the rule of β-conversion is usually specified in this form:

$$(\lambda x.M)\ N =_\beta M\ [x:=N]$$

The right-hand side contractum is the result of substituting the *term N* for all free occurrences of the variable x within the term M. The rule of β-reduction is the left-to-right part of the above equality:

$$(\beta) \qquad (\lambda x.M)\ N \rightarrow_\beta M\ [x:=N]$$

The rule (β) is used to model the application of the function referred to by the term λ*x.M* to the argument denoted by *N*. Using programming-language technical jargon, we can explicate the rule as follows. The term λ*x.M* denotes, or *declares*, a procedure with a formal parameter *x* and the procedural body *M*. Thus the redex on the left-hand side denotes a procedure that consists in calling the procedure λ*x.M* which is to be executed with the actual argument value replacing the formal parameter *x*, and this value is to be provided by the sub-procedure *N*. The contractum term on the right-hand side is schematic. In principle, it can be read as the instruction to execute the procedural body *M* in which the formal parameter *x* has been replaced by the actual argument value provided by the procedure *N*. In case *N* fails to produce an argument value, the procedure body *M* has nothing to operate on, and thus the rule (β) cannot be applied. While in the λ-calculi of total functions this fact is irrelevant, in the λ-calculi of partial functions this eventuality has to be taken into account.[5]

Partiality, as we only know too well, brings about technical complications. However, we do need to work with partial functions, because otherwise we face the problem of a non-recursive explosion of domains that is computationally non-tractable (for details see Duží 2003). Yet just a few results have been obtained in this area. Moggi (1988) would appear to have been the first to put forward a definition of a partial λ-calculus, and Feferman (1995) presents a set of axioms for the *Partial Lambda Calculus* (for details see, e.g., Duží et al. 2010, § 2.7, 261-262). However, they both specify the predicate '↓' which in '*N*↓' means that the term *N* is 'defined' or 'referring'. Consequently, the rule is valid in the sense of weak congruency; if both sides are defined then they denote the same value. However, such a restriction to non-recursively defined cases of *v*-properness would be a serious shortcoming of TIL or indeed any other formal semantics based on the λ-calculus. Hence, we do need a *universally valid* rule regulating β-transformation. TIL is a λ-calculus of partial functions, and in virtue of its procedural semantics we have the technical machinery required to specify a universally valid rule of β-conversion.

---

[5] A partial function is a function with *at most* one value at each argument. Every total function is, therefore, a partial function, but not vice versa.

2.1. Three kinds of β-conversion

Now we are going to examine three kinds of β-conversion using the technical apparatus of TIL. The three kinds are β-*conversion by name* (β_n-conversion), β-*conversion by value* (β_v-conversion), and *restricted* β-*conversion by name* (β_r-conversion). We will use simple examples to illustrate them. Let the calling procedure *M* and the called procedure *N* be $[\lambda x\ \lambda y\ [{}^0{>}\ y\ x]]$ and $[{}^0Div\ {}^0 3\ {}^0 0]$, respectively. Then we have the Composition

(1) $[[\lambda x\ \lambda y\ [{}^0{>}\ y\ x]]\ [{}^0Div\ {}^0 3\ {}^0 0]]$

Types. $x, y \rightarrow_v \tau$; $>/(o\tau\tau)$; $Div/(\tau\tau\tau)$: the division function; $3,0/\tau$.

The Closure $[\lambda x\ \lambda y\ [{}^0{>}\ y\ x]]$ produces a mapping of type $((o\tau)\tau)$, i.e. a function *f* that associates a number *x* with the class of numbers *y* that are larger than the number *x*.

However, as mentioned above, partiality is a complicating factor. Some molecular constructions can be *v-improper* in the sense of failing to produce the sort of object they are typed to construct. There are two kinds of improperness, as we said above. Either a construction is compounded in a type-theoretically incoherent way, or it is the procedure of applying a function to an argument at which the function is not defined. We will now address the latter kind of improperness. Improperness rooted in wrong typing will be examined in Section 3 below.

The Composition $[{}^0Div\ {}^0 3\ {}^0 0]$ is the procedure of applying the division function to arguments 3 and 0. Since dividing any number by 0 is not defined, this Composition does not *v*-construct anything for any valuation *v*; it is *v-improper* for any valuation *v*, or *improper* for short.

The Composition (1) is the procedure of applying the function *f* constructed by $[\lambda x\ \lambda y\ [{}^0{>}\ y\ x]]$ to the argument that is to be produced by the Composition $[{}^0Div\ {}^0 3\ {}^0 0]$. Yet since this Composition does not produce anything, there is no argument to apply *f* to. Hence, the Composition (1) is by Def. 1 also *improper*.

2.1.1. β-conversion by name

The result of applying β_n-conversion to (1) is that the *x* in the 'body' of *M* is replaced by $[{}^0Div\ {}^0 3\ {}^0 0]$. This yields:

(2) $[[\lambda x\, \lambda y\, [{}^0{>}\, y\, x]]\, [{}^0Div\ {}^0 3\ {}^0 0]] \rightarrow_{\beta n} [\lambda y\, [{}^0{>}\, y\, [{}^0Div\ {}^0 3\ {}^0 0]]]$

This result demonstrates the problem of β_n-reduction in the logic of partial functions. While the left-hand side Composition of (2) is improper, the right-hand side contractum is *not improper.* It produces a *degenerate function* undefined at all its arguments. In other words, we obtain an empty class of numbers, the characteristic function of which is undefined at any number. Bizarre as it is, it is still something rather than nothing and therefore an object. Hence the left-hand and the right-hand side constructions of (2) are not strictly equivalent, hence the β_n-rule is not valid.

In this simple case, the absence of strict equivalence might seem harmless. After all, if that bizarre function is applied to a number, the result is an improper construction; hence also a gap comparable to a truth-value gap, and the final result would be the same. Yet our operational semantics reveals that it is not quite as harmless as it might seem. The execution of the left-hand side construction is improper, which is something we already know. It makes no sense to execute this construction, because it fails to produce something. However, in the right-hand side construction this fact is hidden. We end up with a procedure producing a function, and only after calling this procedure a second time is this failure revealed.

This deficiency is best demonstrated by an analysis of an empirical attitude *de re*. Consider:

(3) Tom believes *of* the Pope that he is wise.

On the *de re* reading of (3) the property of being believed by Tom to be wise is ascribed to the individual (if any) that holds the papal office. Thus, the analysis amounts to this construction:

(3*) $\lambda w \lambda t\, [\lambda he\, [{}^0Believe_{wt}\ {}^0Tom\ \lambda w^* \lambda t^*\, [{}^0Wise_{w^*t^*}\ he]]\ {}^0Pope_{wt}]$

Types. Variable $he \rightarrow_\nu \iota$; $Believe/(o\iota o_{\tau\omega})_{\tau\omega}$; Tom/ι; $Wise/(o\iota)_{\tau\omega}$; $Pope/\iota_{\tau\omega}$.[6]

---

[6] For the sake of simplicity, we analyse the attitude of believing intensionally, that is, as a relation-in-intension to a *proposition*, which makes for an *implicit* attitude. The believer is related to the proposition regardless of the particular way the proposition is conceptualized or constructed. This approach yields notorious problems with logical-mathematical omniscience. Thus, a more appropriate analysis would be hy-

The construction of the papal office, ^0Pope, occurs in (3*) extensionally, i.e. with *de re* supposition. Thus if the Pope does not exist (that is, if the papal office is not occupied in world *w* and time *t* of evaluation) then its extensionalization $^0Pope_{wt}$ is *v*-improper, and (3*) constructs a proposition that lacks a truth-value at the relevant ⟨*w*, *t*⟩-pair. This is as it should be, though, because there is an *existential presupposition de re*.[7] Now executing β-reduction by name consists in replacing the 'formal parameter' *he* (that is, the variable *he*) by the Composition $^0Pope_{wt}$, which in turn yields this construction:

(4) $\lambda w \lambda t\, [^0Believe_{wt}\ ^0Tom\ \lambda w^* \lambda t^*\, [^0Wise_{w^*t^*}\ ^0Pope_{wt}]]$

However, in (4) ^0Pope does *not* occur with supposition *de re*. This is because $^0Pope_{wt}$ has been drawn into the λ-generic intensional context of Tom's perspective ($\lambda w^* \lambda t^*$), and an intensional context is dominant over the lower extensional one, which in turn means that the improperness of $^0Pope_{wt}$ is suppressed or irrelevant. If $^0Pope_{wt}$ is *v*-improper, then Tom believes that the degenerate proposition *v*-constructed by the Closure $\lambda w^* \lambda t^*\, [^0Wise_{w^*t^*}\ ^0Pope_{wt}]$ is true, which is a logical possibility. In other words, there is no logical reason for the proposition constructed by (4) to be undefined. Thus β_n-reduction has turned a *de re* occurrence into a *de dicto* occurrence, which is wrong.

For these reasons a necessary condition for the validity of β-reduction by name is usually specified, namely that the procedure that is typed to produce an argument value be proper. For instance, Raclavský (2009) presents the following definition of the validity of β-reduction by name:

> Let *C* be a closure of the form λ*x* […*x*…] that can contain also other variable than *x* (λ-bound or not). Let *C* be composed with the construction *D* in the Composition [*C D*]. Let *D* be a *v*-proper construction. If *D* contains free occurrences of variables and these variables are λ-bound

perintensional believing relating the believer to a hyperproposition, that is, a *construction* of a proposition, which makes for an *explicit* attitude: $Believe^*/(o\iota^*_n)_{\tau\omega}$. Yet as a toy example, demonstrating the invalidity of β-reduction by name the implicit *Believe* suffices.

[7] For details on the analysis of propositional attitudes *de dicto* and *de re* see, for instance, Duží et al. (2010, § 5.1) or Duží & Jespersen (2012).

> in C, then let C be α-expanded into a construction C' that does not contain the variables free in D as λ-bound. Then the construction C'' that is obtained from C' by substituting the construction D for all free occurrences of the λ-bound variable corresponding to x in C'' is the *β-reduced form of the construction C*. (Raclavský 2009, 285)[8]

Hence, the conditions for the validity of β-reduction by name can be summarized as follows:

i) the construction D of an argument value must be *v*-proper;
ii) no collision of variables must arise; if D contains free occurrences of variables that occur λ-bound in C, we must apply α-conversion to avoid collision.

This is a standard way of specifying β-conversion by name. However, in Duží & Jespersen (2013) another shortcoming of β-reduction by name has been identified. Even if β-reduction by name is a valid transformation satisfying conditions (i) and (ii), it can yield *a loss of analytic information* about *which function* has been applied to *which argument*. The authors illustrate this problem by an analysis of the well-known sentence, "John loves his wife, and so does Peter". There are two non-equivalent readings of this sentence. On the so-called sloppy reading, both John and Peter love their own wives, making them exemplary husbands. On the so-called strict reading, John and Peter share the property of loving John's wife, with trouble looming on the horizon. The problem is that β-reduction by name reduces the sloppy reading to the strict one, squeezing out the former. As a result, the anaphor resolution of 'so does Peter' invalidates the natural reading on which Peter loves his own wife whom he is presupposed not to share with John, as would be a possibility in a bigamist culture. As a solution to this problem, the authors define the rule of β-reduction by value, which we are going to examine below.

2.1.2. Restricted β-conversion by name

Above we specified the shortcomings evinced by β-conversion by name. There is, however, a restricted variant of this conversion that

[8] Translated from the Czech original by the authors.

suffers none of them. This variant is *restricted* β-conversion by name. β_r-conversion consists in collision-less substitution of free variables for λ-bound variables ranging over the same types. It is a strictly equivalent, and thus valid, conversion. For instance, $[\lambda x\, [^0{+}\, x\, ^0 1]\, y]$ can be simplified to $[^0{+}\, y\, ^0 1]$. This transformation is nothing but a manipulation with λ-bound variables that has much in common with η-reduction and much less with β-reduction. The latter is the operation of applying a function *f* to its argument *a* in order to obtain the value of *f* at *a* (leaving it open whether a value emerges). No such features can be found in β_r-reduction. It is just a formal simplification of the original construction.

For instance, above we analysed the *de re* attitudinal sentence "Tom believes of the Pope that he is wise" as ascribing the property of being believed by Tom to be wise to the holder of the papal office:

$$\lambda w \lambda t\, [\lambda he\, [^0 Believe_{wt}\, ^0 Tom\, \lambda w^* \lambda t^*\, [^0 Wise_{w^*t^*}\, he]]\, ^0 Pope_{wt}]$$

This is the β_r-restricted form of the literal analysis of the sentence "The Pope has the *property* of being believed by Tom to be wise", which amounts to

$$\lambda w \lambda t\, [\lambda w' \lambda t' [\lambda he\, [^0 Believe_{w't'}\, ^0 Tom\, \lambda w^* \lambda t^*\, [^0 Wise_{w^*t^*}\, he]]]_{wt}\, ^0 Pope_{wt}]$$

Yet we see little reason to differentiate semantically or logically between "The Pope is believed by Tom to be wise" and "The Pope has the property of being believed by Tom to be wise".[9] Hence, this kind of reduction is frequently applied in logical analysis of natural-language expressions.

[9] This is not to say we see no reason at all not to differentiate. For instance, if the believer is a self-assured nominalist then they may protest that while they do believe that the Pope is wise they do not believe that the Pope has any properties. Or it could be argued that one thing is to believe that the Pope is wise and another is to believe that the Pope has the property of being wise, because the latter at least appears to presuppose that the believer have the additional conceptual resources to master the notion of *property*.

2.1.3. β-conversion by value

Above we examined unrestricted β-conversion by name and warned against its undesirable side-effects. The difference between conversion by name and by value has consequences also from the point of view of *computational complexity*. When conversion by name is executed, the called procedure *N* is to be executed as many times as the variable *x* occurs in the calling procedure *M*. Here is a simple example for illustration. Consider the application of the identity function $\lambda x\,[x{=}x]$ to the argument computed by[10] $((1+1)/2)^2$:

$$[\lambda x\,[x{=}x]\ ((1+1)/2)^2]$$

Reduction by name results in the equality

$$\text{(A)}\quad (1+1)/2)^2 = (1+1)/2)^2$$

On the other hand, if we pass the argument by value, then we first obtain the argument value by executing the procedure $(1+1)/2)^2$. This produces the number 1, the Trivialization of which is afterwards substituted for *x*. As a result, we obtain the equality

$$\text{(B)}\quad {}^0 1 = {}^0 1$$

It is readily seen that procedure (A) is much more complicated than (B). Passing the argument to the function by name and by value makes, therefore, a difference to the computational complexity of the resulting procedure.

Hence, we need a universal rule of β-conversion that would not exhibit the above defects. Fortunately, it turns out to be feasible to formulate such a *generally valid logical rule*. The invalid rule by name is moulded on the programming technique of calling a sub-procedure *N* by name: the sub-procedure itself is substituted for the 'local variable' *x* in the 'procedure body' *M*. Programmers are well aware of the fact that this technique can

---

[10] Now we use the usual mathematical notation to make the constructions easier to read. In TIL notation the construction $((1+1)/2)^2$ would be written as '$[{}^0\mathit{Power}\ [{}^0\mathit{Div}\ [{}^0{+}\ {}^0 1\ {}^0 1]\ {}^0 2]\ {}^0 2]$', where *Power*, *Div*/$(\upsilon\upsilon\upsilon)$, υ the type of natural numbers.

have undesirable side-effects, unlike the technique of calling a sub-procedure by value.

The rule of β-reduction *by value* was originally specified logically for TIL in Duží et al. (2010, § 2.7). Unfortunately, there is a typo in Claim 2.6, (cf. Duží et al. 2010, 270) that proves its validity. The correct definition can be found in Duží (2014). Here we recapitulate the correct definition and provide the proof of validity.

Definition 3 *(β-conversion by value)*
Let $Y \to_v \alpha$; $x_1, D_1 \to_v \beta_1, \ldots, x_n, D_n \to_v \beta_n$, $[\lambda x_1 \ldots x_n\ Y] \to_v (\alpha\beta_1 \ldots \beta_n)$. Then the conversion

$$[[\lambda x_1 \ldots x_n\ Y]\ D_1 \ldots D_n] \Rightarrow_\beta {}^2[{}^0Sub\ [{}^0Tr\ D_1]\ {}^0x_1 \ldots [{}^0Sub\ [{}^0Tr\ D_n]\ {}^0x_n\ {}^0Y]]$$

is *β-reduction by value*. The reverse conversion is *β-expansion by value*. □

Claim 1
β-reduction and β-expansion by value are valid conversions. In other words, the constructions

$$[[\lambda x_1 \ldots x_n\ Y]\ D_1 \ldots D_n]$$

and

$${}^2[{}^0Sub\ [{}^0Tr\ D_1]\ {}^0x_1 \ldots [{}^0Sub\ [{}^0Tr\ D_n]\ {}^0x_n\ {}^0Y]]$$

are strictly equivalent.

Proof
Let C be identical to $[[\lambda x_1 \ldots x_n\ Y]\ D_1 \ldots D_n]$ and D to ${}^2[{}^0Sub\ [{}^0Tr\ D_1]\ {}^0x_1 \ldots [{}^0Sub\ [{}^0Tr\ D_n]\ {}^0x_n\ {}^0Y]]$. We are to prove that for any valuation v either both C and D are v-improper, or C and D v-construct the same object.

(a) If for some i, $1 \leq i \leq n$, construction D_i is v-improper then so is the Composition C, according to Def. 1, iii). Then also the Compositions $[{}^0Tr\ D_i]$ and $[{}^0Sub\ [{}^0Tr\ D_1]\ {}^0x_1 \ldots [{}^0Sub\ [{}^0Tr\ D_n]\ {}^0x_n\ {}^0Y]]$ are v-improper according to Def. 1, iii), and thus also the construction D is v-improper according to Def. 1, vi).

(b) Otherwise, let $D_1, \ldots, D_n$ all be v-proper, v-constructing the objects $d_1, \ldots, d_n$, respectively. Then by Def. 1, iv) the Closure $[\lambda x_1 \ldots x_n\ Y]$ v-constructs the function $f/(\alpha\beta_1 \ldots \beta_n)$.

(b1) If Y is $v(d_1/x_1,\ldots,d_n/x_n)$-improper, then f is undefined on $\langle d_1,\ldots,d_n\rangle$ and thus Composition C is v-improper according to Def. 1, iii). We are to show that D is also v-improper. The Composition $[{}^0Sub\ [{}^0Tr\ D_1]\ {}^0x_1\ \ldots\ [{}^0Sub\ [{}^0Tr\ D_n]\ {}^0x_n\ {}^0Y]]$ v-constructs $Y({}^0d_1/x_1,\ldots,\ {}^0d_n/x_n)$, i.e. the construction Y where all the occurrences of the variables $x_1,\ldots,x_n$ have been replaced by ${}^0d_1,\ldots,{}^0d_n$, respectively. Since Y is $v(d_1/x_1,\ldots,d_n/x_n)$-improper, the execution of $Y({}^0d_1/x_1,\ldots,\ {}^0d_n/x_n)$, hence D, is v-improper as well according to Def. 1, vi).

(b2) Otherwise, if Y is not $v(d_1/x_1,\ldots,d_n/x_n)$-improper, then the value of f on $\langle d_1,\ldots,d_n\rangle$ is the α-entity $v(d_1/x_1,\ldots,d_n/x_n)$-constructed by Y. Let this α-entity be a. Then by Def. 1, iii), construction C v-constructs a. We are to show that construction D also v-constructs a. The first Execution of D v-constructs $Y({}^0d_1/x_1,\ldots,\ {}^0d_n/x_n)$. Since the Trivializations ${}^0d_1,\ldots,{}^0d_n$ construct the entities $d_1,\ldots,d_n$, respectively, the second Execution v-constructs the entity a.

Hence, C and D come out strictly equivalent.

In Section 2.1.1 we demonstrated the invalidity of the β_n-conversion of the Composition

$$[\lambda x\ [\lambda y\ [{}^0> y\ x]]\ [{}^0Div\ {}^03\ {}^00]]$$

Using the rule of β_v-conversion defined above, here is a valid conversion of this Composition:

$$[\lambda x\ [\lambda y\ [{}^0> y\ x]]\ [{}^0Div\ {}^03\ {}^00]] \Rightarrow_\beta {}^2[{}^0Sub\ [{}^0Tr\ [{}^0Div\ {}^03\ {}^00]]\ {}^0x\ {}^0[\lambda y\ [{}^0> y\ x]]]$$

It is readily seen that both the left-hand and the right-hand side constructions are improper. Indeed, since $[{}^0Div\ {}^03\ {}^00]$ is improper, by Def. 1, iii) the Composition $[\lambda x\ [\lambda y\ [{}^0> y\ x]]\ [{}^0Div\ {}^03\ {}^00]]$ is improper. For the same reason, the Composition $[{}^0Tr\ [{}^0Div\ {}^03\ {}^00]]$ is improper and thus also the whole Composition $[{}^0Sub\ [{}^0Tr\ [{}^0Div\ {}^03\ {}^00]]\ {}^0x\ {}^0[\lambda y\ [{}^0> y\ x]]]$ as well as its Double Execution are improper. Partiality is strictly propagated up, as it should be.

Similarly, the analysis of the *de re* attitude (3*) can be validly reduced in this way:

$\lambda w\lambda t\,[\lambda he\,[{}^0Believe_{wt}\ {}^0Tom\ \lambda w^*\lambda t^*\,[{}^0Wise_{w^*t^*}\ he]]\ {}^0Pope_{wt}] \Rightarrow_\beta$
$\lambda w\lambda t\,{}^2[{}^0Sub\,[{}^0Tr\ {}^0Pope_{wt}]\ {}^0he\ {}^0[{}^0Believe_{wt}\ {}^0Tom\ \lambda w^*\lambda t^*\,[{}^0Wise_{w^*t^*}\ he]]]$

Remark: The reduced construction is actually the *literal* analysis of the sentence "Tom believes of the Pope that he is wise". The anaphoric reference 'he' referring to the holder of the papal office is resolved by the substitution of this holder (if any) for the variable *he*, that is, by the constituents $[{}^0Sub\,[{}^0Tr\ {}^0Pope_{wt}]\ {}^0he$ ….

Proof

We are to prove that for any world *w* and time *t* of evaluation, the constructions

$[\lambda he\,[{}^0Believe_{wt}\ {}^0Tom\ \lambda w^*\lambda t^*\,[{}^0Wise_{w^*t^*}\ he]]\ {}^0Pope_{wt}]$

and

${}^2[{}^0Sub\,[{}^0Tr\ {}^0Pope_{wt}]\ {}^0he\ {}^0[{}^0Believe_{wt}\ {}^0Tom\ \lambda w^*\lambda t^*\,[{}^0Wise_{w^*t^*}\ he]]]$

v-construct the same truth-value or are both *v*-improper.

1) Let ${}^0Pope_{wt}$ *v*-construct an individual *a*. Then we will show that both constructions *v*-construct the same truth-value as does the Composition $[{}^0Believe_{wt}\ {}^0Tom\ \lambda w^*\lambda t^*\,[{}^0Wise_{w^*t^*}\ {}^0a]]$. In any world *w* and time *t* of evaluation the following steps are truth-preserving:

   ⇒

   a) $[\lambda he\,[{}^0Believe_{wt}\ {}^0Tom\ \lambda w^*\lambda t^*\,[{}^0Wise_{w^*t^*}\ he]]\ {}^0Pope_{wt}]$ ∅
   b) ${}^0Pope_{wt} = {}^0a$ ∅
   c) $[\lambda he\,[{}^0Believe_{wt}\ {}^0Tom\ \lambda w^*\lambda t^*\,[{}^0Wise_{w^*t^*}\ he]]\ {}^0a]$ a), b), SI (Leibniz)
   d) $[{}^0Proper\ {}^0a]$ by Def. 1
   e) $[{}^0Believe_{wt}\ {}^0Tom\ \lambda w^*\lambda t^*\,[{}^0Wise_{w^*t^*}\ {}^0a]]$ β-reduction by name

   ⇐

   f) ${}^2[{}^0Sub\,[{}^0Tr\ {}^0Pope_{wt}]\ {}^0he\ {}^0[{}^0Believe_{wt}\ {}^0Tom\ \lambda w^*\lambda t^*\,[{}^0Wise_{w^*t^*}\ he]]]$ ∅
   g) ${}^0Pope_{wt} = {}^0a$ ∅
   h) $[{}^0Tr\ {}^0Pope_{wt}] = [{}^0Tr\ {}^0a]$ g), SI, Def. of *Tr*
   i) ${}^2[{}^0Sub\,[{}^0Tr\ {}^0a]\ {}^0he\ {}^0[{}^0Believe_{wt}\ {}^0Tom\ \lambda w^*\lambda t^*\,[{}^0Wise_{w^*t^*}\ he]]]$ h), SI
   j) $[{}^0Believe_{wt}\ {}^0Tom\ \lambda w^*\lambda t^*\,[{}^0Wise_{w^*t^*}\ {}^0a]]$ i), Def. of *Sub*, Def.1, vi)

2) Let ${}^0Pope_{wt}$ be *v*-improper. Then by Def. 1, iii), vi) all the Compositions

$[{}^0Tr\ {}^0Pope_{wt}]$,
$[\lambda he\ [{}^0Believe_{wt}\ {}^0Tom\ \lambda w^*\lambda t^*\ [{}^0Wise_{w^*t^*}\ he]]\ {}^0Pope_{wt}]$
$[{}^0Sub\ [{}^0Tr\ {}^0Pope_{wt}]\ {}^0he\ {}^0[{}^0Believe_{wt}\ {}^0Tom\ \lambda w^*\lambda t^*\ [{}^0Wise_{w^*t^*}\ he]]]$

and thus also the Double Execution

${}^2[{}^0Sub\ [{}^0Tr\ {}^0Pope_{wt}]\ {}^0he\ {}^0[{}^0Believe_{wt}\ {}^0Tom\ \lambda w^*\lambda t^*\ [{}^0Wise_{w^*t^*}\ he]]]$

are *v*-improper. □

Remarks: At steps (c), (h) and (i) the rule of substitution of identicals for extensional contexts is applied. More precisely, since ${}^0Pope$ occurs extensionally (*de re*), the *v*-congruent constructions ${}^0Pope_{wt}$ and ${}^0a$ are substitutable salva veritate here. For details, see Duží et al. (2010, §2.7.1) and Duží (2013). Step (e) is justified by step (d). Since the Trivialization of an entity is never *v*-improper, β-reduction by name can be validly applied here.

3. λ_α-Closure and β-conversion

We have so far tacitly applied the definition of λ-Closure as per Definition 1, iv):

> (λ-)*Closure* $[\lambda x_1 \ldots x_m\ Y]$ is the following *construction*. Let $x_1,\ldots,\ x_m$ be pair-wise distinct variables and Y a construction. Then $[\lambda x_1\ \ldots\ x_m\ Y]$ *v-constructs* a function f that takes any members $B_1,\ldots,\ B_m$ of the respective ranges of the variables $x_1,\ldots,\ x_m$ into the object (if any) that is $v(B_1/x_1,\ldots,B_m/x_m)$-constructed by Y, where $v(B_1/x_1,\ldots,B_m/x_m)$ is like v except for assigning B_1 to x_1, …, B_1 to x_m.

According to this definition, λ-Closure is not *v*-improper for any valuation *v*; it always *v*-constructs a function *f* of the following type. Let $x_1 \to \beta_1$, …, $x_m \to \beta_m$, and let Y be typed to *v*-construct objects of type α. Then the function *f* is of type $(\alpha\ \beta_1\ldots\beta_m)$. However, the function *f* can be a *degenerate function*, which takes no argument to a value. This is the case when Y is *v*-improper for any valuation *v*.

Tichý (1988) applies the definition of λ_α-Closure that leaves room for a slightly different procedure than the above. Recapitulating Tichý's definition, we have:[11]

> To generalize, let α be a type, $x_1, \ldots, x_m$ distinct variables ranging over the respective types $\beta_1, \ldots, \beta_m$, and v a valuation. *Any construction Y* can be used in constructing a mapping from $\beta_1, \ldots, \beta_m$ into α; we shall call this latter construction the λ_α-*closure of Y on* $x_1, \ldots, x_m$, or briefly $[\lambda_\alpha x_1 \ldots x_m\, Y]$. For any v, $[\lambda_\alpha x_1 \ldots x_m\, Y]$ v-constructs the mapping which takes any $X_1, \ldots, X_m$ of the respective types $\beta_1, \ldots, \beta_m$ into that member (if any) of α which is $v(X_1/x_1, \ldots, X_m/x_m)$-constructed by Y, where $v(X_1/x_1, \ldots, X_m/x_m)$ is like v except for assigning X_1 to x_1, ..., and X_m to x_m. (Tichý 1988, 65; emphasis ours)

Claim 2
Every λ-Closure is a λ_α-Closure, but not vice versa.

Proof
Let $Y \rightarrow_v \gamma$ and $\alpha = \gamma$. Then λ-Closure and λ_α-Closure are identical procedures producing the same (possibly degenerate) function *f*. However, if $\alpha \neq \gamma$ then λ-Closure is a procedure identical to λ_γ-Closure that produces a function $f/(\gamma\ \beta_1 \ldots \beta_m)$ while λ_α-Closure is another procedure producing a (degenerate) function $g/(\alpha\ \beta_1 \ldots \beta_m)$. □

Claim 3
The λ_α-Closure $[\lambda_\alpha x_1 \ldots x_m\, Y]$, $Y \rightarrow_v \gamma$ and $\alpha \neq \gamma$, v-constructs a degenerate function.

Proof is obvious. If Y is typed to v-construct objects of type γ then the resulting mapping does not take any $X_1, \ldots, X_m$ of the respective types $\beta_1, \ldots, \beta_m$ into that member (if any) of type α which is $v(X_1/x_1, \ldots, X_m/x_m)$-constructed by Y, because there is no such member.

Hence, if $\alpha \neq \gamma$ then λ_α-Closure is not identical to any λ-Closure. According to Tichý's definition, λ_α-Closure is also never v-improper for any

---

[11] Tichý uses here the term 'collection', because his definition of types follows only later in the text. For the sake of simplicity and in the interest of a smooth reading, we use 'type' instead of 'collection'.

valuation *v*. It always *v*-constructs a function *f*; but again, this function can be degenerate. As with λ-Closure, the function *f* is degenerate in case *Y* is *v*-improper for any valuation *v*. But, importantly, this definition leaves room for another way of *v*-constructing a degenerate function. Suppose that *Y* is typed to *v*-construct objects of type γ, where γ ≠ α. Then, since for any valuation *v* no member of the type α is *v*-constructed by *Y*, the function *fl*(α $\beta_1 \ldots \beta_m$) is degenerate even if *Y* itself is not *v*-improper.[12]

A simple example to illustrate the situation. Let *C* be the λ_τ-*Closure*

$$[\lambda_\tau x\ [{}^0{=}\ {}^0 0\ {}^0 1]]$$

where τ is the type of real numbers, $x \rightarrow_v \tau$. Then the function *f* produced by $[\lambda_\tau x\ [{}^0{=}\ {}^0 0\ {}^0 1]]$ is a degenerate function. The reason is this. According to the strict reading of Tichý's definition the λ_τ-Closure *C* produces a mapping of type (ττ). Yet the Composition $[{}^0{=}\ {}^0 0\ {}^0 1]$ produces the truth-value **F**, which is an object of type o. Hence no object of type τ is *v*-constructed by $[{}^0{=}\ {}^0 0\ {}^0 1]$, and thus *f* does not return any value at any number. Note the difference between $[\lambda_\tau x\ [{}^0{=}\ {}^0 0\ {}^0 1]]$ and $[\lambda x\ [{}^0{=}\ {}^0 0\ {}^0 1]]$. While the former constructs a degenerate function *f* of type (ττ), the latter constructs an empty class of numbers, that is, an object of type (oτ).

The difference between λ-Closure and λ_α-Closure affects also the validity of β-conversion. When dealing with the validity of rules of β-conversion, we have considered so far only one problematic issue, namely the case when the procedure *N* that is to produce an argument value is *v*-improper by failing to do so. We have shown that in such a case the unrestricted rule of β-reduction by name is not a valid rule, while the restricted version of β-reduction by name and β-reduction by value are valid rules.

Nonetheless, there is another problematic issue, namely the procedure of applying a degenerate function *f* to an argument value. Trivially, such a procedure fails to produce anything, because a degenerate function returns no value at any argument. We have seen that the λ-Closure $[\lambda x_1 \ldots x_m Y]$ *v*-constructs a degenerate function *f* in case *Y* is *v*-improper for any valuation

[12] True, it is dubious whether Tichý indeed intended the interpretation that we present here, because he does not adduce any example of such a procedure. Thus it seems that he tacitly presupposed that the type α is identical with the type γ, and that the subscript α at 'λ_α' was intended only to indicate objects of which type *Y* is typed to *v*-construct. Yet if we take his definition literally, such an interpretation is possible.

v. Yet even in this case the rule of β-conversion by value is valid, as we have proved in *Claim* 1.

However, the λ_α-Closure $[\lambda_\alpha\, x_1 \ldots x_m\, Y]$ such that $Y \rightarrow_v \gamma$ and $\alpha \neq \gamma$ *v*-constructs a degenerate function *f* even in case *Y* is not *v*-improper, which is a complicating factor.

To illustrate the situation, consider the Composition

(5) $[[\lambda_\tau x\, [{}^0{=}\, {}^0 0\, x]]\, {}^0 2]$

It satisfies both of the conditions (i) and (ii) for the validity of β-reduction by name as specified above; (i) Trivialization $^0 2$ is never *v*-improper, it constructs an argument value, namely the number 2 to which the function $f/(\tau\tau)$ constructed by $[\lambda_\tau x\, [{}^0{=}\, {}^0 0\, x]]$ is applied. The second condition (ii) is trivially satisfied, there being no collision of variables. Hence β-reduction by name would appear to be valid; but alas, it *is not*:

$$[[\lambda_\tau x\, [{}^0{=}\, {}^0 0\, x]]\, {}^0 2] \rightarrow_\beta [{}^0{=}\, {}^0 0\, {}^0 2]$$

The reason is obvious. Since *f* returns no value at any argument, its application to any number is improper. Thus the left-hand side redex, i.e. the Composition (5), is improper. However, the contracted right-hand side Composition $[{}^0{=}\, {}^0 0\, {}^0 2]$ is a proper Composition producing the truth-value **F**.

The main reason for the insufficiency of Raclavský's proposal of the validity conditions for β-reduction (by name) is that he does not take into account the strict literal reading of Tichý's definition of λ_τ-Closure $[\lambda_\tau\, x_1 \ldots x_m\, Y]$, that is, the possibility that *Y* is typed to *v*-construct objects of type α where $\alpha \neq \tau$. In other words, he works with λ-Closure rather than λ_τ-Closure.

Actually, to the best of our knowledge, the possibility that λ_α-Closure $[\lambda_\alpha\, x_1 \ldots x_m\, Y]$ can *v*-construct a degenerate function though the 'body' procedure *Y* is *v*-proper has not been taken into account up to now. Thus, we formu-late:

Claim 4

Let a construction $Y \rightarrow_v \gamma$ be *v*-proper for any valuation *v*, and let $\alpha \neq \gamma$. Further, let $D_1, \ldots, D_m$ be *v*-proper constructions of objects of the respective types $\beta_1, \ldots, \beta_m$. Then the λ_α-Closure $[\lambda_\alpha\, x_1 \ldots x_m\, Y]$ *v*-constructs a

degenerate function $f\!l(\alpha\,\beta_1\ldots\beta_m)$ due to $Y \rightarrow_v \gamma$ and $\alpha \neq \gamma$, and β-reduction by name, symbolized thus:

$$[[\lambda_\alpha\, x_1\ldots x_m\, Y]\; D_1,\ldots,D_m] \rightarrow_\beta Y(x_1{:}D_1,\ldots,x_m{:}D_m)$$

is not a valid conversion.

Proof
By assumption, the left-hand side is the procedure of applying a degenerate function to a tuple-argument provided by $D_1,\ldots,D_m$. Since a degenerate function does not have any value at any argument, this procedure is *v*-improper by failing to produce any value. Yet the right-hand side procedure is proper by assumption. □

Fortunately, β-conversion by value is unaffected by this kind of invalidity. According to Def. 3, β-conversion by value is applicable only if type α is identical to type γ. If they are not, β-conversion by value would not be valid, either. Recall the Composition (5). Reducing this Composition by value, and at the same time ignoring the necessary condition $\alpha = \gamma$, would result in:

$$[[\lambda_\tau x\; [{}^0{=}\; {}^0 0\; x]]\; {}^0 2] \Rightarrow_\beta {}^2[{}^0 Sub\; [{}^0 Tr\; {}^0 2]\; {}^0 x\; {}^0[{}^0{=}\; {}^0 0\; x]]$$

While the left-hand side is improper, because the λ_τ-Closure constructs a degenerate function of type $(\tau\tau)$, there is no logical reason for the right-hand side to be improper. The right-hand side construction constructs the truth-value **F**. This is because the result of the substitution is the Composition $[{}^0{=}\; {}^0 0\; {}^0 2]$, the execution of which yields **F**. In other words, the following constructions are equivalent by constructing the same truth-value **F**:

$${}^2[{}^0 Sub\; [{}^0 Tr\; {}^0 2]\; {}^0 x\; {}^0[{}^0{=}\; {}^0 0\; x]]$$
$${}^{20}[{}^0{=}\; {}^0 0\; {}^0 2]$$
$$[{}^0{=}\; {}^0 0\; {}^0 2]$$

Claim 5
Let λ_α-Closure $[\lambda_\alpha\; x_1\ldots x_m\; Y]$ *v*-construct a degenerate function $f\!l(\alpha\,\beta_1\ldots\beta_m)$ due to $Y \rightarrow_v \gamma$ and $\alpha \neq \gamma$. Then β-reduction by value:

$$[[\lambda_\alpha\, x_1\ldots x_m\, Y]\; D_1,\ldots,D_m] \Rightarrow_\beta$$
$${}^2[{}^0 Sub\; [{}^0 Tr\; D_1]\; {}^0 x_1\; \ldots\; [{}^0 Sub\; [{}^0 Tr\; D_n]\; {}^0 x_n\; {}^0 Y]]$$

is not applicable.

Proof
According to Def. 3, in order that the rule of β-reduction by value be applicable, the types α and γ must be identical, which they fail to be here. □

Let us now examine the validity of β_r-reduction when the λ_α-Closure $[\lambda_\alpha x_1 \ldots x_m Y]$ *v*-constructs a function that is degenerate because $Y \rightarrow_v \gamma$, $\alpha \neq \gamma$. Consider again the λ_τ-Closure $[\lambda_\tau x [^0= {}^0 0 x]]$. Composing this Closure with variable $y \rightarrow_v \tau$, we obtain:

$$[[\lambda_\tau x [^0= {}^0 0 x]] y] \Rightarrow_{\beta r} [^0= {}^0 0 y]$$

Unfortunately, the Composition $[^0= {}^0 0 y]$ *is not* *v*-improper for any *v*, while the redex $[[\lambda_\tau x [^0= {}^0 0 x]] y]$ *is* *v*-improper for any *v*.

Claim 6
Let λ_α-Closure $[\lambda_\alpha x_1 \ldots x_m Y]$ *v*-construct a degenerate function $f/(\alpha \beta_1 \ldots \beta_m)$ due to $Y \rightarrow_v \gamma$ where $\gamma \neq \alpha$. Further, let *Y* be *v*-proper for any valuation *v*, and let $y_1, \ldots, y_m$ be variables ranging over the respective types $\beta_1, \ldots, \beta_m$. Then the restricted β_r-reduction by name:

$$[[\lambda_\alpha x_1 \ldots x_m Y] y_1, \ldots, y_m] \rightarrow_\beta Y(x_1 : y_1, \ldots, x_m : y_m)$$

is not a valid conversion.

Proof is obvious.

4. Concluding remarks

Above we have examined the conditions for the validity of the rule of β-reduction in the hyperintensional, typed λ-calculus of partial functions. While unconditional β-reduction by name is not a strictly equivalent transformation in the logic of partial functions, β-reduction by value and restricted β-reduction by name are strictly equivalent, hence valid conversions. If reduction by name is to be validly applied, then none of the constituents of the application procedure must be *v*-improper. This is the case of restricted β_r-reduction, which merely substitutes variables for λ-bound

variables of the same respective types. Such a reduction is often applied in the analysis of empirical natural-language expressions.

In current TIL as expounded in Duží et al. (2010) and later, only λ-Closure has been considered while Tichý (1988) defined λ_α-Closure, $[\lambda_\alpha x_1 \ldots x_m\ Y]$. We showed that Tichý's definition leaves room for a procedure that produces a degenerate function even if Y is not v-improper for any valuation v. This interpretation is, however, fatal for β-reduction. The rule of β-reduction by value is not applicable, and the rule of restricted β_r-reduction is not valid even in case Y is not v-improper for any v, but Y is typed to v-construct objects of a type different from α. For this reason we recommend working only with λ-Closure, that is, with a λ_α-Closure $[\lambda_\alpha x_1 \ldots x_m\ Y]$ such that Y is typed to v-construct objects of type α.

For background, in programming languages the difference between β-reduction by name and by value revolves around the choice of *evaluation strategy*. Historically, call-by-value and call-by-name date back to *Algol 60*, a language designed in the late 1950s. The difference between call-by-name and call-by-value is often called *passing by reference* vs. *passing by value*, respectively. Strangely enough, purely functional programming languages such as *Clean* and *Haskell* use call-by-name. In our opinion, call by value would be a better evaluation strategy. For instance, *Java* manipulates objects by reference. However, *Java* does not pass arguments by reference, but by value. Call-by-value is not a single evaluation strategy, but rather a cluster of evaluation strategies in which a function's argument is evaluated before being passed to the function. In *call-by-reference* evaluation (also referred to as *call-by name* or *pass-by-reference*), a calling procedure receives an implicit reference to the argument sub-procedure. This typically means that the calling procedure can modify the argument sub-procedure. A call-by-reference language makes it more difficult for a programmer to track the effects of a procedure call, and may introduce subtle bugs.

Our proposal amounts to a *logical specification of an evaluation strategy by-value as adapted to TIL*. We have also developed a computational variant of TIL, the so-called *TIL-Script* language. For the reasons set out above, the grammar of the TIL-Script language does not make it possible to define λ_α-Closure, hence only λ-Closure is used. Finally, only the call-by-value reduction strategy is applied, which is thus universally applicable and valid.

Acknowledgments

This research has been supported by the Grant Agency of the Czech Republic, project No. GA15-13277S, *Hyperintensional logic for natural language analysis*, and also by the internal grant agency of VSB_TU Ostrava, project SGS No. SP2016/100.

References

BARENDREGT, H. P. (1997): The Impact of the Lambda Calculus. *Bulletin of Symbolic Logic* 3, 181-215.

CHANG, S. & FELLEISEN, M. (2012): The Call-by-Need Lambda Calculus, Revisited. *Programming Languages and Systems*. *Lecture Notes in Computer Science*. Vol. 7211, 128-147.

DUŽÍ, M. (2003): Do We Have to Deal with Partiality? In: Bendová, K. & Jirků, P. (eds.): *Miscellanea Logica* V. Prague: Karolinum, 45-76.

DUŽÍ, M., JESPERSEN, B., & MATERNA, P. (2010): *Procedural Semantics for Hyperintensional Logic. Foundations and Applications of Transparent Intensional Logic*. Berlin: Springer.

DUŽÍ, M. (2013): Deduction in TIL: From Simple to Ramified Hierarchy of Types. *Organon F* 20, Supplementary issue 2, 5-36.

DUŽÍ, M., & JESPERSEN, B. (2012): Transparent Quantification into Hyperintensional Contexts *de re*. *Logique et Analyse* 55, No. 220, 513–554.

DUŽÍ, M., & JESPERSEN, B. (2013): Procedural Isomorphism, Analytic Information and β-conversion by Value. *Logic Journal of the IGPL* 21, No. 2, 291-308.

DUŽÍ, M. (2014): Structural Isomorphism of Meaning and Synonymy. *Computación y Sistemas* 18, No. 3, 439-453.

DUŽÍ M., & JESPERSEN, B. (2015): Transparent Quantification into Hyperintensional Objectual Attitudes. *Synthese* 192, No. 3, 635-677.

FEFERMAN, S. (1995): Definedness. *Erkenntnis* 43, No. 3, 295-320.

MOGGI, E. (1988): *The Partial Lambda-Calculus*. *PhD thesis*. University of Edinburg, available as *LFCS report* at http://www.lfcs.inf.ed.ac.uk/reports/88/ECS-LFCS-88-63/.

PLOTKIN, G. D. (1975): Call-by-Name, Call-by-Value, and the Lambda Calculus. *Theoretical Computer Science* 1, No. 2, 125-159.

PLOTKIN, G. D. (1977): LCF Considered as a Programming Language. *Theoretical Computer Science* 5, No. 3, 223-255.

RACLAVSKÝ, J. (2009): *Jména a deskripce: logicko-sémantická zkoumání* [*Names and Descriptions: Logico-Semantical Investigations*]. Olomouc: Nakladatelství Olomouc.

TICHÝ, P. (1988): *The Foundations of Frege's Logic*. Berlin: de Gruyter.

Organon F 24 (1) 2017: 37-53

Species as Individuals: Just another Class View of Species[1]

BRUNO PUŠIĆ

ABSTRACT: In this paper I will present an argument that the view of species having the ontological status of individuals implies that species actually have the ontological status of classes, despite the fact that the representatives of the view that species are individuals (or SAI) claim the contrary. Representatives of the SAI view try to argue that species cannot be classes because classes cannot change. I will show that, according to the representatives of the SAI view, groups of organisms must fulfill four necessary conditions in order to be treated as species. They must be: 1. integrated and continuous spatiotemporal genealogical lineages of organisms that are their constituent elements; 2. separated from the continuous genealogical lineage, from the last known common ancestor to modern organisms, by evolutionary unity; 3. made up of organisms going through the same or similar evolutionary processes; 4. groups of organisms whose members reproduce sexually. I will also show that when these conditions are compared to the list of extrinsic essential properties made by Caplan and Devitt it will be apparent that they are the same. In conclusion I will argue that if, under the SAI view, one of the necessary conditions that groups of organisms must fulfill in order to be treated as species is that members of the species must reproduce sexually, then each member of the species must possess the same *specific mate recognition system* or SMRS, which in turn makes SMRS an intrinsic essential property of each member of the species. What follows from this is that, according to the species and individuals view, species are in fact classes.

KEYWORDS: Essential properties – ontological status of species – species as individuals – species as classes – specific mate recognition system.

[1] Received: 31 August 2016 / Accepted: 19 December 2016

✉ Bruno Pušić
Horvaćanska cesta 28, 10 000 Zagreb, Croatia
e-mail: dna.restless@gmail.com

1. Species as individuals view

According to Wilkins (2009), the view that species have the ontological status of an individual is the only new philosophical position of the species since the modern synthesis. The author who first presented "the species-as-individuals view" (henceforth "SAI view") is a biologist, Michael Ghiselin in the paper called "On Psychologism in the Logic of Taxonomic Controversies" (see Ghiselin 1966). Later in the defense and further argumentation of the SAI view, philosopher David Hull joined in. Ghiselin and Hull are the two main representatives of the SAI view. The theory that species are individuals is allegedly the most widely accepted view on the ontological status of species among biologists (cf. Ghiselin 1992; Ereshefsky 2010).

Species cannot be classes because classes do not change. This illustrates Ghiselin's observation that the species concept is a theoretical concept in the context of evolutionary theory, which indicates the need for an alternative position to the view that species have the ontological status of classes. This is the main motivation for the SAI view and the consequent argument for that position is called the "evolutionary units argument". According to this argument, species are the result of various evolutionary processes that occur at lower levels of the biological hierarchy – genes, individuals, groups. The necessary condition for an entity to participate in any evolutionary process is spatiotemporal continuity and extension, which classes do not have. The very definition of a class entails that members of the class are spatiotemporally unrestricted, which *a priori* excludes them from participation in evolutionary processes. Spatiotemporal continuity and extension is a paradigmatic characteristic of an individual. Hull builds the SAI view on the analogy with the characteristics that are commonly attributed to individual organisms. Hull points out that the concept "individual" can be understood in a narrow and in a broader sense. In the narrow sense the concept "individual" refers to a single organism, while in a broader sense it refers to "any spatiotemporally localized and well-integrated entity" (Hull 1980, 313). For an argument in favor of the SAI view, the concept "individual" must be understood in a broader sense:

> Individuals are spatiotemporally localized entities that have reasonably sharp beginnings and endings in time. Some individuals do not change

> much during the course of their existence, others undergo considerable though limited change, and still others can change indefinitely until they eventually cease to exist. But regardless of the change that may occur, the entity must exist continuously through time and maintain its internal organization. How continuous the development, how sharp the beginnings and endings, and how well-integrated the entity must be is determined by the processes in which these individuals function, not by the contingencies of human perception. (Hull 1980, 313)

Here Hull is trying to show that the concept of individual does not necessarily refer only to individual organisms. A key property of an individual is spatiotemporal extension and location, which makes every species a historical entity. This property can be possessed by other entities as well, such as groups and, in this case, species. The difference between an individual organism and a species is that an individual organism lasts for a short period of time and its ability to change is limited by its genotype, while a species can exist over a long period of time and go through a potentially unlimited number of evolutionary changes. These changes are limited by genetic resources of a species which can potentially go through an infinite number of changes. That can also imply a change in a species without a qualitative change to a new species.

Potentially infinite variability of species does not necessarily make a species the unit of evolutionary change, it is rather the result of selection that takes place at the lower levels of biological hierarchy. It is this property that puts a major constraint on the status of what species can have. The selection at lower levels of the biological hierarchy is not possible if there is no spatiotemporal continuity and contact between members of the species, because selection is the consequence of differential survival and reproduction of members of a certain species. This means that the species must necessarily be an integrated and continuous spatiotemporal genealogical lineage of organisms, which are its constituent elements. That is a necessary property of an individual, not of a class.

Hull points out that this is a necessary but not sufficient condition of conceiving a species as an individual. Without additional requirements, all genes, organisms and species would form one individual because all organisms from the last known common ancestor until today form an integrated and continuous spatiotemporal genealogical lineage. Additional

requirement – which Hull uses to narrow down the ontological status of an individual to the level of the species – is an evolutionary unity.[2] Hull does not explain precisely enough the notion of evolutionary unity, but it could be said that evolutionary unity is "something" that differentiates a species as a particular individual from a continuous genealogical lineage since the last known common ancestor until today and from other species. Hull says that the evolutionary unity of a species is being maintained by internal and external mechanisms. The internal mechanisms are the gene flow and homeostasis, while the environment and the specific ecological niche make up the external mechanisms. In order to ascribe the ontological status of an individual to a species, it must have all these mechanisms,[3] plus it needs to fulfill the conditions mentioned earlier.

According to the first internal mechanism, two populations make one individual if there is gene flow between them at least occasionally – in the evolutionary conception of time. If two populations are long isolated, an additional criterion is required in order for the status of the individual to be ascribed to them, such as a potential breeding, which is in itself problematic since there are good species in nature that form stable hybrid zones but do not form one species. Because of that, ontological status of an individual would not be ascribed to them.

Second internal mechanism – homeostasis – Hull takes over from Eldredge and Gould:

---

[2] In papers titled "Are Species Really Individuals?", "Individuality and Selection" and "Matter of Individuality", Hull uses three concepts that are related to the same condition of individuality: cohesion, coherence and evolutionary unity (see Hull 1976; 1980; 1992). I opted for the concept of evolutionary unity because – despite the fact that it is as imprecise as the other two concepts – it is the least imprecise of all three and that is why it seems the best in pointing out to the condition of individuality for species that Hull is trying to add in order to narrow it down to the species level.

[3] Hull nowhere explicitly stated the aforementioned claim, but it seems that this conclusion can be drawn for two reasons. The first reason I have already explained, and it refers to the fact that without these mechanisms the notion of the individual has too much scope. Another reason we can see from the claims of the representatives of the SAI view when they exclude organisms that reproduce asexually from the species status, which will be discussed later in the paper.

> The answer probably lies in a view of species and individuals as homeostatic systems – as amazingly well-buffered to resist change and maintain stability in the face of disturbing influences […] Lerner (1954, 6) recognizes two types of homeostasis, mediated in both cases, he believes, by the generally higher fitness of heterozygous vs. homozygous genotypes: (1) ontogenetic self-regulation of populations (developmental homeostasis) […] and (2) self-regulation of populations (genetic homeostasis) "based on natural selection favoring intermediate rather than extreme phenotypes". In this view, the importance of peripheral isolates lies in their small size and the alien environment beyond the species border that they inhabit – for only here are selective pressures strong enough and the inertia of large numbers sufficiently reduced to produce the "genetic revolution" (Mayr 1963, 533) that overcomes homeostasis. The coherence of a species, therefore, is not maintained by interaction among its members (gene flow). It emerges, rather, as an historical consequence of the species' origin as a peripherally isolated population that acquired its own powerful homeostatic system. (Eldredge & Gould 1972, 114)

According to Eldredge and Gould, species keep their evolutionary unity in the following way. After peripatric speciation, in the new environment, they create a new balance due to which they undergo small evolutionary changes and survive for as long as they can maintain this balance.

Hull explains that the environment influences an evolutionary unity of a species in a way that all the members of a certain species are affected by the same selection pressures. This implies that all members of a certain species will go through the same or similar evolutionary changes. For example, imagine that all members of the species X are under the selection pressure because of which taller members of the species have more offspring. If selection pressure worked in this way and if it would last long enough, the average height of members of species X at the time t_1 would be lower than the average height of the members of species X at the time t_3.

The second external mechanism that affects the evolutionary unity of a species is an ecological niche. Hull explains ecological niche as a "relation between a particular species and key environmental variables" (Hull 1992, 300). This means that different species in a combination with the same

environmental factors will occupy different ecological niches, and that is an important integration factor for Hull.

It is necessary to clarify one more condition of individuality and that is spatiotemporal continuity. Ghiselin (1992) points out that the notion of an individual in logic refers to a single object at any level of integration. This means that for an object to be an individual, it is not necessary for it to be physically continuous. Ghiselin (1992) explains it by using the following example: The United States are an individual, regardless of the fact that they are physically discontinuous. Between Alaska and the rest of the territory of the United States there is Canada's territory as well as the international waters. Therefore, we can say that for a certain organism or a population it is not important that they are spatiotemporally and physical continuous in order for them to be individuals.

Before going any further to the implications arising from the SAI view, it would be useful to make a summary of all the criteria for the species' individuality. The species is a theoretical concept in the context of evolutionary theory, which implies that species are historical entities and continuous genealogical lineages, which means that they have spatiotemporal location and continuity. The criterion by which one species is separated from other species is an evolutionary unity which, according to Hull, consists of internal mechanisms – gene flow and homeostasis – and external mechanisms, the environment (selection pressures) and ecological niche.

From the SAI point of view, a number of important consequences is entailed. The first consequence is that species can evolve: "If species were not individuals, they could not evolve. Indeed, they could not do anything whatsoever" (Ghiselin 1992, 364). Species, except that they evolve, "they speciate […] they provide their component organisms with genetical resources, and they become extinct" (Ghiselin 1992, 377). Hull and Ghiselin add that species even compete with other species, but point out that it is not as important as competition between members of the same species.

Another consequence of the SAI position is that whether or not an organism is a member of a species is not determined on the basis of the characteristics that an organism possesses, but rather based on the necessary and sufficient conditions. Organism is a member of a certain species if it

belongs to a certain genealogical lineage that meets the criteria of individuality. This entails two consequences: there are no laws about species and the names of species are personal names. "They are meaningless identification tags and nothing else" (Hull 1976, 174). When we use the term *Tatooine* we think of the fictional desert planet in a binary star system. This is not its definition but a description and the name is only a reference to that description. It is the same with species, when we say *Homo sapiens*, we do not refer to two-legged rational animals with a little hair on their bodies. *Homo sapiens* is only a label for a specific group of organisms. People may be two-legged rational animals with a little hair on their bodies as a rule, but we will not say for people without hair on their bodies that they are not people. Hull concludes that – if a species membership is not determined by necessary and sufficient conditions – then there is no human nature. Even if there are characteristics that are common to all and only humans, it would only be a temporary condition that can easily change with evolutionary changes in the future. Thus, individuals can only be described and that description will be temporary and subject to change as described individuals go through evolutionary changes.

Given that individual organisms belong to a certain species if they belong to a certain genealogical lineage and "since they are derived from and contribute to a single gene pool" (cf. Hull 1980, 328), they form parts of the species, and not members of the species. For clarity, it is useful to make an analogy with individual organisms. Different organs of an individual organism are not its members but parts, because they form a single, integrated and spatiotemporally continuous whole that changes as its parts are changing. Classes have members and the change of their members does not affect the determination of the class.

The next consequence of the SAI view is that organisms that reproduce asexually do not form species. Ghiselin makes a comparison to the economy and says that the attribution of the status of the species to the organisms that reproduce asexually would be like starting to create imaginary companies for the self-employed. There are three main reasons why Hull and Ghiselin believe that organisms that reproduce asexually do not form species. Hull says that organisms that reproduce asexually as well as taxa of a higher rank in the biological hierarchy, do not meet all the criteria for evolutionary unity. Organisms that reproduce asexually evolve, they have spatiotemporal continuity and location which makes them historical

entities and forms them into continuous genealogical lineages, they are exposed to selection pressures, they do fill certain ecological niches, but they completely "lack any intrinsic mechanisms for promoting evolutionary unity" such as gene flow and homeostasis (cf. Hull 1976, 183-184). Even if organisms that reproduce asexually had enough evolutionary unity, it should be based on the external mechanisms, and Hull doubts that external mechanisms would be effective for this task. Ghiselin states another reason. He believes that species are individuals which have to evolve separately from each other, and this is possible because they form separate reproductive units and because they are the result of a speciation process. It is clear that species that reproduce asexually do not meet the above criteria mentioned by Ghiselin. Species that reproduce asexually do not constitute separate reproductive units because they do not reproduce sexually. Therefore, they cannot evolve separately from one another, which implies that they cannot form species and therefore cannot be individuals. Hull adds that organisms that reproduce asexually cannot constitute entities of higher levels than those of genealogical lineages because they lack internal mechanisms of evolutionary unity and because they evolve only by processes of replication and interaction. So, the genealogical lineages are the peak of integration that organisms that reproduce asexually can achieve. Genealogical lineages are species of organisms that reproduce asexually, concludes Hull.

Hull and Ghiselin point out that the SAI view entails a stance in the debate on the problem of universals in the species problem. The entailed view is realism. The reason is simple; individuals are concrete objects that really exist. Ghiselin again draws an analogy with economy and says that species are as real as are the companies such as Diamondback or Textile House. We have also seen earlier in the paper that classes cannot evolve because they are abstract objects. The fact that species evolve implies that species have spatiotemporal continuity that is a necessary precondition of evolution and the basic characteristic of an individual. "Now that species are conceived of as individuals, they have to be absolutely concrete, and must be viewed as no more than intellectual constructs organisms are" (Ghiselin 1992, 366).

Another consequence of SAI view is that when a species dies out, it is forever. Hull presents two arguments in support of his claim. The first argument is derived from two basic properties of individuality, and those are

location and spatiotemporal continuity. Each species has its beginning in a certain period of time, at a certain location and its end. This makes it spatiotemporally unique. Once a species becomes extinct, the same species cannot reoccur. Even if we assume that in the future a species will appear that will have all the characteristics identical to the species that is now extinct, it would still be a new species. The difference would be in the spatiotemporal location of the new species. It is the same with organisms. Once an organism dies, the same organism can no longer be recovered. Even if an organism would appear that would be identical in every conceivable characteristic to the organisms that died, it would still be the new and different, spatiotemporally unique organism.

2. Species as classes view

The basic claim made by representatives of this position is very simple. They believe that species are classes because all members of the species possess some properties that are essential (Kitts & Kitts 1979; Kitcher 1992; Devitt 2008; Putnam 1975). For starters, it is necessary to define the notion of a class. For clarity and consistency, in the rest of the paper I will use the term class in the same way as Stamos: "[…] I shall use the term "class" for intensionally defined (therefore abstract) objects […] the members of the class must have common (nontrivial) properties" (Stamos 2003, 21). Common non-trivial properties of members of a certain class are also called "essential properties".

The simplest formulation of the position that species are classes was formulated by Putnam: "Lemon: natural kind word […] associated characteristics: yellow peel, tart taste, etc." (Putnam 1975, 144). All members of the species F have at least one essential property P.

When can properties be regarded as essential properties, and what makes an essence of a certain species, according to essentialism? Devitt explains:

> A property *P* is *an essential property* of being an *F iff* anything is an *F* partly in virtue of having *P*. A property *P* is *the essence* of being an *F iff* anything is an *F* in virtue of having *P*. The essence of being *F* is the sum of its essential properties. (Devitt 2008, 345)

According to Stamos, this way of defining classes and conditions of membership in the classes entails that the class is defined only as the membership conditions. He cites an example: "If x is an atom with seventy-nine protons in its nucleus, then x is an atom of gold" (Stamos 2003, 173). All organisms with the property P belong to the species F which is in perfect analogy with the example with the atom of gold. If this is true, this way of defining the species entails certain implications.

First implication is that species are abstract entities because membership conditions are abstract entities as well. The existence of the membership conditions is completely independent from the fact whether members of a certain class exist or not. It is clear that if we determine that "All organisms with the property P are members of the species F" that there may be a circumstance in which organisms with the property P do not exist. In this case, because species is an abstract entity, we can't conclude that species F does not exist, but only that it does not have any members.

Second implication is that classes defined in this way remain unchanged with changes in the number of its members, because the changes in the membership do not change the membership conditions, which in this case is the class (Stamos 2003, 172-173).

One version of essentialism of interest for the species problem is biological essentialism. This position is specific in that it claims that the necessary properties of species are genetic, as argued by Caplan (1980; 1981), Kitts & Kitts (1979) and Devitt (2008).

However, even this claim is disputed by some representatives of biological essentialism. Devitt and Caplan allow the existence of extrinsic necessary properties in addition to intrinsic necessary properties. According to Caplan, extrinsic necessary properties of a species are the ability to obtain a fertile offspring between group members and the origin from a common ancestor. In addition, Caplan thinks that the claim that species are classes does not imply that essential properties of species are eternal and unchanging. Species are after all entities that arise from evolutionary processes. When organisms do not manifest essential properties of a particular species, it is reasonable to assume that this class has gone extinct or has evolved to a different class. According to Caplan, species are classes that are subject to evolutionary processes and their consequences (see Caplan 1980, 74-75). Devitt, similar to Caplan, points out that it is specific for species, next to their intrinsic genetic properties, that they are also historical

entities and that their members are a part of the genealogical nexus (see Devitt 2008, 368). How can species be classes and have essential properties while undergoing evolutionary changes at the same time? This is explained by Devitt:

> Suppose that S1 and S2 are distinct species, on everyone's view of species, and that S2 evolved from S1 by natural selection. Essentialism requires that there be an intrinsic essence G1 for S1 and G2 for S2. G1 and G2 will be different but will have a lot in common. (Devitt 2008, 372)

The process of gradual evolutionary change, which would be compatible with essentialism, would proceed as follows: from S1 a group of organisms separates and under the circumstances between that group and the rest of the species S1 gene flow is interrupted. At this point in a group that separated, G1 is still its intrinsic essence. Suppose that the separate group is exposed to different selection pressures than species S1. Their essential intrinsic properties will slowly begin to change and move away more and more from G1 and approach more and more to G2, while this process is not completed. The end result of this process will be species S1 with G1 essential properties and species S2 with G2 essential properties. The process of gradual evolutionary change described in this way is compatible with the theory of evolution and is in accordance with basic tenets of essentialism. Let us remember that essentialism does not require that species must have eternal and unchanging essential properties. According to Devitt, species are a special type of classes that participate in the evolutionary processes and have no eternal and unchangeable essential properties. The way that species evolve and that they are classes at the same time, implies that species change classes as they evolve. In the illustrated example, species S2 with essential properties G2 evolved from species S1 with the essential properties G1. As species S1 went through evolutionary process, it gradually changed its essence from G1 to G2. In that way, organisms that at the beginning of the evolutionary process belonged to species S1 with essential properties G1, eventually become species S2 with essential properties G2 and thereby changed the class to which they belong.

3. Species as individuals: just another class view of species

Representatives of the SAI view try to argue that species cannot be classes because classes cannot change. Species participate in the evolutionary processes and that *a priori* makes them entities that are going through changes all the time. In this part of the paper, I will argue that the SAI view nevertheless implies that species have the ontological status of classes, although the representatives of the SAI view claim otherwise.

To begin with, let me remind the reader of the main tenets of the view that species are classes. First, all members of the species must possess some properties that are essential. As explained by Devitt:

> A property *P* is *an essential property* of being an *F iff* anything is an *F* partly in virtue of having *P*. A property *P* is *the essence* of being an *F iff* anything is an *F* in virtue of having *P*. The essence of being *F* is the sum of its essential properties. (Devitt 2008, 345)

What is important for the view that species are classes is that properties that are common to all members of the class must be nontrivial.

Second, representatives of the position that species are classes allow for the possibility that necessary properties can be extrinsic properties and not necessarily intrinsic. Caplan (1980) gives two examples of extrinsic necessary properties: the ability to obtain fertile offspring between group members and the origin from a common ancestor. Devitt (2008) expands the list of extrinsic necessary properties further: species are also historical entities and their members are a part of the genealogical nexus.

In addition, both Caplan (1980) and Devitt (2008) argue that the view that species are classes does not imply that essential properties of species are eternal and unchanging. Caplan (1980) states that species are classes that are subject to evolutionary processes and their consequences while Devitt (2008) explains how species can have the ontological status of a class and yet undergo evolutionary changes at the same time:

> Suppose that S1 and S2 are distinct species, on everyone's view of species, and that S2 evolved from S1 by natural selection. Essentialism requires that there be an intrinsic essence G1 for S1 and G2 for S2. G1 and G2 will be different but will have a lot in common. (Devitt 2008, 372)

With this line of argumentation, Caplan and Devitt intercept the arguments of the representatives of the SAI view according to which: a) species cannot be classes because classes cannot change, and b) their participation in the evolutionary processes *a priori* makes them the entities that are going through changes all the time, which excludes them from the ontological status of a class. Caplan and Devitt intercept the mentioned arguments of the representatives of the SAI view because they offer us a plausible interpretation of how it is possible for species to be classes that change by undergoing evolutionary processes as described in the second part of this paper.

To resume my discussion, the basic position of the representatives of the SAI view implies that species *necessarily* have to be integrated and that continuous spatiotemporal genealogical lineages of organisms and organisms are their constituent elements. The additional requirement of individuality of species is evolutionary unity maintained by internal and external mechanisms. It is evolutionary unity that separates individual species out of the continuous genealogical lineage as separate entities. Without evolutionary unity, it would not be possible to identify individual species, which makes it the second *necessary* condition of specieshood. The third condition for the individuality of species is that all members of a species are going through the same or similar evolutionary changes. This implies that if we have a group of organisms whose members do not go through the same or similar evolutionary changes, they do not belong to the same species, and if so, we have just reached the third *necessary* condition of specieshood. The last *necessary* condition of individuality of species is that species must consist of organisms that reproduce sexually, because organisms that reproduce asexually cannot form a species.

According to SAI view, species are individuals and that entails the following necessary conditions that groups of organisms must fulfill in order to be treated as species:

1. integrated and continuous spatiotemporal genealogical lineages of organisms that are their constituent elements;
2. separated from the continuous genealogical lineage from the last known common ancestor to modern organisms by evolutionary unity;

3. made up of organisms that go through the same or similar evolutionary processes;
4. groups of organisms whose members reproduce sexually.

These are the essential properties of species according to the SAI view. In accordance with the basic tenets of the view that species are classes, the SAI view also implies that for each entity to which "X is a species" applies, it is true that X must necessarily possess all four just mentioned properties.

It should be noted that these properties are not intrinsic properties of individual organisms but they are extrinsic properties of the species. However, as I have shown in the second part of the paper, Caplan and Devitt, both representatives of the view that species are classes, allow for the existence of extrinsic necessary properties for species and they explicitly state them as follows:

1. The ability to obtain a fertile offspring between group members.
2. The origin of a species from a common ancestor.
3. Species are historical entities.
4. Species members are a part of the same genealogical nexus.

If we compare the conditions that groups of organisms must fulfill in order to be treated as species according to the SAI view and extrinsic essential properties of species listed by Caplan and Devitt, it should be clear that they are fundamentally the same, although the formulation made by the representatives of the SAI view is slightly more detailed. If this is true, it is only possible to conclude that, according to the SAI view, species are in fact classes.

Do organisms in the SAI view have some intrinsic essential properties? I think that they must have at least one intrinsic essential property. If one of the necessary conditions that groups of organisms must fulfill in order to be treated as a species is that members of the species must reproduce sexually, the consequence on the level of individual organisms in that species would be that they must have some species *specific mate recognition system* (or SMRS) which is possessed by all and only members of the species. According to Paterson, SMRS is a group of adaptations – specific to each species and in turn to all and only members of a specific species –

which is being used during courtship and reproduction among potential partners. SMRS evolved as the adaptation under the influence of specific selection pressures when the incipient species detached from the ancestral species (Paterson 1992). It is important to note that the way SMRS evolved is a consequence of the fact that the group of organisms in question is: a) integrated and continuous spatiotemporal genealogical lineage, which b) is separated from the continuous genealogical lineage from the last known common ancestor to modern organisms by evolutionary unity. In this case, evolutionary unity consists of a species-specific selection pressures that shaped species SMRS. That makes SMRS an intrinsic essential property that each individual organism in a species must possess since one of the extrinsic essential properties of species is that a species is a group of organisms whose members reproduce sexually. Although the SAI view is relatively new, it does not seem to be a revolutionary position that will fundamentally change our understanding of species. It is only an interesting new version of the view that species have an ontological status of a class, the very thing that representatives of the SAI position wanted to avoid.

4. Conclusion

Representatives of the SAI view try to argue that species cannot be classes because classes cannot change, and since species are subject to evolutionary changes, they undergo changes all the time. According to the view that species are individuals, species are treated as:

1. integrated and continuous spatiotemporal genealogical lineages of organisms that are their constituent elements;
2. separated from the continuous genealogical lineage from the last known common ancestor to modern organisms by evolutionary unity;
3. made up of organisms that go through the same or similar evolutionary processes;
4. groups of organisms whose members reproduce sexually.

I argued that these are necessary conditions for groups of organisms to be treated as species, according to the representatives of the SAI view.

When these conditions are compared to the list of extrinsic essential properties made by Caplan and Devitt, it is clear that they are the same. If these are indeed necessary conditions for specieshood, then the conclusion that species are classes in the SAI view is unavoidable. I have also argued that, if under the SAI view one of the necessary conditions that groups of organisms must fulfill in order to be treated as species is that members of the species must reproduce sexually, then each member of the species must possess the same SMRS, which in turn makes SMRS an intrinsic essential property of each member of the species. This makes SAI view a new version of the old position that species have an ontological status of a class, not a revolutionary new position that will fundamentally change our understanding of species.

References

CAPLAN, A. L. (1980): Have Species Become Declasse? *Philosophy of Science Association* 1, 71-82.

CAPLAN, A. L. (1981): Back to Class: A Note on the Ontology of Species. *Philosophy of Science* 48, 130-140.

DEVITT, M. (2008): Resurrecting Biological Essentialism. *Philosophy of Science* 75, 344-382.

ERESHEFSKY, M. (2010): Species. In: Zalta, E. N. (ed.): *The Stanford Encyclopedia of Philosophy* (Spring 2010 Edition), available at: http://plato.stanford.edu/entries/species/

ELDREDGE, N. & GOULD, S. J. (1972): Punctuated Equilibria: An Alternative to Phyletic Gradualism. In: Schopf, T. J. M. (ed.): *Models in Paleobiology.* San Francisco: Freeman, Cooper and Company, 82-115, available at: http://www.somosbacteriasyvirus.com/phyletic.pdf.

GHISELIN, M. (1966): On Psychologism in the Logic of Taxonomic Controversies. *Systematic Zoology* 15, No. 3, 207-215.

GHISELIN, M. (1992): Species Concepts, Individuality, and Objectivity. In: Ereshefsky, M. (ed.): *The Units of Evolution*. Cambridge: MIT Press, 363-380.

HULL, D. L. (1976): Are Species Really Individuals? *Systematic Zoology* 25, 174-191.

HULL, D. L. (1980): Individuality and Selection. *Annual Review of Ecology and Systematics* 11, 311-332.

HULL, D. L. (1992): A Matter of Individuality. In: Ereshefsky, M. (ed.): *The Units of Evolution*. Cambridge: MIT Press, 293-316.

KITCHER, P. (1992): Species. In: Ereshefsky, M. (ed.): *The Units of Evolution*. Cambridge: MIT Press, 317-342.

KITTS, D. B. & KITTS, D. J. (1979): Biological Species as Natural Kinds. *Philosophy of Science* 46, 613-622.

PATERSON, H. (1992): The Recognition Concept of Species. In: Ereshefsky, M. (ed.): *The Units of Evolution*. Cambridge: MIT Press, 139-158.

PUTNAM, H. (1975): *Mind, Language and Reality*, Cambridge: Cambridge University Press.

STAMOS, D. N. (2003): *The Species Problem: Biological Species, Ontology, and the Methaphysics of Biology*. New York: Lexington Books.

WILKINS, J. S. (2009): *Species: A History of the Idea*. Berkeley, Los Angeles: University of California Press.

Organon F 24 (1) 2017: 54-84

Toward a Demarcation of Forms of Determinism[1]

VLADIMÍR MARKO

ABSTRACT: In the current philosophical literature, determinism is rarely defined explicitly. This paper attempts to show that there are in fact many forms of determinism, most of which are familiar, and that these can be differentiated according to their particular components. Recognizing the composite character of determinism is thus central to demarcating its various forms.

KEYWORDS: Determinism – fatalism – logical determinism – scientific determinism – logical fatalism.

Determinism is a basic philosophical concept. It is usually assumed that both the term "determinism" and determinism as a philosophical conception or theory are clear and obvious. In the literature, however, the precise contours of determinism are not explicitly defined – an obscurity that often leads to inconsistencies and misunderstandings.

In this article, I put to the side questions concerning the soundness or adequacy of the philosophical views I shall consider. Instead, I am interested only in the basic conceptual contours of different kinds of determinism and whether it is possible to sort them into some kind of interrelated order for the purposes of better demarcating varieties of determinism. My

[1] Received: 24 October 2016 / Accepted: 2 January 2017

✉ Vladimír Marko
Department of Logic and Methodology of Sciences, Comenius University in Bratislava
Gondova 2, 814 99 Bratislava, Slovak Republic
e-mail: vladimir.marko@uniba.sk

main thesis consists of three claims: (i) there are *many forms* of determinism; (ii) each form of determinism, as a philosophical conception, has a *composite character*; and (iii) conceptions of determinism can be *differentiated* according to the particular elements used in their composition.

In what follows it is suggested that a) similar (or even the same) conceptions of determinism may be designated by different names; and that b) various formulations labeled with the same name represent substantially different conceptions. Below I discuss different deterministic conceptions and emphasize some of their essential components that enable us to make their distinct features more vivid.

1. The origin of the term

When philosophers wish to label a certain philosophical conception "deterministic", they do not usually feel the need to additionally clarify or explicate what they mean. They simply take our understanding for granted. The meaning of this term varies, however, in both historical and contemporary texts.

"Determinism" has its origins in Latin. In Roman authors, we encounter use of *"determino"* or "*determinatio*", which means "to enclose within boundaries, to bound; to limit, to prescribe, to determine; to fix, to settle".[2] The Greek equivalent of the Latin *determinare* (syn. *definire*) is ἀφωρισμένης, which was used in approximately the same way. Ancient and medieval uses of these terms departed greatly from our use today.

Although philosophical conceptions of determinism have had their advocates throughout history, the specific term "determinism" arrived on the scene much more recently. A survey of Krug's *Allgemeines Handwörterbuch* reveals various uses of "*Determinismus*" *(Bestimmung*, *Predeterminismus*) and "*die Deterministen*" (see Krug 1827, vol. i, 500-501). Indeed, it contains a note on the first appearance of these terms (cf. Krug

[2] Livy uses this as a technical term to describe the augurs' division of the parts of heaven into regions (*determinavit regiones*) and for marking their boundaries [Liv. *Ab urbe condita libri*, i, 18, 7, 32]. A similar example can be found in Gellius [*Att. n.* 13, 14]. In Cicero [*Inv.* 1, 52, 98], "the conclusion [i.e. *peroration*] brings to a close and delimits the whole oration (*determinatio totius orationis*)".

1829, vol. v, 100), made by Christian Wilhelm Snell (1789) in a commentary on Kant's moral themes in *Über Determinismus und moralische Freiheit*. At several other places in the *Allgemeines Handwörterbuch*, determinism is used in the sense of "philosophical necessity". This use is related to an English source: Joseph Presley's (1799; 1780) concept of "determination"; and Krug quotes John Presley's correspondence with John Palmer (cf. Krug 1827, vol. iii, 128, 299, 303). A year after Snell (in 1790), Carl Friedrich Bahrdt in (1790, 291) also employs determinism as a theoretical concept. Soon after, the term appears in Kant's treatise on religion (see Kant 1793). In a footnote, Kant considers determinism in the context of the opposition between agency and determination by external forces. Here, it is described as *predeterminism*, and it is ultimately rejected as an "illusion" (cf. Kant 1793, 58A). That same year, Heydenreich published his *Über Freiheit und Determinismus* (see Heydenreich 1793).

Herbart uses the term once at the end of his text on Pestalozzi (1804, 281) and several times subsequently (cf. Herbart 1842). He claims that determinism is a prerequisite for action: "Determinismus ist Voraussetzung des Handelns" (Herbart 1843, 147; 152). Hegel (1816, ii, 206; 236) treated the term as a standard philosophical notion (in the context of mechanical processes, but also with respect to religion and freedom). An extensive record of the term's use in German can be found (with minor shortcomings) in the *Deutsches Fremdwörterbuch* (see Schulz et al. 1999, 442-443). Until the second part of the nineteenth century, the term was regularly used in the context of free will and its determination by antecedent circumstances, which were usually conceived as "external causes" that determine agents' decisions in the traditional sense of a "causa finalis".

2. Early demarcations

In the opening pages of his *Determinism and Indeterminism in Modern Physics*, Ernst Cassirer dates the rebirth of determinism to 1872, the year in which Emil Du Bois-Reymond (1886, 107) held a public speech on the limits of our knowledge of nature (see Cassirer 1956). Du Bois-Reymond reflected on the Laplacean roots of the notion and attempted to revive a genuine philosophical conception of determinism in the sense of complete

physical causal determinism. In fact, he simply repeated the formulation from the key passage of *Essai philosophique sur les probabilités*. Laplace's determinism, based on the principle of universal causal concatenation, was inspired by Leibniz's principle of sufficient reason. In a famous passage, Laplace writes the following:

> We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes. (Laplace 1902, 4)

This form of determinism identifies causation and lawfulness with determinism. Laplace wants to say that predictability (p) must at least in principle be grounded in the following postulation: There exists an intellect of some sort (i) that has access to and is able to analyze all relevant data (d) – where (d) consists of information about all forces (l) and all states (the position of all items at time t) in the system (s) – and that can bring this data under a single formula (f). In short, predictability (p) on this view is the result of the ability to apply a unique function (calculability) to the relevant data. In particular, Laplace highlights the following conditions for predictability: $p=\langle i, d, f\rangle$, where (d) consists of subset $\langle l, s, t\rangle$.

Laplace's determinism is a philosophical conception, built from different components. Central to it is the idea of a system governed by causation, which in turn proceeds according to laws. In addition, it relies on the notion of exceptional abilities (to obtain, analyze and calculate data via the application of a function, on the basis of which to make predictions). The data consists of laws, states and time indices, where laws are understood as active forces that are able to cause occurrences.

Cassirer (1956) pursued a different option and sought to distinguish between a new "critical" form of determinism and the old "metaphysical" determinism. The former is based on the belief that causal relations and laws originate in the mind – i.e. their source is our experience, not nature itself. Natural laws apply not to objective things, as metaphysical determinism

conceives them, but to cognitions and their ordering. In this sense, causal relations have a necessarily epistemological foundation.[3]

At the beginning of the twentieth century, William James (1907), motivated to make space for free will, also attempted to demarcate contemporary conceptions of determinism. He identifies the old determinism with the following view:

> [P]arts of the universe already laid down absolutely appoint and decree what the other parts shall be … Any other future complement than the one fixed from eternity is impossible. The whole is in each and every part, and welds it with the rest into an absolute unity, an iron block, in which there can be no equivocation or shadow of turning. (James 1907, 150)

He calls this "old" conception "hard determinism". Hard determinism does "not shrink from words like fatality, bondage of the will, necessitation, and the like." This contrasts with the "new" determinism – "soft determinism" – according to which the following is true:

> Nowadays, we have a soft determinism which abhors harsh words, and, repudiating fatality, necessity, and even predetermination, says that its real name is freedom; for freedom is only necessity understood, and bondage to the highest is identical with true freedom. (James 1907, 149)

3. Scientific determinism and scientific fatalism

Susanne Langer associates scientific forms of determinism with fatalism (see Langer 1936, 474). Fatalism is seen as the outcome of a full-fledged determinism. Determinism, which is based on the assumption that every event has an immediate cause, is a useful thesis for scientific pur-

[3] Cassirer (1956, 114): "We find the essential significance of the causal relation, if interpreted in a critical rather than a metaphysical sense, to be that it contains a statement not immediately about things but about experience, by which and in virtue of which alone things, as objects of knowledge, can be given us. It expresses something about the content of empirical knowledge."

poses. Problems arise, however, when this thesis is connected with predictability, which leads to *scientific fatalism*. Modern scientific fatalism, according to Langer, is "the assumption that there is a theoretically knowable collection of causes for any act". The thesis is derivable, on this view, from the false assumption inherent in determinism (and illustrated via Laplace's demon) of the ability to obtain knowledge of the "total state of the universe". The assumption was thought by Russell & Whitehead (1910, 40) to involve an "illegitimate totality", since "a whole cannot be theoretically constructed"; because of this, the doctrine of determinism in its philosophical form was taken to constitute "a modern version of belief in Fate" (Langer 1936, 478). Scientific fatalism is the view that there is a theoretically knowable collection of causes for any act (see Langer 1936, 478). Although "pure" determinism and fatalism commonly posit a causal connection between the past and the future, such that the latter can be predicted on the basis of the former, they do not entail the predictability of the future, for causality does not necessary imply predictability. Even in the case of a completely causal universe, the unpredictability of human agency undermines general predictability, given the unknowability of human agency (cf. Langer 1936, 472).

Mario Bunge also views the idea "that causality is fatalistic" as mistaken and draws a distinction between scientific determinism and fatalistic determinism (going as far as to argue that the two are "incompatible"; see Bunge 1959, 101-102). His view on fatalism, causality and determinism differs slightly from Langer's, however. While causal determinism is a theory that is grounded in reason and argument and offers "the means for knowing, predicting, and consequently changing the course of events", fatalism assumes that a lawless, supernatural power (fate) drives our unknowable and inescapable destinies – a power that is above the law, works with unconditional necessity, and directs the course of events. Causality need not entail any such transcendental or supernatural agency. Moreover, it does not entail inevitability: causes can interfere with each other, background or hidden causes and conditions may obtain, human agency may intervene, and so on. Bunge inclines toward a conception known as agent-causation, according to which the presence of the elements listed above can result in different outcomes (which he interprets as a source of probability). Thus, "general determinism" need not be

viewed as holding unconditionally. It enables us to use our knowledge of laws to change or modify courses of events while leaving room for chance and freedom. In addition, Bunge firmly believes that statistical laws exclude determinism completely and are indeed incompatible with it since they are based not on causal principles but on probability and generalized correlations obtained from data. As he puts this idea, "statistical law and probability destroys determinism".

4. Determinism in terms of predictability

Karl Popper, who prefers to interpret determinism as an epistemological thesis, sums up the doctrine of scientific determinism ("the doctrine much stronger than common sense") in his *Open Universe*. On his view, scientific determinism is a view with which "most physicists would have agreed at least prior to 1927" (Popper 1982, xx). According to scientific determinism, "the structure of the world is such that *any event can be rationally predicted, with any desired degree of precision, if we are given a sufficiently precise description of past events, together with all the laws of nature*" (Popper 1982, 1-2).

On Popper's account, scientific determinism has its roots in "religious determinism" and seems to be "a kind of translation of religious determinism into naturalistic and rationalistic terms" (Popper 1982, 6). It is contrasted with the metaphysical doctrine of determinism, which holds that

> [A]ll events in this world are fixed, or unalterable, or predetermined. It does not assert that they are known to anybody, or predictable by scientific means. But it asserts that the future is as little changeable as is the past. Everybody knows what we mean when we say that the past cannot be changed. It is in precisely the same sense that the future cannot be changed, according to metaphysical determinism. (Popper 1982, 7)

Metaphysical determinism differs from scientific determinism. It is entailed by both religious and scientific determinism. However, metaphysical determinism (along with metaphysical indeterminism) is not testable, since

it lacks empirical content. With regards to testability, another distinction drawn by Popper is that between a weak version of scientific determinism and its stronger form (cf. Popper 1982, 36*ff*).

The weak version presupposes the possibility of predicting any future instant of time in a closed physical system ("even from within") "by deducing the prediction from theories in conjunction with initial conditions" (i.e. with knowable initial conditions). Theories here play the instrumental role of describing the world – asserting that it has certain properties. However, this does not mean that, if the theory that describes certain properties of the world is true, everything that can be deduced from it has a corresponding property in the world. This "stronger" kind of determinism, criticized by Popper as false, subscribes to the predictability of "any given state, *whether or not the system in question will ever be in this state*" (Popper 1982, 37). Popper is not always consistent. This part of his book seems to identify determinism with causation (cf. Popper 1982, 149), while in other places he asserts that they are different (cf. Popper 1982, 4, 19, 23). Although predictability contributes to the testability of scientific theories, Popper is critical of metaphysical and stronger forms of determinism.

Indeed, Popper is not alone in criticizing the conflation of determinism and predictability (see, e.g., Earman 1986, 9-10; Suppes 1993; Kellert 1993; and Stone 1989). Predictability, which is just one component of (Laplacean) determinism, is an epistemological concept; determinism, on the other hand, should be analyzed as an *ontic* or *physical* thesis. Thus, it is necessary to distinguish determinism proper from determinism in the sense of the ability to make predictions. Patrick Suppes appeals to the three-body problem and Turing machine examples: both are illustrations of deterministic systems *par excellence*. As is well known, there is no algorithm (which could allow for prediction) for determining whether an arbitrary Turing machine in an arbitrary configuration will ever halt (see Suppes 1993, 245-246). Suppes therefore insists on the *conceptual separation of two notions*: predictability and determinism. We have good reason to interpret certain systems as "deterministic" even though we may not be able to predict events occurring within it, which would suggest that determinism need not come hand in hand with the predictability thesis.

5. The rise of so-called "syntactical determinism"

Russell joins the discussion in his well-known lecture on the obscurity of the concept of a cause (see Russell 1917). On his view, the concept of determinism can be demystified by revealing its true nature – its standing as *a functional relation*:

> A system is said to be "deterministic" when, giving certain data, e_1, e_2, …, e_n at times t_1 t_2,…, t_n respectively [*viz.* "determinants"], concerning this system, if E_t is the state of the system at any time t, there is a functional relation of the form $E_1 = f(e_1, t_1, e_2, t_2, …, e_n, t_n)$.
>
> The system will be "deterministic throughout the given period" if t, in the above formula, may be any time within that period, though outside that period the formula may be no longer true. If the universe, as a whole, is such a system, determinism is true of the universe; if not, not. (Russell 1917, 199)

> Determinism in regard to the will … Whether this doctrine is true or false, is a mere question of fact; no *a priori* considerations (…) can exist on either side. (Russell 1917, 205).

> We were unable to find any *a priori* category involved: the existence of scientific laws appeared as a purely empirical fact, not necessarily universal, except in a trivial and scientifically useless form. (Russell 1917, 208)

Russell thus insists on revising the concepts of cause and necessity – two fundamental elements of what had been the dominant approach to science. Since "there is no *a priori* category of causality" – but merely certain observed uniformities (cf. Russell 1917, 205) – the notion of necessity is "a confused notion not legitimately deducible from determinism" (Russell 1917, 207); it must be viewed simply as a logical necessity driven by constitutive determinants as arguments of a necessary propositional function.

Although Russell turns away from the confused notions of cause and causality, his formulation leaves room for a connection between determinism and predictability: a system is deterministic if its previous states determine its later states in the precise sense in which the arguments of a function determine its values. One of Russell's important suggestions (which

concerns the traditional view that scientific laws hold absolutely) reflects "the principle of the irrelevance of time": the fact that time, in an absolute sense, cannot enter into our formulae (see Russell 1917, 205). Our laws are not *a priori* principles that, because they have held for the past and present, will necessarily hold in the future. Our formulas are "methodological precepts", not Laws of Nature that hold absolutely and eternally.

Russell was not completely satisfied with his formulation of determinism. There are several reasons for this. For any set of data points that are describable by some function, those points are in fact describable in other ways, by infinitely many functions. Further, in dynamic systems, the past state of a system, to which a formula was hitherto applicable, may well change in the future, and what is selected as the simplest way to capture the facts may therefore change as well. In addition, the way our system has hitherto been described may advance, such that formulas we have thus far relied on no longer apply. For this reason, we must bear in mind the principle of the irrelevance of time.[4]

Russell's revision seriously shook the traditional image of science held by the scientific community. As a result of his disruption, however, his critique of causation and natural laws made room for the concept of logical determinism.

6. Schlick's logical determinism

Russell's observations provided the basis for a developing conception, which Moritz Schlick would later call *logical determinism* (the first use of this term). Schlick outlines this position as follows:

> Let us see how the scientist uses the word determination – then we shall find out what he means by it. When he says that the state E at the time

[4] "In fact we might interpret the 'uniformity of nature' as meaning just this, that no scientific law involves the time as an argument, unless, of course, it is given in an integrated form, in which case *lapse* of time, though not absolute time, may appear in our formulas" (Russell 1917, 205). An extension of Russell's formula with regards to determinism in dynamical or evolutive systems is given in van Fraassen (1989, 254). Russell's function must be extended to cover all possible trajectories of the system, i.e. to encompass changes to successive states of the system.

> t_1 is determined by the state C at the time t_0, he means that his differential equations (his Laws) enable him to calculate E, if C and the boundary conditions are known to him. Determination therefore means Possibility of Calculation, *and nothing else*. (Schlick 1932, 114)

The "natural law" of science, however, "is not a prescription as to how something should behave, but a formula, a description of how something does in fact behave" (Schlick 1939, 147). Natural laws are just descriptions without any force, which do not make things to move according to their prescriptions.

The necessity of logical determinism is not the necessity of *causal nomological* determinism. It is the necessity of *functional* determination, which enables us, on the basis of a function and its determinants, to calculate the necessary relational dependencies among determinants with respect to the selected function.

Russell's and Schlick's formulations share a crucial assumption: namely that determinism is firmly linked with predictability (and, conversely, with the ability to make retrodictions). Schlick's "possibility of calculation" corresponds to Laplace's condition for making predictions (although Laplace had in mind a singular function over a complete universe). If one state of affairs is determined in the above functional sense, there is room for this state to be predicted (or to be calculated in advance) on the basis of knowledge of its previous states and the function that connects them to it.

Schlick's *calculability* (predictability) is a form of deducibility. It represents a standard understanding of logical determinism according to which one state is propositionally connected to another state via inferential power. Logical necessity must be distinguished from physical necessity and causation: "[W]hat is called causal necessity is absolutely different from logical necessity […] [F]ormer philosophers […] frequently made the mistake of confusing the two and believing that the effect could be logically inferred from the cause" (Schlick 1932, 108). Schlick calls the relationship between logical principles and reality "a problem of logical determinism" – or "a paradox" (see Schlick 1931, 159) – and locates its origin in Aristotle: "[T]he Principle of the Excluded Middle could not be applied to future events unless we assume the truth of Determinism." Most likely with Jan Łukasiewicz in mind, Schlick adds that "there are

even modern logicians who follow him [*viz.* Aristotle] in this" (Schlick 1932, 115).

7. The bivalent nature of logical determinism

Jan Łukasiewicz's formulation of determinism, developed more than a decade before Schlick's, runs as follows: "By determinism I understand the belief that if *A* is *b* at instant *t* it is true at any instant earlier than *t* that *A* is *b* at instant *t*" (Łukasiewicz 1990, 113). On this formulation, the future must be treated like the past; it differs from the past "only in so far as it has not yet come to pass". Everything is fixed in advance. The way out of determinism consists in abandoning the beliefs that lead to a conception of eternal truth and the absence of free will.

Łukasiewicz offers two arguments against determinism. One is based on "logical principles", while the other is based on "the principle of causality". Here, I wish only to emphasize his commitment to the view that the bivalent nature of propositional calculus leads to determinism. This argument relies on identifying two principles: the principle of bivalence and the law of excluded middle. Even though the argument is ostensibly valid (on the basis of propositional calculus), it must be rejected for other reasons: "Although this solution appears to be logically valid, I do not regard it as entirely satisfactory, for it does not satisfy all my intuitions" (Łukasiewicz 1990, 124).

The rejection of determinism "finds its justification both in life and in colloquial speech" (Łukasiewicz 1990, 125). The principle of bivalence is not applicable to future-oriented propositions that describe possible future (not yet generated) states of affair. Such propositions do not have "real correlates" like propositions about the present and the past. A third, "neutral" value would be more appropriate to future contingents, and intermediate sentences would "ontologically have possibility as their correlate".

The argument for determinism is logically valid. Thus, determinism is a self-consistent view; to the extent that it rests on the assumption of bivalence, however, not only is it unable to deal with future contingents but it also has unintuitive consequences with regards to human agency.

Friedrich Waismann uses the more expressive term "logical predestination" since, according to this conception, it seems "that indeed the entire

future is somehow fixed, logically preordained" (Waismann 1959, 352). Following Waismann, Zbigniew Jordan interprets logical determinism as the semantic formulation of strict determinism, "where the strict causal determinism implicitly assumes that an unending sequence of events has no limit" (Jordan 1963, 23). The principle of causality is not a necessary consequence of the principle of bivalence, but it provides a firm connection to real correlates, which secures the necessary truth of future propositions and justifies the thesis of eternal truth. In this sense, "strict determinism" is the outcome of (a) the principle of bivalence and two additional assumptions: (b) the correspondence theory of truth and (c) the timelessness or absolute character of truth (cf. Jordan 1963, 1). On Jordan's view, "strict determinism" occupies the following relative place in the transitive chain of principal dependence: "If the principle of bivalence entails strict determinism and strict determinism entails fatalism, the principle of bivalence entails fatalism" (Jordan 1963, 3). In the same spirit, Jan Wołenski has recently interpreted logical determinism as a form of radical determinism (see Wołenski 1996).

8. Inevitability

The transitivity chain traced by Jordan led to the standard representation of logical determinism as logical fatalism. This conception finds support in Aristotle's sea battle example and the case of future contingent propositions. Gilbert Ryle's lecture "It Was to Be" (see Ryle 1953), Richard Taylor's articles and the widespread discussion that followed during the sixties (cf. Wallace 2011), A. J. Ayer's "Fatalism" (see Ayer 1963) and Michael Dummett's "Bringing About the Past" (see Dummett 1964) are among the many texts on fatalism that have contributed to this tradition. Logical necessity began to be more frequently interpreted in terms of inevitability.

Even though Raymond D. Bradley warned against the confusion of logical determinism with fatalism in the late fifties, the tradition of interpreting logical determinism as fatalism (or at least a kind of fatalism) continues. In "Must the Future Be What It Is Going to Be?", Bradley restates some of Schlick's earlier warnings to the effect that logical necessity must be distinguished from causal necessity and that the truth of logical propositions and their relations has a different character than the

truth of empirical evidence (see Bradley 1959). He criticizes the common assumption that logical determinism *implies* (logical) fatalism. On his view, what is timeless differs from what is empirical. The failure in this inference consists in ascribing logical necessity to causal necessity, and causal necessity to fatalism. We can accept as valid that if *x* is causally determined, then *x* is logically determinate. However, *x*'s being logically determinate does not imply that *x* is causally determined. These two claims are not equivalent; one concerns causality and the other concerns logical necessity. Three logical principles that are to be found in Aristotle's discussion of the sea battle – the law of identity, the law of noncontradiction and the law of excluded middle – which form the crux of logical determinism, do not provide a sufficient basis for the projection of logical necessity onto causal necessity or the (actual) necessity of future truths.

9. Logical fatalism

The term "(logical) fatalism" (a view according to which time is symmetrical and all possible worlds are reduced to the actual world) has, over time, completely replaced the term "(logical) determinism". In his articulation of what is referred to as the standard argument for (logical) fatalism, for example, Taylor nowhere mentions determinism, logical or otherwise (cf. Taylor 1962). Interestingly, laws are not mentioned anywhere in the first version of the argument. Instead, he emphasizes causes. He later suggests that the only difference between the fatalist and the determinist is that the former explicitly holds that there is no difference between universal causation and inevitability. The distinction between fatalism (which claims that the future is unavoidable) and determinism (which relies on the causal assumption) seems superfluous. Fatalism as the claim that certain events are going to happen *no matter what* and *regardless of their causes* is, for Taylor, "enormously contrived": "it would be hard to find in the whole history of thought a single fatalist, on that conception of it" (Taylor 1974, 55). Fatalistic claims about *unavoidability* and deterministic claims about *truth* and *necessity* coincide and differ only with regards to the perspectives from which they are made. Like Taylor, Steven Cahn identifies fatalism with the thesis that:

> [T]he laws of logic alone suffice to prove that no man has free will, suffice to prove that the only actions which a man can perform are the actions which he does, in fact, perform, and suffice to prove that a man can bring about only those events which do, in fact, occur and can prevent only those events which do not, in fact, occur. (Cahn 1967, 8)

This attempt is supported by many authors. According to Peter van Inwagen, for example, fatalism is "the thesis that it is a logical or conceptual truth that no one is able to act otherwise than he in fact does; that the very idea of an agent to whom alternative courses of action are open is self-contradictory" (van Inwagen 1983, 23). Similarly, Paul Horwich describes fatalism as follows:

> What was true in the past logically determines what will be true in the future; therefore, since the past is over and done with and beyond our control, the future must also be beyond our control; consequently, there is no point in worrying, planning and taking pains to influence what will happen. (Horwich 1988, 29)

Finally, J. M. Fischer conceives of fatalism as "the doctrine that it is a logical or conceptual truth that no person is ever free to do otherwise" (Fischer 1989, 8).

10. Determinism and fatalisms

Taylor is only partly correct when he writes that "it would be hard to find in the whole history of thought a single fatalist", for there is a very reasonable sense in which we might hold that certain events are "unavoidable" even though they are not subject to strict causal necessity. In ancient texts, for example, we find a wide range of conceptions in which fate differs from necessity. Examples of this can be found in Cicero's *De fato* and *De divinatione* (among other of his works). These different forms of ancient fatalism can be distinguished according to certain 'topological points'. In the case of so-called event fatalism, future events are unavoidable in relation to either *time*, *place*, *means* (*the way* they are realized) or *kind* (arising from some necessary realization of disposition, etc.; cf. Marko

2011a; 2011b). In some cases, unavoidability hinges on the correlation between a sign and that which is signed, as in the Stoic example of the conditional predictive sentence "If Fabius was born during the Dogstar, he will not die at sea," where the relation between the antecedent state and the consequent state is to be interpreted neither strictly causally nor via classical propositional implication but in terms of a certain sort of *connectedness* (e.g. in terms of relevant connection or a "responsibility" relation).

Since fatalism is not always about fixed points in time, in many cases it is not connected to the examination of causes, laws, logical laws, and the like. Many of these conceptions do not make room for the possibility of agency, as is the case with conditional fate. What ancient fatalisms have in common is that they generally concern truth in advance of a given happening – once in the past, it was true that at least one kind of entity (event, occurrence, disposition or truth of proposition) would inevitably be actualized (in this or that way).

Although logical determinism (at least in Bradley's sense) and logical fatalism (in Taylor's or Cahn's sense) seem to correspond to each other conceptually, they do not necessarily equally correspond to all forms of fatalism. Some forms of ancient fatalism correspond to, for example, John Earman's naturalistic fatalism – the view that an event will occur in every physically possible world, "no matter what happens" – "for instance, that the laws of biology dictate that I am naturalistically fated to die". Earman also claims, however, that "naturalistic fatalism in this sense neither entails nor is entailed by determinism" (Earman 1986, 18).

Susan Haack argues that theological fatalism (theological determinism) is an upgraded version of the argument for logical fatalism (see Haack 1974). An additional proposition with theological content is added (for example 'God is omniscient' or 'God is omnipotent') – one that is formally inessential to the proof. Since the logical premisses are independent of theological content, the additional premise plays a superfluous role in the argument, a redundant detour from its logic.

11. Nomological determinism

Several modern arguments for incompatibilism rely on explicit deterministic assumptions (e.g. *The Direct Argument* and *The Consequence*

Argument – see van Inwagen 1983). In his exposition of determinism, van Inwagen (1983, 184-188) starts from the simple assumption that the past determines precisely one physically possible future. On his view, determinism as a thesis about propositions need not be identified with determinism based on the principle of universal causation. He does not feel obliged to accept the principle of universal causation and doubts that this principle even entails determinism or that determinism entails causation (so there is still space for immanent causation despite the fact that in complex physical events it is an open question how and whether causation can be distinguished). Laws, however, are firm constraints that limit our abilities.

Van Inwagen understands a law of nature as "any set of worlds that has as a subset the set of all worlds in which the laws of nature are the same as those of the actual world, or, as we might say, are *nomologically congruent* with the actual world" (van Inwagen 1983, 65).

Determinism is presented as a conjunction of the following two theses:

> For every instant of time, there is a proposition that expresses the state of the world at that instant.
> If p and q are any propositions that express the state of the world at some instants, then the conjunction of p with the laws of nature entails q.[5] (van Inwagen 1983, 65)

Determinism consists in the antecedent conjunction of *past truths* (Po) and *the laws of nature* (L), while agency (A), is conditioned by that conjunction [*i.e.* $\Box((\mathrm{Po} \,\&\, \mathrm{L}) \rightarrow \mathrm{A})$].

In addition to this conception, according to which human agency is determined by past truths and natural laws, we find a wide range of approaches that interpret this basis in a compatibilistic manner, allowing for agent-causation as an intervening link in the deterministic chain. The notion of causality plays a central role here. Some compatibilists, continuing in the tradition of the "soft determinism" held by James, accept both causal determinism and logical determinism, while others do not fully accept either the former or the latter.

---

5 Cf. also van Inwagen (1983, 58; and 2004, 344).

Contrary to the above formulations from Schlick, the laws of nature used to be frequently understood and qualified as causes. The so-called ontic conception (cf. Salmon 1998, 54), widely accepted in scientific practice, claims that since laws are the explanatory engines of occurrences in the physical world, they can be interpreted as being responsible for occurrences: "laws of nature stand in no need of 'philosophical analysis'; they ought to be posited as ontological bedrock" (Maudlin 2007, 1).

When we appeal to laws of nature, our starting point is the conviction that they are in some sense fundamental and cannot be reduced to other, more primitive notions. They are, we assume, basic ontological notions, since "our world is governed by laws". In this sense, as Carl Hoefer puts it, a law is a cause "that makes things happen in a certain way" (Hoefer 2010).

12. Determinism without laws and causation

Is it possible to represent determinism without laws? This depends on how we interpret laws. Some interpretations of laws of nature do not depend on the notion of a cause. For example, Ernest Nagel's syntactical formulation of laws of nature conceives of them according to their logical function. On Nagel's view, a set of laws is deterministic with regards to an isolated system of bodies, relative to a definite class of properties, if, given the state of that system at any initial time, those laws logically determine a unique state of that system for any other time (cf. Nagel 1999, 281).[6] Laws are theoretical notions, and according to Nagel, "a theory is deterministic if, and only if, given its state variables for some initial period, the theory logically determines a unique set of values for those variables for any other period" (Nagel 1999, 292).

If we wish to see the relation between two states as causally connected and thus to assume a causal version of determinism, this would seem to lead us toward *ontological determinism*. For this reason, Nagel insists that causality should be kept apart from determinism. Some authors pre-

[6] We find a similar opinion in Hempel: "a deterministic theory provides a system of laws which, given the state of an isolated system at one time, determine its state at any other time" (Hempel 1962, 107).

fer to keep causality in their accounts of determinism as "a pure theoretical notion" – a useful concept that has indubitable explanatory advantages (cf. Tooley 1987, Chap. 9).

13. Deterministic theory and possible worlds

A more precise definition – one that may partly avoid the above difficulties – comes from Richard Montague (see Montague 1974; later slightly reformulated by Earman 1986, 12-14). Montague develops Nagel's (1953) earlier formulations. His idea is that determinism can be seen as a property of theories. Briefly, a theory is interpreted by a formal semantics approach and is associated with a class of models. The objects of a theory are represented as systems, properties are states, and regularities can be represented as functions that ascribe a value to some point on the *t* axis (and can be interpreted as "the laws of the theory"). Thus a theory *T* is deterministic if any of at least two possible histories (*S* and *S'*) that realize (satisfy – *Rl*)[7] *T* and which are identical at a given time t_0 are identical at all times *t*. *T* is deterministic if and only if all models of that theory that agree on the state of the world at one time also agree at certain other times.

Let us suppose that *S* is a history, where $S = \langle D_1, \ldots, D_n \rangle$ and where *D* is one argument function defined at least for all real numbers *R*, and that the state of *S* at *t* ($st_S(t)$) is defined as $st_S(t) = \langle D_1(t),\ldots, D_n(t) \rangle$. Then,

- a theory is historically determined:
  If $S, S' \in Rl(T)$, $t_0, t \in R$, $t_0 < t$, and $st_S(t_0) = st_{S'}(t_0)$,
  then $st_S(t) = st_{S'}(t)$;

- a theory is futuristically determined:
  If $S, S' \in Rl(T)$, $t_0, t \in R$, $t < t_0$, and $st_S(t_0) = st_{S'}(t_0)$,
  then $st_S(t) = st_{S'}(t)$.

---

[7] A formula ϕ of *L* is realized *by* a history *S* (i.e. *Rl*) just in case there is a standard model *M* of language *L* such that history *S* is a partial model corresponding to $M(S = Pm(M))$ and formula ϕ is true in model *M*. *S* realizes a class of formulas or theory *K* (in symbols, $S \in Rl(K)$) if there is a single standard model M such that $S = Pm(M)$ and theory K holds in model M.

A theory is *deterministic* if it is both *historically* determined and *futuristically* determined. That is:

If S, S′ ∈ *Rl*(*T*), t_0, $t \in R$, and $st_S(t_0) = st_{S'}(t_0)$ *then* $st_S(t) = st_{S'}(t)$.

Earman interprets Montague's formulations in terms of physically possible worlds. Earman's modification of Montague allows for additional alternative approaches and different modes, where determinism can be interpreted not only as a property of the theory but also as a property of, for example, the set of laws, the world, the actual state of the universe (where the history is settled by the laws even though they do not determine future states of the universe), and so on.

14. Deterministic systems

Russell's observation concerning the principle of the irrelevance of time suggests an attempt to define determinism in a changing system in terms of *actual* trajectories alone rather than *possible* trajectories (which could be infinite in number and must be avoided). According to Montague, this approach will preclude the deterministic condition that a given state is always followed by the same history of state transition. Taking determinism as a modal notion, Bas van Fraassen tries to refine Russell's formulation, taking into account not only actual but also possible trajectories. On his view, a system is deterministic if two possible worlds have the same history of state transitions: "If *u* and *v* are possible histories, and $u(t) = v(t')$ then for all positive numbers *b*, $u(t + b) = v(t' + b)$" (van Fraassen 1989, 254).

Mark Stone (see Stone 1989) and Stephen Kellert (see Kellert 1993), via an analysis of Laplacean determinism, attempt to identify and extract key properties of determinism, which can serve as necessary and sufficient conditions for determinism:

a) there exists an algorithm which relates a state of the system at any given time to a state at any other time, and the algorithm is not probabilistic;

b) the system is such that a given state is always followed by the same history of state transitions;
c) any state of the system can be described with arbitrarily small (non-zero) error (cf. Stone 1989, 125).

On this view, determinism is a necessary condition for predictability, *but not vice versa*.

Stone, Clark (1989) and Kellert extend deterministic interpretation from linear to non-linear systems (systems usually interpreted as not fully stable or not transforming continually because they are affected by occasional "jumps"). These systems are deterministic, although they are only globally, not locally, predictable. Their defining feature is that, even though they behave chaotically, they periodically jump into patterned (deterministic) behavior: Although movements in these systems are characterized by infinite possibility, they oscillate within steady and predictable macro patterns. The chaotic behavior of the system is due to epistemic considerations (or to the lack of Laplacean "demonic" abilities on the part of the observer) with respect to computability and our inability to give precise initial conditions. Determinism is here accepted as an explanatory tool because some aspects of a system's evolution are coverable (not statistically or probabilistically, but) by strictly deterministic differential equations that allow for the "predictability of higher order characteristics" with respect to certain deterministic aspects of the system (related, for example, to its qualitative or topological character – cf. Kellert 1993, 56-57).

Deterministic properties can also be analyzed within the scope of quantum theory, especially quantum field theory. Some recent results support the thesis that quantum theory can also be interpreted as deterministic and that such an interpretation is entirely coherent (see Butterfield 2005).

15. The components of determinism

As we have seen in the previous sections, there are many forms of determinism and fatalism. I have referred to more than twenty different terms related somehow to these kinds of conceptions. Obviously, the list is far

from being exhaustive – it represents only a selection, here reduced to those conceptions that are in some sense dominant or frequently discussed in modern philosophy. In philosophy as well as in other disciplines (from a computer programming to economy), we could easily find other labels with some specific designations useful for a given discipline. Moreover, other expressions are frequently used to point out quite specific circumstances. For example, consider the term "fatalism" as it is nowadays used in medicine or psychology: An event, with respect to a patient's conditions, is unavoidable regardless of any treatment (or no treatment at all) – the event in question is related to some future fixed point irrespective of a way to be reached. It corresponds to the conception characterized here as event-fatalism or to Earman's naturalistic fatalism ("which neither entails nor is entailed by determinism" and it presents conception that has a very little in common with other forms of fatalism, either in a sense of scientific or logical fatalism).

Given the distinctions discussed so far, it is easy to find out that

a) similar (or even the same) conceptions of determinism may be designated by different names;

and that

b) various formulations labeled with the same name represent substantially different conceptions.

For example: a) deterministic conceptions in a sense of Russell's functional determinism, Montague's syntactical determinism, Nagel's nomological determinism (and, mostly, Popper's weak version of scientific determinism) have many elements in common. The same case is with a strong version of scientific determinism characterized by Popper and scientific fatalism sketched in Langer. Moreover, the terms like logical predestination, strict determinism, radical determinism, fatalism, etc., frequently refer to the same conception. From the other side, b) the metaphysical determinism is differently characterized by Łukasiewicz and Popper and these two are conceptually different. In addition, there are some vivid differences among the conceptions of logical determinism as characterized by Schlick, Waismann, Jordan, Wołenski, Bradley, Taylor and Cahn. In literature, these conceptions are frequently referred to by the same name. The discussion on

fatalism is alike – logical or scientific, naturalistic or event-fatalistic conceptions do not always coincide conceptually, since they do not share the same components. The component of ceaselessness, that is a substantial part of determinism, is not a necessary component of all forms of fatalism. All these cases witness a need for a more appropriate way for demarcating these conceptions.

As we have seen, the term "determinism" has been used in a very broad way to capture specific ideas and concepts that are conflicting and even mutually exclusive. The demarcation of different forms of determinism might well offer a solution – given an emphasis on their substantial differences. Since the conceptions discussed so far are complex, their differences (and varieties) can be elucidated via a matrix of some essential or fundamental properties (or their absence).

I suggest moving towards a classification of different forms of determinism from bottom-up rather than via a simple (theoretically heterogeneous) typology. Such a classificatory project would provide a more appropriate demarcation. It would enhance our understanding of the determinism's sub-forms and promote further research on the subject. Furthermore, a sorting of distinct forms of determinism that focuses on the mutual dependency of their components would offer an informative insight into the nature of these components. Simple typology only provides us with distinctions motivated by certain dominant properties. However, it cannot provide us with a satisfying picture of the mutual dependencies among singular types. For this reason, the search for a minimal common denominator of different forms of determinism is a worthwhile direction of inquiry.

It seems that this basic level must consist in the notion of "functional determination", for this represents the common core of all further elaborations. This core layer comprises only the simple "order" of variables and it is free of excessive additional features, such as, for example, a kind of relation (or nature of impact) among the entities or temporal character of direction (which is introduced when we refer to "state plus time (of occurrence)"). Such a characterization can preserve the basic order of entities, and it guarantees that entities are sorted according to a linear function. This core level, which involves only transitivity and continuity (i.e. ceaselessness), can serve as a basis for further developments of determinism, such

as those discussed by philosophers today. In addition, this feature enables opposition to ceased character of indeterminism. This basic level can be compared to John McTaggart's idea of a "flat" series of time, or a *C-series* (cf. Mc Taggart 1908, 462). It need not be understood as determinism itself; rather, it serves as a basis for the development of different forms of determinism.

If we add to this basic level a further conceptual component – the universal principle of causation – we get causal determinism. Further, by adding to this composition of causal determinism yet another element – "the laws of nature" – we have nomological causal determinism (the most appropriate with Laplace's view). If we omit the principle of causality but retain the laws of nature, we come to nomological determinism. As we have already seen, some forms of nomological determinism are advocated by Russell (who conceives of laws as functions), Nagel, Schlick and Montague (who construe laws as functions alike but with an additional emphasize on the role of the laws of thought). A form of nomological determinism is also present in van Inwagen's conception (with an additional claim that laws of nature are nomologically congruent with the actual world and that the real world inevitability can be understood as or related to a logical necessity). If we think of Salmon's, Maudlin's and Hoefer's conception as alike in their claiming that laws are responsible for occurrences and play the role of causes, then these conceptions will interfere with nomological causal determinism, too.

If we step back to the basic level and add both the principle of causation and the so-called "Aristotelian laws of thought" (the principles of non-contradiction, excluded middle and identity), this brings us to a form of logical determinism criticized by Jordan. If we enrich this new composition of the principle of correspondence (correspondence to so-called "real correlates"), then we obtain a metaphysical version of logical determinism which Łukasiewicz had in mind in his criticism of determinism. Other forms of determinism can be composed in a similar manner, with the basic lower layer amounting to a core that contains the minimal features of transitivity and continuity (ceaselessness) – the latter property serving as a component which is necessary both for the principle of causality and for the laws of nature (as well as for functions that serve as their substituents).

Another recommendation concerns the possibility of a minimal formulation of fatalism – one that has the essential property held in common by the above-listed forms of fatalism: inevitability. The minimal conditions for determinism obviously differ from the minimal conditions for fatalism. Inevitability is not a part of the minimal deterministic core (at least, as we saw above, it is not necessarily related to ceaselessness). So, it is possible to develop a minimal core of fatalism in accordance with the traditional forms mentioned above. However, it is also possible for these two core layers to intersect at a higher level. Such an intersection would require further additional assumptions related to causal, nomological or other specific properties. Some forms of the so-called logical fatalism can include layers that also belong to some forms of determinism – by reading implication causally or by interpreting inevitability as logical (or functional or lawful) necessity. The term "logical fatalism", usually used as a unique extension of logical determinism, may also be appropriate in case it includes, for instance, the principle of causality together with the laws of thought, and if it identifies necessity with inevitability. According to Jordan (contra Bradley), logical determinism (equipped with the principle of causality and the laws of thought, especially with the law of excluded middle identified with the principle of bivalence) leads to fatalistic inevitability. Taylor's formulation of fatalism also accepts causes and in the later versions he identifies universal causality with inevitability. However, the conception that accepts only the laws of thought but omits the principle of causality – frequently referred to as logical determinism – is not necessarily based on minimal core needed for determinism (i.e. ceaselessness). Bradley used to characterize such conception as timeless and Jordan relates it to timeless truths. The conception is considered by Ayer, Cahn, Ryle, Dummett and discussed under the term "fatalism" also by van Inwagen and Fischer.

Now it can be seen that if we take into account only some of the dominant components of the above mentioned conceptions and their combinations, we are able to recognize their common features and point to some basic differences between them. Let me illustrate these differences with the following table. Although incomplete, it includes the most fundamental components. Here, the names of particular conceptions are given temporarily and

marked with asterisks to avoid possible confusion with terms introduced hitherto.

	ceaselessness	*inevitability*	*the principle of causation*	*the laws of nature*	*covering function*	*"laws of thought"*
*causal determinism**	+		+			
*nomological causal determinism**	+		+	+		
*nomological determinism**	+			+		
*(functional) nomological determinism**	+				+	
*logical determinism**	+		+			+
*Taylor's "fatalism"**	+	+	+			+
*(genuine) fatalism**		+				
*logical fatalism**		+				+

With both determinism and fatalism, we are able to supply other features as building blocks (giving rise to further conceptions): time component (character of time, temporal direction, time symmetry or asymmetry), causality, logical, physical, epistemic or other properties (laws of nature, statistical laws, probabilistic laws, linear or nonlinear changes to the system, etc.). Adding any of these distinct elements results in still further distinct philosophical conceptions.

What I have tried to show above is that these combinations stem from more elementary layers that must be further investigated as the composite elements of complex conceptual structures. Each of these combinations, no

matter how far they resemble other compounds, has its own meaning and leads to a different philosophical and conceptual standpoint. My primary interest in this paper was not to provide an exhaustive, systematic list of the various forms of determinism on offer but rather to draw attention to the fact that determinism is not a unitary philosophical conception as it is frequently used in the literature. I have outlined traditional and modern philosophical approaches to determinism and have provided a sketch of the means by which we might achieve a deeper understanding of their contours and points of intersection. The forms of determinism outlined above are composite in character. Reducing them to their more elementary building blocks helps us to better understand their composite conceptual structures. A deeper understanding of these elements is crucial to our ability to assess the theoretical consistency and sustainability of particular conceptions of determinism.

Acknowledgments

The work on this paper was supported by a grant from the *Slovak Research and Development Agency*, Grant No. APVV-0149-12.

References

ACKRILL, J. L. (1963): *Aristotle's Categories and de Interpretatione*. Oxford: Oxford University Press.

AYER, A. J. (1963): Fatalism. In: *The Concept of a Person and Other Essays.* New York: St. Martins Press, 235-268.

BAHRDT, C. F. (1790): Leb*en und Thaten des weiland hochwürdigen Pastor Rindvigius: Ans Licht gestellt von Kasimir Renatus Denarée*. Vol. 2. Ochsenhausen [i. e. Liebau]: Friedrich.

BRADLEY, R. D. (1959): Must the Future Be What It Is Going to Be? *Mind* 68, No. 270, 193-208.

BUNGE, M. (1959): *Causality*. Cambridge, (Mass.): Harvard University Press.

BUTTERFIELD, J. (2005), Determinism and Indeterminism. In: *Routledge Encyclopedia of Philosophy*, Vol. 3. London: Routledge.

CAHN, S. M. (1967): *Fate, Logic, and Time*. New Haven: Yale University Press.

CASSIRER, E. (1936): *Determinismus and Indeterminismus in der modernen Physik. Historische and systematische Studien zum Kausalproblem*. Göteborg: Elanders

Boktryckeri Aktiebolag. [English translation: Cassirer, E. (1956): *Determinism and Indeterminism in Modern Physics*. New. Haven: Yale University Press].

CLARK, P. (1989): Determinism, Probability and Randomness in Classical Statistical Physics. In: Gavroglu, K., Goudaroulis, Y. & Nicolacopoulos, P. (eds.): *Imre Lakatos and Theories of Scientific Change*. *Boston Studies in the Philosophy of Science* 111. Springer, 95-110.

DU BOIS-REYMOND, E. (1886): *Über die Grenzen des Naturerkennens* (Versammlung Deutscher Naturforscher und Aerzte, 14 August 1872). In: Du Bois-Reymond, E. (1886): *Reden von Emil Du Bois Reymond*. Leipzig: Veit und Comp, 105-140.

DUMMETT, M. (1964): Bringing About the Past. *Philosophical Review* 73, No. 3, 338-359.

EARMAN, J. (1986): *A Primer on Determinism*. Springer.

FISCHER, J. M. (1989): *God*, *Foreknowledge*, *and Freedom*. Stanford: Stanford University Press.

FRANKFURT, H. (1969): Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy* 66, No. 23, 829-839.

GLYMOUR, C. (1971): Determinism, Ignorance, and Quantum Mechanics. *The Journal of Philosophy* 68, No. 21, 744-751.

HAACK, S. (1974): On a Theological Argument for Fatalism. *The Philosophical Quarterly* 24, No. 95, 156-159.

HACKING, I. (1983): Nineteenth Century Cracks in the Concept of Determinism. *Journal of the History of Ideas* 44, No. 3, 455-475; reprinted in: Hacking, I. (1990): *The Taming of Chance.* Cambridge University Press, Chap. 18, 150-159.

HAMILTON, W. (1846): *The Works of Thomas Reid, D.D. Now Fully Collected, with Selections from His Unpublished Letters*. Edinburgh: MacLachlan, Stewart & Co.

HEGEL, G. W. F. (1816): *Wissenschaft der Logik*. Nürnberg: Schrag.

HEMPEL, C. G. (1962): Deductive-Nomological vs. Statistical Explanation. In: Feigl, H. & Maxwell, G. (eds.): *Minnesota Studies in the Philosophy of Science* III. Minneapolis: University of Minnesota Press, 98-169.

HERBART, J. F. (1804): *Pestalozzi's Idee eines ABC der Anschauung als ein Cyklus von Vorübungen im Auffassen der Gestalten*. Göttingen: J. F. Röwer.

HERBART, J.F. (1812), *Bemerkungen über die Ursachen, welche das Einverständnis über die ersten Gründe der prakt. Philosophie erschweren*. Königsberg: Fr. Nicolovius.

HERBART, J. F. (1842): *Kleinere philosophische Schriften und Abhandlungen: nebst dessen wissenschaftlichem Nachlasse*. Bd. 1. Hartenstein, G. (ed.), F. A. Brockhaus.

HERBART, J. F. (1843): Aphorismen und kürzere Fragmente, a. Zur *Einleitung in die Philosophie*. In: Hartenstein, G. (ed.): *Kleinere philosophische Schriften und*

Abhandlungen: nebst dessen wissenschaftlichem Nachlasse. Bd. 3. F. A. Brockhaus.

HEYDENREICH, A. L. CH. (1793): *Ueber Freyheit und Determinismus und ihre Vereinigung*. Erlangen: Johann Jakob Palm.

HOEFER, C. (2010): Causal Determinism. In: Zalta, E. N. (ed.): *The Stanford Encyclopedia of Philosophy.* Available at: http://plato.stanford.edu/entries/determinism-causal/#LawNat

HORWICH, P. (1988): *Asymmetries in Time, Problems in the Philosophy of Science*. London: MIT Press.

JAMES, W. (1884/1907): The Dilemma of Determinism. *The Unitarian Review* 22 (September 1884), No. 3, 193-224. Reprinted in: James, W. (1907): *The Will to Believe and Other Essays in Popular Philosophy*. New York: Longmans Green and Co., 145-183.

JORDAN, Z. (1963): Logical Determinism. *NJDFL* IV, No. 1, 1-38.

KANT, I. (1793): *Die Religion innerhalb der Grenzen der bloßen Vernunft*. Leipzig: Druck und Verlag von Philipp Reclam.

KELLERT, S. H. (1993): *In the Wake of Chaos: Unpredictable Order in Dynamical Systems*. Chicago: University of Chicago Press.

KRUG, W. T. (1827-1829): *Allgemeines Handworterbuch der philosophischen Wissenschaften nebst ihrer Literatur und Geschichte*. Vol. i-v, Leipzig: F. A. Brockhaus.

LANGER, S. (1936): On a Fallacy in 'Scientific Fatalism'. *International Journal of Ethics* 46, No. 4, 473-483.

LAPLACE, P. S. (1902): *A Philosophical Essay on Probabilities.* Truscott, F. W. & Emory, F. L. (transl.), London: Wiley.

ŁUKASIEWICZ, J. (1990): On Determinism. In: Borkowski, L. (ed.): *Selected Works. Studies in Logic and the Foundations of Mathematics*. Amsterdam, London: Nort-Holland, 110-128.

MARKO, V. (2011a): Looking for the Lazy Argument Candidates 1. *Organon F* 18, No. 3, 363-383.

MARKO, V. (2011b): Looking for the Lazy Argument Candidates 2. *Organon F* 18, No. 4, 447-474.

MAUDLIN, T. (2007): *The Metaphysics within Physics*. Oxford: Clarendon Press.

MCTAGGART, J. E. (1908): The Unreality of Time. *Mind* 17, No. 68, 457-474.

MILL, J. S. (1843): *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence, and Methods of Scientific Investigation*. London: J. W. Parker.

NAGEL, E. (1953): The Causal Character of Modern Physical Theory. In: Feigl, H. & Brobeck, M. (eds.): *Readings in the Philosophy of Science*. New York: Appleton Century Crofts, 419-437.

NAGEL, E. (1999): Alternative Descriptions of Physical State. In: *The Structure of Science: Problems in the Logic of Scientific Explanation.* 2nd ed., Indianapolis: Hackett: 285-292.

POPPER, K. (1982): *The Open Universe: An Argument for Indeterminism*. Bartley III, W. W. (ed.), London: Hutchinson.

PRIESTLEY, J. (1779): *A Letter to the Rev. Mr. John Palmer. In Defense of the Illustrations of Philosophical Necessity*. Bath – London: printed by Cruttwell, for J. Johnson.

PRIESTLEY, J. (1780): *A Second Letter to the Rev. Mr. John Palmer: In Defense of the Doctrine of Philosophical Necessity*. London: printed by H. Baldwin, for J. Johnson.

RUSSELL, B. & WHITEHEAD, A. N. (1910-1913): *Principia Mathematica*. Cambridge: University Press.

RUSSELL, B. (1912-1913): On the Notion of Cause. *Proceedings of the Aristotelian Society* 13, 1-26. Reprinted in: Russell, B. (1917): *Mysticism and Logic and Other Essays*. Auckland: Floating Press, Chap. ix, 180-208.

RYLE, G. (1953): *Dilemmas: The Tarner Lectures*. Cambridge: University Press.

SALMON, W. (1998): *Causality and Explanation*. Oxford: University Press.

SCHLICK, M. (1931): Das Kausalität in der gegenwärtigen Physik. *Naturwisseschaften J. H.* 7, No. 19, 145-162.

SCHLICK, M. (1932): Causality in Everyday Life and in Recent Science. *University of California Publications in Philosophy* 15, 99-125.

SCHLICK, M. (1939): When is a Man Responsible? In: Rynin, D. (ed.): *Problems of Ethics*. New York: Dover, 143-156.

SCHULZ, H., BASLER, O. & STRAUSS, G. (eds.) (1999): *Deutsches Fremdwörterbuch*. Band 4. Berlin, New York: de Gruyter.

SNELL, CH. W. (1789): *Über Determinismus und moralische Freiheit*. Offenbach: bey Ulrich Weiss und Carl Ludwig Brede.

STONE, M. A. (1989): Chaos, Prediction and Laplacean Determinism. *American Philosophical Quarterly* 26, No. 2, 123-131.

SUPPES, P. (1993): The Transcendental Character of Determinism. *Midwest Studies in Philosophy* 18, No. 1, 242-257.

TAYLOR, R. (1962): Fatalism. *Philosophical Review* 71, No. 1, 56-66.

TAYLOR, R. (1974/1992): *Metaphysics*. 4th ed. Englewood Cliffs, N.J.: Prentice-Hall.

TOOLEY, M. (1987): *Causation: A Realist Approach*. Oxford: Clarendon Press, New York: Oxford University Press.

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

VAN INWAGEN, P. (1983): *An Essay on Free Will*. Oxford: University Press.

VAN INWAGEN, P. (2004): Freedom to Break the Laws. *Midwest Studies in Philosophy* 28, 334-350.

WALLACE, D. F. (2011): *Fate, Time and Language – An Essay on Free Will*. Columbia University Press.

WAISMANN, F. (1959): How I See Philosophy? In: Ayer, A. J. (ed.): *Logical Positivism*. Glencoe, (Ill.): The Free Press, 354-380.

WOŁENSKI, J. (1996): An Analysis of Logical Determinism. (Paper delivered at the conference *Łukasiewicz in Dublin*, held at University College Dublin, July 1996).

Organon F 24 (1) 2017: 85-104

Meaning-Constitutive Inferences[1]

MATEJ DROBŇÁK

ABSTRACT: A traditional objection to inferentialism states that not all inferences can be meaning-constitutive and therefore inferentialism has to comprise an analytic-synthetic distinction. As a response, Peregrin argues that meaning is a matter of inferential rules and only the subset of all the valid inferences for which there is a widely shared corrective behaviour corresponds to rules and so determines meaning. Unfortunately, Peregrin does not discuss what counts as "widely shared". In the paper, I argue for an empirical plausibility of Peregrin's proposal. The aim of the paper is to show that we can find examples of meaning-constitutive linguistic action, which sustain Peregrin's response. The idea is supported by examples of meaning modulation. If Peregrin is right, then we should be able to find specific meaning modulations in which a new meaning is publicly available and modulated in such a way that it has a potential to be widely shared. I believe that binding modulations – a specific type of meaning modulations – satisfy this condition.

KEYWORDS: Inferentialism – meaning – meaning-constitutive inferences – meaning modulation – normative inferentialism.

1. Introduction

Despite the progress in making inferentialism more rigorous, accomplished thanks to the work of Robert Brandom (1994, 2000) and others,

[1] Received: 26 November 2016 / Accepted: 12 January 2017

✉ Matej Drobňák
Department of Philosophy and Social Sciences, University of Hradec Králové
Náměstí Svobody 331, 500 02 Hradec Králové, Czech Republic
e-mail: matej.drobnak@gmail.com

inferentialism still faces many objections.[2] One of the traditional objections focuses on the analytic–synthetic distinction. According to inferentialism, meaning depends on inferences held valid by speakers. Clearly, so the objection goes, not all of the inferences we make can be meaning-constitutive and therefore inferentialism has to include a satisfactory version of the analytic–synthetic distinction. Since this is a Sisyphean task, the reputation of inferentialism seems to be corrupted.

A promising attempt to answer the objection can be found in Peregrin (2014b).[3] Even though answering the objection is not his main objective, Peregrin argues that meaning is a matter of inferential rules and only the subset of all the valid inferences for which there is a *widely shared corrective behaviour* among members of some community corresponds to rules (and is therefore meaning-constitutive). Unfortunately, Peregrin does not discuss what counts as "widely shared". This opens a way to a possible objection if we make use of Peregrin's proposal in the metasemantic debate on meaning constitutiveness. Someone can claim that the criterion of "widely shared corrective behaviour" may be an interesting theoretical proposal, but it is excessively vague and therefore it is of no use.

In this paper, I argue for an empirical plausibility of Peregrin's proposal. The aim of the paper is to show that we can find examples of meaning-constitutive linguistic action, which take place in specific communication situations. The idea is supported by examples of meaning modulation from Ludlow (2014). Meaning modulation is, first and foremost, a phenomenon which can be observed in communication. Speakers often change, adjust and discuss meanings of the words they use for various purposes. In general, we can understand meaning modulation as a tool, which facilitates successful communication by deciding open questions about a meaning of

---

[2] For more references on progress in inferentialism see, e.g., Boghossian (2003, 2012), Peregrin (2006, 2010, 2012), Shapiro (2004).

[3] Brandom (2007) offers a response to this objection, but Fodor and Lepore find it unsatisfactory. Brandom builds on Sellars's (1949, 296) idea that "conceptual connections are just the lawful ones" (2007, 661). However, this is a weak response as it leads to a consequence that if speakers are wrong about laws, then the words they use mean something else as what the speakers intend them to mean. The view that our words can mean something else as what we intend them to mean is highly controversial. For criticism of such a view, although in a different context, see Schwarz (2013).

a word, by making some of the features of a meaning more explicit or by changing some of the features of a meaning. A paradigmatic example of a meaning modulation is the discussion of whether Pluto should be a planet. It can be seen – from the semantic point of view – as a discussion and clarification of the meaning of the word 'planet'.

What I find interesting about meaning modulations is the way how they can be (and often are) used within communities. Some of the modulations not only facilitate successful communication, but also serve to settle precedents, which are subsequently adopted and followed by other members of a linguistic community. They are part of more general social mechanisms which operate on the level of whole communities and which constitute new meanings.

If Peregrin is right and meaning is determined by inferential rules, then situations of meaning modulation should support his criterion of widely shared corrective behaviour: we should be able to find a specific type of modulations in which an outcome of the modulations is publicly available and modulated in such a way that it has a potential to establish a widely shared corrective behaviour. I believe that binding modulations – a specific type of modulations – satisfy this condition.

2. Preliminaries

The objection mentioned above was explicitly formulated by Fodor and Lepore (2001, 2007), who attempt to criticize several aspects of inferentialism. Among others, they argue that inferentialism has a problem stating which inferences are meaning-constitutive.

According to them, there are many inferences which are actually made but which are not/should not be semantically relevant. If inferentialism is a doctrine that meaning is an inferential role, i.e. a set of inferences in which an expression plays a role, an inferentialist needs to delineate clear boundaries of meaning-constitutive inferences. Fodor and Lepore believe that this means that an inferentialist has to revive the well-known analytic-synthetic distinction to distinguish between meaning-constitutive and "utterly contingent" inferences. As Quine (1951) persuasively showed, this seems to be a task doomed to failure.

> We're also not clear what Brandom thinks about the status of utterly contingent inferences like "If it's a plant in my backyard and it's taller than 6 feet, then it's a tree". He does apparently endorse the idea that "[the concept-constitutive inferences] must include … those that are materially [sic] correct" (MIE, p. 657). But what he gives as examples are two he borrows from Sellars: "A is to the East of B" ⊢ "B is to the West of A" and "Lightning is seen" ⊢ "Thunder will be heard soon". We find this puzzling since the first of these strikes us as arguably conceptually necessary (whatever that means) and the second strikes as arguably nomologically necessary (whatever that means). So even if we granted that both are concept-constitutive, we would still want to know whether clear cases of purely contingent hypotheticals are too; and, if they aren't, how Brandom proposes to do without an analytic/synthetic distinction. (Fodor & Lepore 2007, 680-681)

2.1 Inferentialism and Inferential Role Semantics

It is important to understand the difference between Inferential Role Semantics (IRS) and inferentialism as advocated by Peregrin before responding to the objection.[4] According to IRS as proposed by Boghossian (1993, 2012), the meaning of an expression can be understood as its inferential role. The inferential role is then explained as a set of all the valid inferences in which the expression takes part. Therefore, to understand a meaning of a sentence is to know which other sentences are inferentially connected to the sentence. If the inferential role is understood as a set of *all* the valid inferences related to a sentence and meaning is an inferential role, then all the inferences should be meaning-constitutive – and so such a view is problematized by the objection mentioned above.

---

[4] Despite the fact that Fodor and Lepore address their criticism to Brandom, I will mention his work only to a very limited extent in this paper. The reason is exegetical. Fodor and Lepore present Brandom's views in a way that more or less fits IRS (Boghossian's approach). Peregrin argues that this is a misinterpretation of Brandom and builds a response to the objections on what he sees as a more "Brandomian view". Instead of entering an exegetical discussion, I will talk about Boghossian's inferentialism (IRS) as a target of criticism and Peregrin's inferentialism as a response to the objection.

Such a view is problematic for more reasons. The approach of IRS is individualistic in nature – inferences, which are part of the inferential role of an expression, are determined by the dispositions of a *particular speaker* and her ability to distinguish valid and invalid inferences. As Boghossian puts it, it is determined by an ability to "*infer* from S1 to S2, but not to S3" (Boghossian 1993, 73). Such an approach opens the way once more for the objection mentioned above. If the inferential role depends on the practices of particular speakers, it is not clear how to delineate the boundary between meaning-constitutive and non-constitutive inferences. In particular contexts, some inferences that we would normally be inclined to call meaning-constitutive may be less important (e.g. for successful communication) than some contingent inferences. More importantly, if an inferential role of a sentence is the set of *all* the valid inferences in which a sentence appears, then different speakers ascribe (slightly) different meanings to the same sentence.[5] Which inferences a speaker includes in the inferential role of a sentence depends on his personal experience, and this is a highly subjective factor.

On the other hand, inferentialism as advocated by Peregrin is in some sense independent of the abilities of particular speakers. As Peregrin puts it: "Language is essentially public, and as such it cannot rest on private associations." (2014b, 45) Meaning is established in the social interactions of many speakers. Additionally, meaning persists within a community only through the existing normative attitudes of speakers – speakers hold some inferences to be correct and by their corrective behaviour force others to conform. If someone ascribes a set of inferences to a sentence which is not in accordance with the publicly established meaning, then she is just wrong and misunderstands the expression (and so she is a legitimate target of criticism).

---

[5] It is an open question if small differences in inferential roles are acceptable. Fodor and Lepore build their objection on the assumption of fully shared meanings because they do not see any viable similarity-based alternative. See Fodor & Lepore (1999) for their discussion of meaning similarity. I believe that the assumption of fully shared meanings is problematic because it does not correspond to the actual linguistic practice of speakers; regardless of the fact whether there is any alternative. However, this is not the topic of this paper and I postpone the discussion for another occasion.

It must be emphasized that Peregrin's version of inferentialism is not completely independent of the abilities and practices of speakers. I agree that the meaning of a word is independent of the inferential practices of each particular speaker – I cannot change what a word means within a linguistic community solely by changing my own inferential practices. However, the meaning of an expression still depends on what the majority of speakers (a minority with high semantic authority, maybe)[6] holds and protects as correct – i.e. it depends on the actual practices of *many* speakers.[7]

Of course, even in this "communal" setting, the sets of all the valid inferences related to particular sentences by individual speakers can vary. So how can Peregrin avoid the objection and distinguish meaning-constitutive and non-constitutive inferences? According to Peregrin, inferential roles should be understood as sets of *inferential rules* or, in some sense, as sets of inferences which correspond to inferential rules. An example of inferential rules can be 'X is a dog ⊢ X is an animal' or 'X is a dog ⊢ X is not a cat'. Peregrin also accepts inferential rules linking a sentence to some extralinguistic factors, which can have the form 'X is a dog ⊢ …' in which the three dots indicate some action that is inferable from the sentence, e.g. not irritate X. The inferential role of the sentence 'Laika is a dog' includes the inferences 'Laika is a dog ⊢ Laika is an animal'; 'Laika is a dog ⊢ Laika is not a cat'; 'Laika is a dog ⊢ …' (not irritate Laika). However, the sentence also appears in "utterly contingent" inferences such as 'Laika is a dog ⊢ Laika cannot enter John's apartment' or 'Laika is a dog ⊢ Laika can be off-leash in many areas of Central Park'.

Peregrin's key to deciding which inferences correspond to inferential rules (i.e. are meaning-constitutive) and which are "utterly contingent" lies in the widely shared corrective behaviour of speakers. Corrective behaviour is any kind of behaviour by which speakers respond to the language

---

[6] For the sake of simplicity, I will talk about a majority of speakers from this point onward. But in many contexts (e.g. in the case of scientific terminology), we cannot expect that a majority of speakers really knows all the correct inferences. Semantic authority plays a significant role in language distribution and preservation and it has to be taken into account. In fact, we can understand 'majority of speakers' as a group of speakers with higher semantic authority or we can simply talk about a majority of speakers whose opinion is semantically relevant.

[7] A similar point was emphasized by Koreň (2017a; 2017b).

use of other speakers. This includes positive as well as negative reactions – rewards in the case of correct inferences and warnings and punishments in the case of incorrect inferences.

> There is an inferential rule in force for a given language if the speakers of the language tend to see some inferences that violate the rule as incorrect. (Peregrin 2014b, 58)

and elsewhere:

> And what I call a propriety, or an (implicit) rule, grows out of such attitudes resonating throughout the surrounding society. (Peregrin 2014b, 10)

In the second quote, Peregrin talks about attitudes, but the attitudes of speakers matter only because they can be expressed behaviourally via corrective behaviour. If I tend to see some inferences as valid, then I tend to correct speakers who violate them. What is even more important is the phrase 'resonating throughout the surrounding society'. While inferences such as 'Laika is a dog ⊢ Laika is an animal' are *publicly well known* and widespread and there is an established *widely shared corrective behaviour* of speakers related to such inferences, inferences such as 'Laika is a dog ⊢ Laika cannot enter John's apartment' depend on the knowledge of particular speakers and so the (relevant) majority of speakers is not able to evaluate their validity. If the speakers are not able to evaluate the validity of such inferences, then they are not able to use corrective practices either and such inferences cannot be meaning-constitutive.

To sum up, we can decide which inferences are meaning-constitutive (i.e. correspond to inferential rules) by evaluating for which inferences there is a widely shared corrective behaviour among the members of a community. In some sense, Peregrin's proposal serves as a criterion of meaning-constitutiveness.

3. Meaning modulations

The question is if we can find something that corresponds to the inferential roles as proposed by Peregrin at the level of natural languages and

linguistic communities: If there are some inferences which are "widely shared" or if Peregrin's proposal is an unreasonable abstraction. I believe that we can find examples of meaning modulations that show that there are social mechanisms important for establishing new meanings at the level of whole communities. However, before I focus on the empirical plausibility of the criterion in more detail, I will present the topic of meaning modulations in general and briefly sketch its relation to inferentialism. Let us have a look at the conversation from the TV series *The Apprentice* in which Donald Trump (and his aide Caroline) discuss an incident that involved a contestant (Ivana) which happened while she was dealing with a given task – to sell a candy bar:

01 Trump: Ivana. You flashed a group of people.
02 Ivana: Look (…) This…
03 Trump: No, no, no. Did that happen?
04 Ivana: It happened? But it happened for a reason.
05 Trump: Why?
06 Ivana: Because I knew (…) Okay we had gone through a lot of product (…) We only had…
07 Trump: What does flash mean? You ripped down your pants? What does that mean?
08 Ivana: I was wearing (…) I was wearing a bikini (…) and (…) and let's not blow this out of proportion. I was wearing bikini shorts.
09 Caroline: We haven't said anything yet so relax.
10 Ivana: More: I know. I know. I'm just really defensive about this because…
11 Trump: Go ahead I'd like to hear that.
12 Ivana: Um.
13 Trump: But you did flash.
14 Ivana: I did. But it was a gimmick. It was a gimmick, just like (…)[8]

[8] The example first appeared in Sidnell (2010), who used it to show how we use the communicational tool of repair – how we go back in conversation to deal with troubles in understanding. The original transcript conventions used by Sidnell are not

Ludlow (2014) uses the example to show how we – more or less implicitly – modulate/litigate meanings within conversations. In this particular case, the word 'flash' has been questioned. In lines 02 and 04, Ivana accepts Trump's accusation of flashing with slight hesitation. Probably, she hesitates for more reasons but as the conversation shows later on, she does not agree that what she did is an evident case of flashing. In line 07, Trump indicates that he is not sure about the meaning of the word (despite the fact that he introduced it into the conversation) and Ivana tries to cash in on Trump's doubts: in line 08, she indicates that flashing should not apply in cases in which someone is wearing a bikini (and so she discusses the boundaries of the meaning of 'flash'). However, Trump does not accept her modulation and forces Ivana to admit that her behaviour was clearly a case of flashing (line 13). Ivana finally defers to Trump and admits that she flashed (14).

Situations like this are interesting for metasemantics in several ways. Most importantly, such situations are quite common and, as Ludlow argues, they should show that meanings are in general underdetermined and meaning modulations serve to specify the meanings for particular conversations and speakers. In other words, the shared language of community is a myth. There are only microlanguages that are created and modulated on the fly and very often include only the speakers who are present, without any impact on other speakers.[9] In the Trump example, it does not matter if there is a correct meaning of 'flash'. It may even happen that an act counts as flashing only if the exposed body is naked and so Ivana did not flash. But Ivana deferred to Trump's understanding of the word and her acceptance settles what the word means within their conversation.[10]

---

important in this context and I decided to use a much simpler transcript: '(…)' indicates a pause made by a speaker and '…' indicates interruption of the speaker by another speaker.

[9] Even though Ludlow focuses on different phenomena, he basically follows Davidson (1986, 1994) in his conclusion about shared language as a myth.

[10] It would be interesting to look at how rational the game of giving and asking for reasons is if understood in terms of meaning modulations. As far as I can see, Ivana did not defer to Trump because he was right or because he offered rational reasons for why her behaviour counts as flashing. She deferred because he was an authority in general – it was his show; he was a judge and her prospective employer. However, this is not the aim of this paper.

Examples like this can easily be "translated" in the inferentialist's terms. We can say that speakers discuss or disagree on the validity of some inferences. In this particular case, it can be an intralinguistic inference 'You flash ⊢ You are completely naked'. Obviously, Trump does not accept the inference, but Ivana would be happy to accept it. An advantage of inferentialism is that it can find meaning litigations in even less obvious circumstances. Let us have a look at the conversation from the TV series Gilmore Girls where Lorelai and her mother Emily dispute whether the offer of a lunch is still on if someone changes their previous plans:

01 Emily: Stop being so dramatic. I just showed up for lunch…
02 Lorelai: What do you mean you showed up for lunch?
03 Emily: Our lunch, at 1:00. You, me, Rory – the three of us. We're having lunch, aren't we?
04 Lorelai: I didn't think so.
05 Emily: You didn't?
06 Lorelai: Well, no, but (…)
07 Emily: When you invited your father and me for the weekend, you said it included a lunch with you and Rory.
08 Lorelai: Well, yes, I know, but that was before you left.
09 Emily: What does my leaving have to do with anything?
10 Lorelai: Well, when you left, you weren't here anymore. You were gone, so we just assumed lunch was...
11 Emily: Where's Rory?
12 Lorelai: Okay, see, you left, so (…)
13 Emily: She's not here, is she?
14 Lorelai: No.
15 Emily: Didn't she know about the lunch?
16 Lorelai: Yes, mom, she knew about the lunch, but you (…) so we (…) and she (…) I'll call her.
17 Emily: I'll wait.

From the inferentialist's perspective, the conversation can be reconstructed as a dispute over the validity of the inference 'You cancel your previous plans ⊢ You cancel the rest of the plans as well'. The validity of the inference is proclaimed in line 08 by Lorelai and challenged by Emily in the next line 09. Emily ignores Lorelai's repeated appeal to accept it and

Lorelai finally defers to Emily in line 16. The example is clearly a case of meaning modulation/litigation from the perspective of IRS. According to IRS, meaning depends on all the inferences held valid by particular speakers and the validity of an inference is in question here, therefore we can conclude that the meaning is in question. What is more, we can conclude that Lorelai has changed/adjusted her understanding of the sentences 'You cancel your previous plans' and 'You cancel the rest of the plans as well' during the conversation.

3.1 Meaning-constitutive modulations

However, the situation is less obvious from the perspective of Peregrin's inferentialism. Not all inferences are meaning-constitutive, i.e. not all inferences are maintained and reinforced by the widely shared corrective behaviour of a community of speakers. In the same manner, not all modulations can be meaning-constitutive as well. If we want to show that Peregrin's criterion presented earlier is empirically plausible, then we should be able to find litigations/modulations that establish meaning-constitutive inferences. In short, we should be able to discern meaning-constitutive modulations.[11]

The modulations presented in the previous examples are made within small groups of people (the conversation between Trump and Ivana is followed by a small group of contestants and judges, the conversation

[11] It is generally accepted in the philosophy of language that meaning change is a long-term, unconscious process. If there are any changes in meanings, they are usually implicitly adopted by speakers in the same way as most of the expressions of a language are learnt. Such a view is typical of Wittgenstein (1953), but also of Peregrin (2014b) and discussed in more detail in Peregrin (2014a). A similar view on meaning change and acquisition, discussed in the context of deciding signalling systems, can also be found in Lewis (1969, 129). In the following sections, I focus on examples of explicit meaning modulations and intentional acceptance of their results. By doing so, I do not intend to claim that this is the only way in which meaning can be adopted by speakers and become widely shared. I focus on explicit examples because the social mechanisms which are applied in the distribution and adoption of a new meaning in such cases are much more evident and so easier to document and analyse. I even think that both views are partially compatible. I can imagine a situation in which a meaning is settled in an explicit modulation, but after some time the modulation is forgotten and the meaning is adopted implicitly by new speakers.

between Emily and Lorelai is private) and there is no indication that these modulations should be applied globally as a precedent for other speakers.[12] On the contrary, it is likely that even Ivana and Lorelai will not follow the results of those modulations in future and their deference is only pretended. Since the modulations that were presented did not establish widely shared corrective behaviour, these modulations cannot be meaning-constitutive. Of course, in some cases, similar modulations which take place in personal communication may play an important role in the concept formation of particular speakers, but they are not important from the perspective of entire linguistic communities.

Now let us have a look at a different example. In 2000, Hayden Planetarium demoted Pluto from the status of a planet in their newly opened exhibition.[13] The decision was unusual at that time and it triggered a wave of criticism. One year later, the New York Times published a front page article called "Pluto Not a Planet? Only in New York", in which the author calls the decision "unilateral" and cited several astronomers who criticized the head of the planetarium, Dr. Neil de Grasse Tyson. The article started a "witch hunt" – Dr. de Grasse Tyson received many letters and emails from ordinary people demanding an explanation and renouncement of his view. Nevertheless, the article triggered an academic debate about the definition of 'planet' as well, and as the debate very soon showed, there were no clear criteria for calling an astronomical object a 'planet'. The International Astronomical Union therefore decided to redefine the term and the new definition did not apply to Pluto anymore: Pluto was officially relegated and pronounced a "dwarf planet" in 2006. Despite the fact that this decision raised a new wave of discussions, after 10 years we can say that it is generally accepted by the vast majority of astronomers, as well as non-experts.

When Ludlow presents the examples of Trump and Pluto, he admits that there is a difference – namely in the explicitness of the modulation. In the case of Pluto, astronomers explicitly discussed the meaning of the term 'planet', while in the case of Trump's conversation with Ivana, the litigation

---

[12] Both conversations are from TV series and they both have been seen by millions of people. However, this does not change the main aim of those examples – to show that there is an everyday phenomenon which *is* usually private.

[13] For a longer overview of the case see Weintraub (2007).

over the meaning of 'flash' was to a large extent implicit. As Ludlow puts it, in cases like Pluto "we are consciously aware of disputes about word meaning" (Ludlow, 2014, 39). As far as I can see, we can identify more differences and all of them are surprisingly well suited to a delineation of the class of meaning-constitutive modulations. The differences lie in

a) the intentions of the speakers;
b) the number of participants in a modulation;
c) information flow and its general accessibility.

a) *The intentions of the speakers*. Even though Lorelai and Emily were engaged in modulation, they do not have any reason to look for the most acceptable modulation. Emily wants to have a lunch with her daughter and granddaughter and her position in the litigation follows from this aim. Lorelai defers to Emily's modulation because she knows she has no chance of convincing her. Neither the intentions of Emily nor the intentions of Lorelai are directed towards the most plausible solution. In fact, it does not matter if there is any plausible modulation; even if there were, it would most probably be ignored. On the other hand, in the case of Pluto, the members of the International Astronomical Union try to find an acceptable modulation – acceptable with regard to the future use of the term within the whole community and with regard to possible future discoveries. In fact, in 2006, there was at least one known object of a size similar to Pluto and potentially there are more such objects in our solar system. The decision that Pluto is a planet would therefore lead to ad hoc decisions about the status of objects in our solar system or to a possible extreme increase in the number of planets. The declassification of Pluto is therefore a result of a reasonable debate looking for plausible solutions[14] for the whole astronomical community, and this was part of the intentions of the committee which was responsible for a redefinition.

[14] I admit that the talk about intentions and the most plausible solutions may be somewhat loose. A reformulation of Dennett's idea of an ideal agent could be used to make the talk rather more rigorous. Dennett (1971) suggests that we can predict someone's behaviour by treating her as an ideally rational agent who uses the best means to attain her aims. Similarly, we can define the most plausible modulation as the one which best suits the aims of the community, where aims are a result of general consensus. In

In general, we can distinguish two types of modulations on the basis of the intentions of the speakers. On the one hand, we have modulations that are intended to serve personal aims, with no intention of attaining a plausible consensus with other speakers. On the other hand, we have modulations that aim at plausible solutions with regard to generally acceptable objectives. We can call the first kind ad hoc modulations and the second kind binding modulations. From the perspective of Peregrin's inferentialism, we can say that the ad hoc modulations are not meaning-constitutive, while the binding modulations are meaning-constitutive – only binding modulations are full-fledged *meaning* modulations.

However, the intentions of speakers are important for the distinction between modulations only because they lead to a difference in the expected consequences in the behaviour of the speakers. Since the litigants in ad hoc modulations follow particular personal aims, we can expect that even a speaker who enforces a modulation will not be consistent in the use of the expression when compared to her past and future conversations. As a result of achieving her aim, a speaker has no reason to follow the modulation any more. Moreover, a speaker may not follow the modulation in the context of her different aims. Since other speakers do not expect that a speaker will follow the modulation, they do not have any reason to adjust and apply their own corrective behaviour so as to be in accordance with the modulation in future conversations as well.

On the other hand, the reasons, which led to the decision about Pluto, are a result of a debate with regard to generally acceptable objectives. Binding modulations are intended from the outset to settle a *widely shared* consensus followed by a majority of a linguistic community in the future and this means that some individuals have to adapt from time to time. However, since the outcome of binding modulations is supposed to be generally acceptable, we have good reason to assume that most speakers will systematically follow the modulation in future conversations, regardless of their initial position in meaning litigation.

---

the case of Pluto, the decision to declassify it was the most plausible solution with regard to more aims – it avoids a possible extreme enlargement of the number of planets and it allows a more rigorous definition of a 'planet' and a more accurate classification of objects in our solar system in general.

b) *The number of participants in a modulation*. Another notable difference between ad hoc and binding modulations lies in the number of speakers who participate in the modulations. Ad hoc modulations are usually incidental and appear in small groups of people, even in one-to-one conversations very often. On the other hand, binding modulations are usually open to all the speakers of a relevant linguistic community. In the case of Pluto, a part of the astronomical community decided which modulation would be in use, but the discussion was open to non-experts as well. Even small children sent letters to Dr. de Grasse Tyson. It does not matter whether their opinion was taken into consideration or not. What is most important is that they took part in the litigation and, by doing so, they designated themselves as members of a relevant community to which the litigation – and its result – applies. This is an important point when compared to ad hoc modulations. If Lorelai refuses to follow the modulation proposed by Emily, her status as a member of any linguistic community will not be harmed in any sense regardless of the fact that she took part in the litigation. Nevertheless, someone's refusal to follow the decision about Pluto can be seen as a reason for the enforcement of corrective practices and, in an extreme case, a reason for her detachment from a linguistic community.

Moreover, the example of Pluto is a rather specific binding modulation. The term 'planet' belongs almost exclusively to astronomy and so astronomers have some semantic authority in litigations. This is why the opinion of non-experts was not taken seriously. However, there are many examples in which the authority is not so clear and the role of "vox populi" is much bigger. This is the case of words such as 'marriage' or 'person'. These words became the centre of attention as they appeared in the press, at academic conferences, and in courtrooms. But the question whether an unborn child is a person is not only a legal, medical, or religious matter. It is, first and foremost, a social matter. The results of these modulations will directly influence the everyday lives of many people and therefore a wide public discussion plays an important role in the final decision.

c) *Information flow and its general accessibility*. Since 2001, the New York Times has published more than twenty articles and commen-

taries on the topic of Pluto's demotion, most of them written by the journalists Kenneth Chang (KC) and Dennis Overbye (DO). This is only a short list:

Jan. 29, 2002	"Planet or No, It's On to Pluto" (KC)
Jul. 30, 2005	"Planet or Not, Pluto Now Has Far-Out Rival" (KC/DO)
Oct. 4, 2005	"9 Planet? 12? What's a Planet, Anyway? (DO)
Feb. 2, 2006	"Icy Ball Larger Than Pluto. So, Is It a Planet?" (KC)
Aug. 16, 2006	"For Now, Pluto Holds Its Place in Solar System" (DO)
Aug. 22, 2006	"Pluto Seems Poised to Lose Its Planet Status" (DO)
Aug. 24, 2006	"Pluto Is Demoted to Being a Dwarf Planet" (DO)
Aug. 25, 2006	"Vote Makes It Official: Pluto Isn't What Is Used to Be" (DO)
Aug. 25, 2006	"And Now There Are Eight" (Editorial)
Sep. 1, 2006	"Debate Lingers Over Definition for a Planet" (KC)
Dec. 24, 2006	"Dwarf Planet" (DO)
Jun. 12, 2008	"Not a Planet, but a Plutoid" (KC)
Jan. 12, 2009	"How Many Planets Do You Want in the Solar System?" (KC).[15]

The focus here is on the New York Times because it is one of the most influential newspapers in the world, but it is basically arbitrary. We can find a similar list of articles about Pluto in practically any newspaper. The interest of journalists in the topic caused an extensive information flow, which ensured that the information about the current status of Pluto (and so about the current state of the meaning of 'planet') was distributed among the members of the relevant linguistic community. This is hardly an accidental feature. Of course, even an ad hoc modulation *can* exceptionally become the centre of attention. However, this fact does not change the main point: an extensive information flow is an important component of meaning-constitutive modulations because it creates favourable conditions for a

[15] It is worth noticing that the opinions of the New York Times journalists changed radically as the discussion proceeded. The journalists adopted the view of the International Astronomical Union without much hesitation, despite their initial criticism.

distribution of the new meaning within a relevant linguistic community. A modulation cannot become widely shared if it is not generally accessible by a majority of speakers.

When combined, the three points related to binding modulations (the intentions of speakers to follow a modulation, a large number of participants in a modulation, and an extensive information flow) constitute ideal conditions for their results to become widely shared and so to establish a widely shared corrective behaviour with regard to a particular set of inferences. Of course, the fulfilment of these conditions does not necessarily guarantee that the new meaning will be adopted and we can easily find borderline situations. This is, for example, the case of the word 'polyarchy', promoted by Robert Dahl within the field of political science. Dahl (1956, 1971, 1984) argued that the contemporary political system in the USA is not democracy, but polyarchy. Democracy is a system in which all the citizens are considered to be equal in political decisions, while in polyarchy control over governmental decisions is constitutionally vested in elected officials. While his distinction was well known, globally discussed, and later on generally accepted in the field, the word 'polyarchy' has never replaced the word 'democracy' within the "linguistic community of political scientists". It is hard to say why this was the case. A possible explanation might be that there was no need to start using the new word because political scientists in 1956 knew very well that 'democracy' did not mean anymore what it used to mean in Ancient Greece. The meaning of the word 'democracy' has changed with emerging modern republics and so there was no need to adopt 'polyarchy'.

However, even if 'polyarchy' is an example of an unsuccessful modulation, I do not think that the existence of borderline cases causes any problems in our current context. What is sufficient for the purpose of supporting Peregrin's criterion is that there are at least some examples in which binding modulations were successfully adopted by a community. The existence of such examples shows that there are general mechanisms for establishing new meanings on the level of whole communities – even though they might fail from time to time. Mechanisms that are related to binding modulations are exactly those mechanisms that guarantee that there is a widely shared corrective behaviour for particular inferences and so the inferences correspond to inferential rules.

4. Conclusion

The conclusion of this paper may seem rather subtle: an outcome of a meaning modulation can hardly become widely shared if a majority of speakers does not know about the modulation, if the speakers do not take part in it, and if the speakers do not intend to follow it. However, the mere fact that there are modulations the outcomes of which are widely shared has interesting consequences for the discussion of meaning-constitutive inferences. It shows that Peregrin's criterion of meaning-constitutiveness can be empirically supported – that there are social mechanisms thanks to which a set of inferences corresponding to inferential rules can become widely shared.

I agree that this is not exactly what Fodor and Lepore had in mind when they discussed the analytic–synthetic distinction. Peregrin is not able to give fixed and finite lists of analytic and synthetic inferences. Nevertheless, actually, this can be seen as an advantage. Peregrin can get rid of the analytic–synthetic distinction in its traditional (problematic) form. His criterion does not depend on any "intrinsic" semantic properties of sentences/words and so it is not circular in defining semantic properties and analyticity.[16] On the contrary, the criterion for meaning-constitutiveness based on the notion of widely shared corrective behaviour follows the dynamics of natural languages and this is a feature worth keeping.

Acknowledgments

The research leading to this work was funded by The Specific Research project "Cooperative aspects in human communication, cognition and emotion" at the Faculty of Arts, University of Hradec Králové (2016).

References

BOGHOSSIAN, P. (1993): Does an Inferential Role Semantics Rest upon a Mistake? In: Villanueva E. (ed.): *Philosophical Issues 3*. Atascadero: Ridgeview Press, 73-88.

[16] In other words, it does not define semantic properties as those which correspond to analytic inferences and analytic inferences as those which are part of meaning.

BOGHOSSIAN, P. (2003): Blind Reasoning. *Proceedings of the Aristotelian Society* 77, No. 1, 225-248.

BOGHOSSIAN, P. (2012): Inferentialism and the Epistemology of Logic: Reflections on Casalegno and Williamson. *Dialectica* 66, No. 2, 221-236.

BRANDOM, R. (1994): *Making It Explicit: Reasoning, Representing, and Discursive Commitment.* Cambridge, MA: Harvard University Press.

BRANDOM, R. (2000): *Articulating Reasons: An Introduction to Inferentialism.* Cambridge, MA: Harvard University Press.

BRANDOM, R. (2007): Inferentialism and Some of Its Challenges. *Philosophy and Phenomenological Research* 74, No. 3, 651-676.

DAHL, R. (1956): *A Preface to Democratic Theory.* Chicago: The University of Chicago Press.

DAHL, R. (1971): *Polyarchy: Participation and Opposition.* New Haven: Yale University Press.

DAHL, R. (1984): Polyarchy, Pluralism, and Scale. *Scandinavian Political Studies* 7, No. 4, 225-240.

DAVIDSON, D. (1986): A Nice Derangement of Epitaphs. In: Lepore E. (ed.): *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Cambridge: Blackwell, 433-446.

DAVIDSON, D. (1994): The Social Aspect of Language. In: McGuinness, B. & Oliveri, G. (eds.): *The Philosophy of Michael Dummett*. Dordrecht: Kluwer Academic Publishers, 1-16.

DENNETT, D. (1971): *The Intentional Stance*. Cambridge, MA: MIT Press.

FODOR, J. & LEPORE, E. (1999): All at Sea in Semantic Space: Churchland on Meaning Similarity. *Journal of Philosophy* 96, No. 8, 381-403.

FODOR, J. & LEPORE, E. (2001): Brandom's Burdens: Compositionality and Inferentialism. *Philosophy and Phenomenological Research* 63, No. 2, 465-481.

FODOR, J. & LEPORE, E. (2007): Brandom Beleaguered. *Philosophy and Phenomenological Research* 74, No. 3, 677-691.

KOREŇ, L. (2017a): Peregrinova inferencialistická koncepcia jazyka a významu (1. časť). *Filosofický časopis* 65, No.1.

KOREŇ, L. (2017b): Peregrinova inferencialistická koncepcia jazyka a významu (2. časť). Forthcoming in *Filosofický časopis* 65.

LEWIS, D. (1969): *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

LUDLOW, P. (2014): *Living Words: Meaning Underdetermination and the Dynamic Lexicon.* Oxford: Oxford University Press.

PEREGRIN, J. (2006): Meaning as an Inferential Role. *Erkenntnis* 64, No. 1, 1-35.

PEREGRIN, J. (2010): Inferentializing Semantics. *Journal of Philosophical Logic* 39, No. 3, 255-274.

PEREGRIN, J. (2012): Inferentialism and the Normativity of Meaning. *Philosophia* 40, No. 1, 75-97.
PEREGRIN, J. (2014a): Implicitní pravidla. *Organon F* 21, No. 3, 381-398.
PEREGRIN, J. (2014b): *Inferentialism: Why Rules Matter.* London: Palgrave Macmillan.
QUINE, W. V. O. (1951): Two Dogmas of Empiricism. *The Philosophical Review* 60, No. 1, 20-43.
SCHWARZ, W. (2013): Against Magnetism. *Australasian Journal of Philosophy* 92, No. 1, 17-36.
SELLARS, W. (1949): Language, Rules, and Behavior. In: Hook S. (ed.): *John Dewey: Philosopher of Science and Freedom*. New York: Dial Press, 289-315.
SHAPIRO, L. (2004): Brandom on the Normativity of Meaning. *Philosophy and Phenomenological Research* 68, No. 1, 141-160.
SIDNELL, T. (2010): *Conversation Analysis: An Introduction.* New York: Wiley-Blackwell.
WEINTRAUB, D. (2007): *Is Pluto a Planet?* Princeton: Princeton University Press.
WITTGENSTEIN, L. (1953): *Philosophical Investigations*. Oxford: Blackwell.

Organon F 24 (1) 2017: 105-113

In Defence of Δ-TIL[1]

DANIELA GLAVANIČOVÁ

In 2015-2016, my two papers on deontic modalities analysed in terms of Transparent Intensional Logic (TIL) were published in *Organon F*. The first of them, Glavaničová (2015), is based on the results of my bachelor thesis. This paper stands at the beginning of my (ongoing) research into deontic logic. The second one, Glavaničová (2016), suggests a small amendment to the analysis provided in the first paper.

In short, I have argued (in the first paper) that deontic logic should be hyperintensional, since deontic propositions lead to the failure of the extensionality principle (we cannot always inter-substitute necessary equivalents into deontic formulas). The framework I used was Tichý's Transparent Intensional Logic (hyperintensional partial lambda calculus with types). The suggested analysis consists mostly in providing type-theoretical analysis along with truth-conditions for deontic propositions and some axioms and inferential rules.

Moreover, the semantic distinction between implicit and explicit deontic modalities was introduced. Informally, consider some normative text (e.g., Decalogue) and a normative sentence that is explicitly contained in that text (e.g., "Thou shalt not steal"). Suppose, for the sake of argument, that this normative sentence can be translated into deontic sentence

(1) It is obligatory that people do not steal.

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

There is an *explicit* deontic construction that can be assigned to this sentence as an analysis. Now there are many deontic sentences that are not "translations" of any of the sentences explicitly contained in the Decalogue, for instance

(2) It is obligatory that people do not dream

is such sentence, but also

(3) It is obligatory that people do not steal or dream
(4) It is obligatory that people do not steal and that 2+2=4
(5) It is obligatory that people do not steal and that bachelor is an unmarried man.

Now the sentence (2) denotes the proposition that is not entailed by the proposition denoted by (1), so (2) is not even implied by the Decalogue. However, the propositions denoted by (3), (4) and (5) are implied by the Decalogue, because they are implied by the proposition denoted by (1). These implied consequences of something explicitly stated are what I call *implicit* deontic propositions.

I attempted to show that this distinction can be useful in resolving some of the paradoxes of deontic logic. This approach to resolving the deontic paradoxes was inspired by a similar approach in epistemic logic (cf. Levesque's 1984 'Logic of Implicit and Explicit Beliefs').

The ideas that deontic modals fail to be extensional and that hyperintensional deontic logic can be useful in resolving deontic paradoxes occurred to me during the summer of 2013, and today, more than three years later, I still hold these beliefs, and I am ready to defend them.[2] Of course, I don't think that I provided a comprehensive, satisfactory account: it is still a work in progress.

Recently, Vladimír Svoboda (2016) responded to my papers with a sharp criticism. I agree with many of his points. Indeed, I was aware of most of them even before reading the criticism. This is so because I have been intensively discussing my approach to deontic logic with various relevant researchers (including Svoboda himself) since writing the bachelor

[2] Recently, Faroldi (2016) defended these claims, though his approach is different from mine.

thesis. Despite this, his criticism contains many points I cannot agree with, and these will be central to this defence. Yet, one should keep in mind (while reading the papers or this defence) that both of the papers are to be understood as first steps in my long-term project of developing a hyperintensional deontic logic, and neither of them as the final proposal.

To begin with, Svoboda explains my approach. It is not surprising that he is not sympathetic to my limitation of deontic logic to (descriptive) logic of deontic propositions, since he himself advocates prescriptive deontic logic. He writes that "when Glavaničová speaks about deontic logic what she has in mind is *narrowly* conceived deontic logic" (Svoboda 2016, 540; italics mine); and that it is "not clear what is meant by the phrase 'implicit command described by the sentence'" (Svoboda 2016, 544). Yet this narrowly conceived deontic logic seems to be prevalent in the current literature, so I do not feel guilty of making this common simplification. Moreover, the phrase he quotes occurred within an informal explanation, not as a part of some definition. So I do not, and need not, presuppose the existence of any logic of commands, contrary to what Svoboda suggests in his criticism (though, I think that there can be one). All that is needed is some informal understanding of the term *command* (similarly for the other undefined terms, such as *sentence*).

Second, Svoboda (2016, 541) claims that the sentence "It is obligatory that Pavel is silent" would be analysed in Δ-TIL either as

$$[{}^0O_{wt}\ [\lambda w \lambda t\ [{}^0Silent_{wt}\ {}^0Pavel]]]$$

or as

$$[{}^0O^*_{wt}\ {}^0[\lambda w \lambda t\ [{}^0Silent_{wt}\ {}^0Pavel]]].$$

I almost agree, but there is a missing element "$\lambda w \lambda t$" at the beginning of the both constructions.

Subsequently, Svoboda discusses my definitions:

$${}^0T : [{}^0O_{wt}\ C] \text{ iff } C \in O_{wt}$$
$${}^0F : [{}^0O_{wt}\ C] \text{ otherwise.}$$
$${}^0T : [{}^0O^*_{wt}\ {}^0C] \text{ iff } {}^0C \in O^*_{wt}$$
$${}^0F : [{}^0O^*_{wt}\ {}^0C] \text{ otherwise.}$$

(The crucial types are $O/(\mathrm{oo}_{\tau\omega})_{\tau\omega}$ and $O^*/(\mathrm{o}^*\mathrm{n})_{\tau\omega}$.)

He writes: "The definitions, in fact, seem somewhat suspicious to me. I, for example don't see how a *construction* of a proposition could be a member of the set of (in this case obligatory) propositions" (Svoboda 2016, 541, footnote 3). The construction in question is not mentioned, but *used* (written *without* trivialization). This means that the semantic content is the proposition constructed, not the construction itself (and this proposition is a member of the set of obligatory propositions). Yet, I agree that I could made it more clearly.

Next, Svoboda claims that "these definitions would be entirely uninteresting if they were not supplemented by some logical principles" (Svoboda 2016, 542). Though I provided some logical principles, this declaration seems to be too strong. A definition (or explication, analysis) can be theoretically interesting even without adding any logical principles. Claiming otherwise is according to me a "narrowly conceived logic": For instance, some definition or some explication can enlighten the relationship between some important concepts. To give an argument for this claim, it is needed to explain the notion of theoretical explication first. My understanding of theoretical explication derives from Tichý (1988, 194-195):

> To explicate a system of intuitive, pre-theoretical, notions is to assign to them, as surrogates, members of the functional hierarchy over a definite objectual base. Relations between the intuitive notions are then represented by the mathematically rigorous relationships between the functional surrogates.

With this understanding of explication in mind, let me now demonstrate that, for instance, my definition

$^0T : [^0O_{wt}\ C]$ iff $C \in O^*_{wt}$
$^0F : [^0O_{wt}\ C]$ otherwise

connected with the type-theoretical analysis of the operator O, i.e. type-theoretical analysis of the "coarse-grained oughts" has some explanatory value. Recall that the type of O is $(\mathrm{oo}_{\tau\omega})_{\tau\omega}$, that is $(((\mathrm{o}((\mathrm{o}\tau)\omega))\tau)\omega)$ unabbreviated. The last "ω", for example, captures the *modal variability* of oughts: What is obligatory differs with respect to possible worlds. In other

words, what is in fact obligatory is usually not obligatory as a matter of logical necessity. If talking about normative systems, one can say that a normative system can change. Next to the last type is "τ", which captures the *temporal variability* of oughts: What is obligatory differs with respect to time. In other words, what is currently obligatory is usually not eternally obligatory. If talking about normative systems, one can say that a normative system in fact changes as time goes by. The type $(o((o\tau)\omega))$ stands for the set of "coarse-grained propositions", so coarse-grained oughts are analysed in terms of coarse-grained propositions and this is captured by the resulting type $(oo_{\tau\omega})_{\tau\omega}$. Similarly for the second part of the definition, or, in general, for the explanatory value of the type-theoretical analysis.

After a concise exposition of my approach, Svoboda starts to assess it. Obviously, he is not satisfied with my way of *testing* the framework: "Somewhat surprisingly, the whole testing consists in a discussion of how the inferential scheme called the *Ross paradox* fares with respect to her distinction between implicit and explicit obligation" (Svoboda 2016, 542-543). Yet, more testing is suggested in footnote 12 of my paper (see Glavaničová 2015, 224), and some generalization is provided by its conclusion (see Glavaničová 2015, 226-227). However, I agree with Svoboda that my presentation of the Ross paradox was *quite misleading*, because the *intuitively invalid* entailment is, of course, the entailment from *It is obligatory that Pavel is silent* to *It is obligatory that Pavel is silent or kills Richard*. Moreover, Svoboda seems to be suspicious of the relevance of The Ross Paradox to descriptive deontic logic:

> The original version of Ross' paradox was presented in the form of the inference *Mail this letter!,* hence *Mail this letter or burn it!*, which was valid according to the prevailing accounts of the logic of imperatives (cf. Ross 1941). It was thus not straightforwardly relevant for *statements* about obligations. (Svoboda 2016, 543)

Contrary to Svoboda, I think that the paradox is relevant to descriptive deontic logic as well as to prescriptive deontic logic (indeed, it is often discussed within the current descriptive deontic logic). It matters little that it was originally formulated within the latter.

Next, I do not see any reason why the distinction between implicit and explicit should be problematic in the moral discourse, and Svoboda does

not provide us with any argument against it (it just "seems quite strange" to him, see Svoboda 2016, 544). It does not seem strange to me: We can talk about moral codes, or about the Decalogue, or about some codes of conduct of some organization, and so on. These usually exist in a written form, so it makes sense to talk about their explicit content, and surely, it also makes sense to talk about their implicit consequences.

One of the important issues under discussion is my use of the implicit-explicit distinction in resolving deontic paradox(es). Svoboda finds it problematic. The main idea is that if we distinguish between deontic propositions that are just propositional contents of some explicitly given commands/permissions prefixed with appropriate deontic operator, and their logical consequences, we wouldn't be misled by these logical consequences (in the sense of reading into them more that is in fact provided by these consequences). For instance, suppose we have a command "Set the prisoner free!"; the relevant explicit deontic proposition would then be "It is obligatory that the prisoner is set free". Now we can derive a logical consequence "It is obligatory that the prisoner is set free or executed". However, if we keep in mind the distinction between explicit and implicit deontic modals/deontic propositions, it is clear that (i) explicit obligations are of utmost importance and are to be preferred over their implicit (implied) consequences; (ii) the implicit deontic proposition "It is obligatory that the prisoner is set free or executed" was derived from the explicit deontic proposition "It is obligatory that the prisoner is set free" only because "or" in the latter is a just the non-exclusive disjunction, so we cannot read it as anything else (not in the least as the free choice disjunction); (iii) explicit deontic propositions correspond to explicitly given commands/permissions; and (iv) if there is an explicitly given command/permission, one should not follow some implicit consequence that would made it impossible to obey the given command.

The remark that Δ-TIL is a weak logic is not very surprising, since virtually every system of hyperintensional logic faces this problem: The more *fine-grained* the meanings are, the weaker the logic is. Creating satisfactory hyperintensional logic is a balancing act; cf. e.g. Mark Jago's *problem of bounded rationality*: real agents are rational but at the same time cognitively bounded (see Jago 2014a, 163-192). Put differently "[i]t seems that (i) rational agents seemingly know the trivial consequences of what they know, but (ii) they do not know all logical consequences of what they

know. The problem of rational knowledge is that (i) and (ii) are incompatible" (Jago 2014b, 1152). To account for this problem, one has to devise a hyperintensional logic that can master the abovementioned balancing act. Glavaničová (2015) uses a weak logic for explicit deontic modals and strong logic for implicit ones. Semi-implicit and semi-explicit deontic modals were introduced exactly to proceed to this balancing act. However, it is still a work in progress. In fact, the aim of the paper was simply to propose an analysis (or explication) of deontic modals and to motivate it. Recall that I understand explication, in accordance with Tichý, as providing us with a type-theoretical analysis. Of course, that does not mean that I don't have intentions to create full-blooded (hyperintensional) deontic logic. Hopefully, I will develop such a(n) (more satisfactory) account in the upcoming years.

Svoboda's criticism contains also a short discussion of my second paper, Glavaničová (2016). I don't agree with Svoboda's suspicion that my "argumentation in favour of deontic relativism (a significant part of the second article consists of this argumentation) appears to be close to trivial and the adherents of objectivism that she mentions appear to be mere straw men" (Svoboda 2016, 547). It is true that I just took the position from MacFarlane's discussion in his (2014) book on relativism, merely presupposing that it may be interesting to argue against the position. Despite this, it seems to me that the position is quite common in the metaethical literature. It was not clear to me whether Svoboda thinks that it is not worthwhile to argue for deontic relativism at all, or just does not find my arguments compelling. Be it as it may, deontic relativism seems natural and obvious to me, mostly because the truth-conditions of deontic propositions vary with respect to normative systems. Despite the appropriateness of deontic relativism, there are many researchers who ignore relativistic nature of deontic propositions, or explicitly argue for other positions than relativism.

Next, Svoboda thinks that using relativism to avoid Chisholm's paradox is a bizarre solution, though the inconsistency is avoided (as it should be). Surely, to be philosophically plausible, one needs to say more about this solution. My idea was that some sort of deontic relativism can mimic the distinction between "categorical oughts" and some "corrective oughts" that presuppose the violation of some categorical oughts. For instance, it is categorically forbidden to murder people; the ideal state is that there are no murders, no murderers. Yet, such categorical oughts are often violated and

in such cases, corrective oughts come into the play. For instance, if the categorical forbiddance of murder was violated, there is a corrective ought that the murderer should be sent to the jail. If the solution is understood in this way, it is not that far from many standard solutions (see "Multiple operators" and "Prima facie and all-things-considered oughts" in Goble 2013, 251-296).

In footnote 19, Svoboda claims:

> The terms 'explicit attitudes' and 'implicit attitudes' were introduced as technical terms in TIL (probably in Duží 2004), and this terminology is probably not ideal. It would be, perhaps, more suitable (though less concise) to distinguish between 'attitudes to the construction of an object' and 'attitudes to the constructed object' or 'coarse-grained attitudes' and 'fine-grained grained attitudes'. (Svoboda 2016, 549)

I do not think that the *explicit – implicit* terminology used in current TIL is "not ideal" (Svoboda 2016, 549). I think that these terms are really apt, because the terminology is used in epistemic logic at least since the official introduction by Levesque (1984). This distinction appears in contemporary deontic logic (informally; see, e.g., Hansen 2013, 159-160, 164 or Hansson 2013, 201-204). Furthermore, the terms explicit and implicit have their intuitive meanings that are useful for an informal explanation of the distinction. Finally, the distinction was used in TIL already in Duží & Materna (2001), so it was not introduced in Duží (2004).

Acknowledgments

I would like to thank Federico L. G. Faroldi and Marián Zouhar for helpful comments.

References

DUŽÍ, M. & MATERNA, P. (2001): Propositional Attitudes Revised. In: Childers, T. (ed.): *The Logica Yearbook 2000*. Praha: Filosofia, 163-173.

DUŽÍ, M. (2004): Intensional Logic and the Irreducible Contrast between de dicto and de re. *ProFil* 5, No. 1, 1-34, http://profil.muni.cz/01_2004/duzi_de_dicto_de_re.pdf.

Faroldi, F. L. G. (2016): *Deontic Modals are Hyperintensional*. Manuscript.

Glavaničová, D. (2015): K analýze deontických modalít v Transparentnej intenzionálnej logike. [On the Analysis of Deontic Modalities in Transparent Intensional Logic.] *Organon F* 22, No. 2, 211-228.

Glavaničová, D. (2016): Δ-TIL and Normative Systems. *Organon F* 23, No. 2, 204-221.

Goble, L. (2013): Prima Facie Norms, Normative Conflicts, and Dilemmas. In: Gabbay, D. M. et al. (eds.): *Handbook of Deontic Logic and Normative Systems*. College Publications, 241-351.

Hansen, J. (2013): Imperative Logic and Its Problems. In: Gabbay, D. M. et al. (eds.): *Handbook of Deontic Logic and Normative Systems*. College Publications, 137-191.

Hansson, S. O. (2013): The Varieties of Permission. In: Gabbay, D. M. et al. (eds.): *Handbook of Deontic Logic and Normative Systems*. College Publications, 195-240.

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

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

Levesque, H. J. (1984): A Logic of Implicit and Explicit Belief. In: *Proceedings of the National Conference on Artificial Intelligence*. Cambridge: AAAI Press/MIT Press, 198-202.

MacFarlane, J. (2014). *Assessment Sensitivity: Relative Truth and its Applications*. Oxford: Clarendon Press.

Svoboda, V. (2013): *Logika pro Pány, Otroky a Kibice. Filosofický průvodce světem deontické logiky*. [*Logic for Masters, Slaves and Kibitzers. A Philosophical Guide through the World of Deontic Logic*.] Praha: Filosofia.

Svoboda, V. (2016): Δ-TIL and Problems of Deontic Logic. *Organon F* 23, No. 4, 539-550.

Tichý, P. (1988): *The Foundations of Frege's Logic*. Berlin – New York: de Gruyter.

Organon F 24 (1) 2017: 114-134

Giacomo Borbone & Krzysztof Brzechczyn (eds.): *Idealization XIV: Models in Science* Brill, Leiden, 2016, 318 pages[1]

This latest addition to the Poznań Studies brings together thirteen contributions by fourteen authors, preceded by a foreword of the editors. As is tradition, the scope of the papers ranges from general problems in the philosophy of science to case studies from within particular disciplines, including sociology, historiography, economics and philosophy. The general tone of the volume is set by the subtitle: all of the studies deal, to varying degrees, with the relation between idealization and modeling, where the former is mostly understood in line with the so-called idealizational theory of science developed by the Poznań School centered around Leszek Nowak.

Reviewing a collection of studies on such a broad selection of topics can be demanding. To make this task easier for myself, I will proceed as follows. First, I shall briefly summarize the contents of each of the papers and, where possible, provide more detailed comments on topics related to my own areas of competence. I will conclude with some general observations on the collection.

The papers

The volume is divided into three parts, the first of which contains four papers dedicated to "General Problems" of idealization and modeling. The opening paper, by Xavier de Donato Rodríguez and José L. Falguera, applies Zalta's well-known theory of abstract objects to scientific theories and the theoretical entities referred to by theoretical terms. The authors propose to view theories as a specific kind of "stories", i.e., abstract objects, which (i) were deliberately authored (in this case, by members of the scientific community) and (ii) encode only propositional properties. They are distinguished from other kinds of stories (e.g., literary fiction) both by the fact that they contain generalizations and that there exists a non-abstract domain to which the "story" is intended by the scientific

[1] ✉ Juraj Halas
Department of Logic and Methodology of Sciences, Comenius University in Bratislava
Gondova 2, 814 99 Bratislava, Slovak Republic
e-mail: juraj.halas@uniba.sk

community to apply. Similarly, theoretical entities are viewed as abstract objects which do not exemplify existence (or any other properties of non-abstract objects) but which may encode certain properties exemplified by real-world, non-abstract objects.

However, the authors do not always abide by the crucial distinction between exemplifying and encoding. For example, on p. 36, they write that "the ideal gas is just an abstract object exemplifying certain properties attributed to it in the kinetic theory of gases", e. g., the property of *being composed of perfectly spherical particles*. However, as an abstract object, the ideal gas surely does not exemplify any such properties, encoding them instead.

Notwithstanding such minor issues, De Donato Rodríguez and Falguera's paper presents a promising framework for further thinking about models and the method of idealization, not least for its expressiveness. Among the notions that the authors attempt to explicate is that of the degree of idealization of a proposition. According to them, a proposition q is more idealized than p if the number of "defeaters" of p (propositions that are incompatible with p) is less than the number of defeaters of q. The definition (p. 35) apparently requires further work, as there seem to be an infinite number of defeaters in either case (if r is a defeater of p, then the proposition $r \vee s$ is also a defeater of p etc.).

Igor Hanzel's study is based on a critique of Nowak's reconstruction of Marx's *Capital*, as well as on an original analysis of Newton's *Principia*. It argues for a distinction between three types of scientific laws – the "pure idealized type" (of scientific law), the "inherent type" and the "inherent idealized type". Hanzel also proposes a typology of measures divided into "external measure", "immanent measure" and "manifestation of the immanent measure of the ground's cause", which is related to the threefold classification of laws.

The antecedents of the "inherent type" and the "inherent idealized type" involve the so-called "inherent conditions" which, according to Hanzel, necessitate the existence of the underlying cause (principal factor, "ground") itself. The knowledge of these conditions, Hanzel argues, enables two specific kinds of inference unrecognized by Nowak: in the first case, the inference from the knowledge of the origins of the principal factor to the characteristics of the principal factor itself, and in the second case, the derivation of phenomena (possibly including "new" ones, as yet unrecognized) from the principal factor that generates them.

However, the analysis is not completely satisfactory on both technical and textual grounds. With respect to the former, some crucial pieces of the puzzle are treated in a rather cursory way. For example, Hanzel argues that the expression "$E^{(k)} = f_k(H)$" ("the phenomenon E in its k-th degree of idealization is functionally

dependent on the principal factor H") in the consequent of a "pure idealized type of scientific law" cannot be simply "turned around", so that the left-hand side of the equation is swapped with its right-hand side (p. 48-49). But then the expression clearly is *not* an equation, and the use of the equals sign is misplaced. Similarly, the symbol $\rightarrow_n$ appears in both of the "inherent" kinds of laws. Hanzel characterizes it as a sentential connective meaning "if …, then necessarily comes into being" (p. 52). Again, the semantics of the symbol is left unspecified. One is led to doubt whether – given its characterization ("…*comes into being*") – it indeed is a *sentential* connective. Finally, the symbol "$\Rightarrow$" is introduced on p. 55 as a shorthand for "explanatory derivation", but the nature of this derivation is left undetermined.

Turning to textual issues, while I agree with the general drift of Hanzel's criticisms of Nowak's reconstruction of Marx's *Capital*, I think his proposals do not correspond to Marx's views all that more closely. Hanzel ascribes to Marx the view that the value of a commodity y produced in an enterprise owned by x ($V(y, x)$) depends on the socially necessary abstract labor expended on y in the enterprise owned by x ($L(y, x)$) (p. 52). However, the indexation of L by x is superfluous precisely because it is *social* labor that counts as value-determining (the same point applies to more complicated expressions involving surplus labor and surplus value on p. 55).

Moreover, according to Marx, the question to what extent a particular concrete labor is recognized as social is only ever settled *ex post facto*, in exchange. Therefore, the "inherent conditions" which, according to Hanzel, necessitate the transformation of products into commodities with value are not sufficient. For a product y to be a commodity and to have value, it is not enough that the enterprise producing y is privately owned by x and that x intends to exchange y for other products. In the extreme case, if y is never exchanged because it is not recognized socially (i.e., on the market) as useful, then there is no value of y to speak of, and indeed no social labor at all had actually been performed in x's enterprise. Hence, even though Hanzel emphasizes that L stands for abstract/social labor, the reconstruction ultimately ends up with something more akin to pre-Marxian labor theories of value where commodities are "impregnated" with value, once and for all, in the production process.

The paper by Lidia Godek deals with Max Weber's ideal types. Returning to Weber's original writings, Godek proposes a new reconstruction of his method of the construction and heuristic use of ideal types based on the Poznań idealizational framework. In contrast with the previous reconstruction within the same tradition, due to Izabella Nowakowa, she argues that Weber's principal method is that of positive potentialization – i.e., the counterfactual assignment of properties of maximum intensity to possible objects (ideal types).

Godek's paper is on the right track when it emphasizes the deficits of Nowakowa's analysis of ideal types. However, I think that it does not go far enough. Both authors miss the fact that when Weber discusses the construction of ideal types, he includes abstraction ("reduction" in Poznań parlance) and transcendentalization (the ascription of new properties) among the methods (Weber 1990, 30). Moreover, Godek's paper does not overcome the view that ideal types are chiefly classificatory instruments (p. 68). However, as noticed already by Hempel (1965), ideal types were intended as heuristic tools that should enable the explanation of social action. Godek provides no details about how this would work.[2]

The first part of the volume is brought to an end with Mieszko Ciesielski's paper on reduction. Ciesielski provides a case study, which tests the conception of reduction of idealized theories, originally developed by Katarzyna Paprzycka. The subjects of the test are the theory of a rational act and the theory of habitual-rational action. Ciesielski notes that on a strict approach, reduction between them is impossible. This leads him to weaken the conditions for reduction, arguing for a special treatment of theories in the humanities.

The focus of the volume's second part is "Idealization in the Social Sciences". Its five papers deal with economics, historiography and linguistics. Adolfo García de la Sienra's paper approaches the topic of models and idealization from a structuralist point of view. Using examples from economics, he shows how idealized models, via their concretization, are used to make empirical claims about real systems. De la Sienra's conception pays close attention to the distinction between a real system, a model system, the set-theoretical structure attached to the latter, a model of data ("empirical structure") and the "Gedankenkonkretum", which is the Marxian term for an initial representation of the target system.[3]

The paper by Łukasz Hardt develops an account of economic models as "believable worlds" which reconciles the view of models as isolations (Mäki, Nowak) with that of models as parallel realities or credible worlds (Sugden). On Hardt's account, economic models (such as Varian's model of sales which serves here as an illustration) are representations of mechanisms which provide us with justifiable

---

[2] For an attempt at reconstructing explanation based on ideal types, see Halas (2016).

[3] However, de la Sienra also makes the rather controversial claim that the representation of an economic agent as rational (in the sense of transitivity of preferences) is not idealized. However, empirical studies of consumer behavior show that transitivity is routinely violated in the real world. It is not clear from de la Sienra's paper why the counterfactual assumption of rationality does not count as a case of idealization.

beliefs about the real world. As such, they are not simply true or false, but are used to maximize truth and minimize falsity in a wider system of beliefs about the real world.

Adam Czerniak's study links the "fallacy of reification of idealization" in economics to the global financial crisis. The fallacy occurs when the concretization of highly idealized models is omitted and the model is applied in a crude, direct way to real-world phenomena. Czerniak discusses the technical problems faced by attempts to concretize value-at-risk (VaR) and dynamic stochastic general equilibrium (DSGE) models in finance, and points out three more general reasons for the prevalence of reification of idealization in economics: the close ties between economics and policy-making, the lack of controlled experiments in (macro-)economics and the absence of firm theoretical foundations comparable with those of physics or chemistry. As one of the possible ways out, Czerniak suggests closer interaction of mainstream economics with heterodox traditions.

One of the editors, Krzysztof Brzechczyn, contributed a paper of his own. It is concerned with the reconstruction of methods of comparative analysis in historiography using the instruments of the idealizational theory of science. The source material for the reconstruction is provided by Skocpol's *States and Social Revolutions*. Brzechczyn arrives at a classification of comparative methods into those that compare cases of different kinds ("contrast-oriented method") and those that focus on cases of the same kind ("parallel method"). In both cases, he argues, the goal is to identify the main factors influencing a magnitude of interest. Brzechczyn concludes that this identification is never purely "inductive" and is always determined, at least in part, by theoretical preconceptions.

The second part of the volume concludes with Barbara Konat's study of the use of idealization in Chomsky's generative grammar. Already in Nowak's earlier work, Chomsky was viewed – along with Galileo, Marx and Darwin – as a pioneer of idealization in his respective discipline.[4] Konat provides a more detailed justification of this claim, focusing on the assumption of the ideal speaker-hearer. She concludes that Chomsky is indeed the "Galileo of linguistics".

The four papers which form the third part focus on "Idealization in the Humanities" – namely, in philosophy (metaphysics, political philosophy), strategic studies (scenario planning) and history.

---

[4] Incidentally, I think this long-standing part of the Poznań School's web of belief is in need of revision. Marx was certainly not the first to use idealization in political economy, nor the first to reflect on its use methodologically. See, for example, the remarks made by John Stuart Mill in (1837), quoted in and discussed by Blaug (1992, 55-59).

Krzystof Kiedrowski analyzes the uses of idealization in Nowak's later project of "negativistic unitarian metaphysics". According to Nowak, this doctrine was itself constructed using the methods of idealization and concretization. Kiedrowski's paper refutes this claim and shows that the methods used are abstraction (the elimination of factors) and its converse, disabstraction (the re-introduction of factors). Given the complex (and perhaps overcomplicated) nature of Nowak's metaphysics, Kiedrowski's paper can be somewhat difficult to follow. However, the main message of Kiedrowski's paper, that there is much to be said in favor of abstraction and disabstraction as methods of theory construction, is commendable – including, I think, vis-à-vis Nowak's earlier project of reconstructing Marx's *Capital*.

Piotr Przybysz focuses on the role of idealization in Rawl's political philosophy. He reconstructs the idealizing assumptions involved in the "original position" and in Rawls' model of the person. He shows that the sequential introduction of the principles of justice can be seen as a process of concretization, i.e., the gradual elimination of idealizing assumptions. This leads him to view Rawls as yet another 20th century pioneer who introduced idealization into his discipline.

I do not find the parallel between Rawls and Galileo entirely convincing in the details. The methods of idealization and concretization, as discussed by Nowak, are concerned with quantitative assumptions about the (causal) influence of certain factors. On the other hand, the assumptions identified by Przybysz in Rawls are all qualitative, and rather inexact at that. The idealizational theory of science was at the outset formulated as a theory about how theories in *empirical science* are built, tested and used for explanatory purposes. The process of concretization was made dependent on empirical evidence about the phenomena. However, in a non-empirical enterprise like political philosophy, the criteria for concretization (e.g., approximation) would seem to have to be different. Przybysz simply presupposes that the concretized versions of Rawls' principles of justice are "more realistic". One is inclined to ask – more realistic on what standards, absent empirical testing?

The contribution by Zenonas Norkus discusses the role of idealization in scenario planning. It contains an interesting, albeit rather long review of the history of the field, including several examples. Norkus argues that scenarios involve a specific, "discursive" kind of idealization, which results in a stylized, hypothetical narrative about future developments based on the identification of certain key causal factors and their possible effects.

The final paper in the volume, by Piotr Szwochert, reviews and extends Brzechczyn's earlier contributions on the role of idealization in historical narration. The analysis of several examples leads him to distinguish two aspects of historical

narration, the "factographic" and the "persuasive", and to discuss the role of axiological assumptions in organizing the narrative.

Concluding remarks

Turning to the volume as a whole, a minor quibble has to do with its structuring. Of the four papers in the first part, only the first really deals with a general problem concerning modeling and idealization. The others approach the topic from the point of view of particular case studies (in physics, social science and philosophy) with less clear consequences for the general framework. As regards parts two and three, the underlying classification into social sciences and the humanities is not quite obvious: one paper dealing with historiography is located in the former part (Brzechczyn), while another in the latter (Szwochert). I should note that some of the papers would have benefited from stricter editing (e.g., Norkus' remark on aesthetics appears twice, verbatim, on p. 285 and p. 293).

Seven of the papers include a restatement of the basic principles of Nowak's idealizational theory. Although they differ stylistically, re-reading the elements of idealization does get tiresome after a while. Given that the tenets of Poznań School are already well established, perhaps the space would have been better used to extend the authors' own contributions. Nonetheless, the fourteenth volume of Idealization succeeds in showing that the tradition is alive, well, and fruitful as ever.

Juraj Halas

References

BLAUG, M. (1992): *The Methodology of Economics or How Economists Explain*. Cambridge: Cambridge University Press.

HALAS, J. (2016): Weber's Ideal Types and Idealization. *Filozofia nauki* 2016, No. 1, 5-26.

HEMPEL, C. G. (1965): Typological Methods in the Natural and the Social Sciences. In: Hempel, C. G.: *Aspects of Scientific Explanation and Other Essays in Philosophy of Science*. New York: The Free Press, 155-171.

WEBER, M. (1990): *Grundriss zu den Vorlesungen über allgemeine ("theoretische") Nationalökonomie*. Tübingen: J. C. B. Mohr.

Andrea Bianchi (ed.): *On Reference*
Oxford University Press, Oxford, 2015, 415 pages[1]

On Reference is a collection of eighteen new essays on topics related to *reference*, with relevance for philosophy of language, philosophy of mind and linguistics. The topics have, of course, been extensively discussed for years, and if nothing else this new collection – which covers a remarkable range of issues and questions discussed from a wide range of perspectives – will help the reader understand *why*. It should, however, be pointed out that *On Reference* is not an introductory book. Background knowledge of the central issues and arguments is assumed, and it is, even for those who have experience with the topics, sometimes tricky to situate particular contributions in the traditions to which they aim to contribute. The book is, however, invaluable for anyone interested in getting up to date on these issues.

The collection is divided into three parts – "The Nature of Reference", "Reference and Cognition", and "Reference and Semantics" – though the overlap is substantial and the allocation of articles to sections seems a bit arbitrary. Section II consists of only three articles, and although those articles are interesting it is hard not to notice some questions it does *not* cover, such as issues related to the nature of *singular thought*. Elsewhere, readers will look in vain for substantial discussions of e.g. the semantics of empty names (a few contributions touch on them) or the relationship between classical referential semantics and recent developments in discourse semantics, such as variabilist theories or theories that accommodate reference in discourse semantic frameworks (Cumming's article is to some extent an exception). There is – with some exceptions – also little explicit discussion of whether there are necessary metaphysical or epistemic criteria for referring, and no discussion of the *acquaintance* condition, which seems to have received renewed attention in recent years. On the other hand, perhaps the fact that even a substantial and comprehensive collection is forced to forgo discussions of some central questions should be taken as testament to how important the topic of *reference* is for contemporary philosophy, and how wide-ranging the implications.

The first section, "The Nature of Reference", adds up to a good overview of recent discussions. The first two chapters discuss two potential rehabilitations of non-referential views, perhaps most provocatively in Christopher Gauker's "The

[1] ✉ Fredrik Haraldsen
Institute of Philosophy, Slovak Academy of Sciences
Klemensova 19, 811 09 Bratislava, Slovak Republic
e-mail: fredrikhar@yahoo.no

Illusion of Semantic Reference". Gauker argues that there is no such thing as semantic reference that can be used to underpin a notion of *knowing what expression e refers to*; instead, he defends what amounts to a *skeptical solution* and attempts to spell out knowledge of meaning in terms of the social status we grant to someone when saying that she knows the meaning of a word. In "Reference and Theories of Meaning as Use" Diego Marconi attempts – rather successfully and thoroughly, if not entirely surprisingly – to show that Horwich's *use theory of meaning* cannot easily be made compatible with the types of meaning externalism that has become orthodoxy after Burge, Putnam, Kaplan and others.

Edouard Machery, Justin Sytsma and Max Deutsch's "Speaker's Reference and Cross-Cultural Semantics" provides new empirical data on cross- and intracultural intuitions concerning Kripke's famous *Gödel case*, offered as part of his *semantic argument* against descriptivism in *Naming and Necessity* (see Kripke 1980). The article presents results of five new experiments designed to circumvent worries raised over the findings reported in Machery et al.'s influential (2004) paper, in particular that the original experiments failed to distinguish intuitions concerning semantic meaning from intuitions concerning speaker meaning (cf. Ludwig 2007), and show that intuitions still exhibit striking cross-cultural variations when these worries are taken care of. I admit that I remain skeptical that the data really show anything particularly interesting about descriptivist accounts of reference fixing (which is what they seem to concern – not descriptivist views of semantic content). First, it is still entirely consistent with a causal-historical theory to hold that uses of "Gödel" in these cases refer to Schmidt. Second, the Gödel case requires a complex setup and relies on complex intuitions; surely, a well-designed test of descriptivism should rather start by testing intuitions about *simple* cases, or at least include a control question to ensure that the test subjects have understood the complex case properly (anecdotally, my own students often don't, and I see no evidence that such control questions were included). A test of Kripke's *Einstein case* – whether someone who associates only the description "inventor of the atomic bomb" with Einstein and says "Einstein invented the atomic bomb" is saying something false about Einstein or something true about someone else – would seem to fit the bill better.

The final three articles of Section 1 defend the referentialist picture. Indeed, they all – in different ways – argue against a current tendency to view thought as prior to reference, that referring with the use of a name is secondary to having the referent in mind. In "Reference without Cognition" Genoveva Martí argues against this "neo-cognitive" trend, according to which referring requires a "cognitive fix" but not necessarily mediation of linguistic rules, and discusses cases where linguistic conventions determine reference independently of users' cognitive fixes; the

paper is a convincing contribution to a central issue. A similar motivation guides Andrea Bianchi's "Repetition and Reference", which aims to construct a substantive *theory* of reference based on the *picture* offered in *Naming and Necessity*. Bianchi requires of the theory that it makes no non-eliminable reference to intentions; instead, he provides metaphysical grounding in terms of causally linked repetitions. Finally, Michael Devitt's "Should Proper Names Still Seem So Problematic?" offers a fine-tuning of his familiar non-Millian theory of proper name meaning, according to which the meaning of a name is (just) its causal mode of referring. Although he doesn't address the worry that his theory conflates the meaning of a term with a meta-semantic story of how it *came to have* that meaning, at least he shows that his theory can stand on its own as a theory of meaning – albeit with the worrisome consequence that the standing meaning of a name changes from occasion to occasion. It is interesting to note how (neo)Fregean – though not descriptivist – the view ultimately is, at least as it is laid out in the present article.

Section II deals with issues related to cognition. Antonio Capuano's "Thinking About an Individual" defends an "outside–in" view of cognition on which (natural) processes bring objects to mind, as opposed to more traditional "inside–out" views according to which cognition proceeds by mental representations. The picture ostensibly amounts to a more fundamental shift in perspective than externalism or anti-individualism, and much of the article is concerned with comparing it to Burge's anti-individualist view; I am less sure how to compare it to the (to some extent) more radical externalist positions defended by, say, McDowell or some central contributors to the phenomenological tradition. Marga Reimer, in "Drawing, Seeing, Referring: Reflections on Macbeth's Dagger", starts by noting that "drawing a dagger" is ambiguous between an ontic (derivative) reading, which entails the existence of a dagger, and a non-ontic (creative) reading, which does not, and argues that "referring" is ambiguous in the same manner. In the non-ontic case an abstract object is (perhaps unintentionally) created that can serve as the target of thought and reasoning. John Perry, in "The Cognitive Contribution of Names", argues that the "direct" cognitive contribution of a name – what is determined by semantics – is just how it *looks* or *sounds*, an observation that he uses to help explain the apparent cognitive significance of "Hesperus is Phosphorus." Ultimately, the view Perry defends is reminiscent of Frege's *Begriffsschrift* view, and he does a compelling effort to defuse central worries associated with this type of approach.

The theme of Section III seems a bit nebulous. It is supposed to cover various semantic problems related to reference, but I have trouble seeing by what principle articles are assigned to Section III rather than Section I. That said, the section does

contain some of the most interesting contributions to the volume. The discussion of predicativism is a case in point: An exchange between Robin Jeshion ("Names Not Predicates" and "A Rejoinder to Fara's '"Literal" Uses of Proper Names'") and Delia Fara ("'Literal' Uses of Proper Names"), as well as an article ("Names As Predicates?") by Ernesto Napoli. According to Napoli, predicativists must claim that a name *n* means *being a bearer of n*, where *n* is an expression (arbitrarily) assigned to individuals by a stipulation/baptism where it is used quotationally; however, such accounts face serious challenges insofar as they assume that a stipulation/baptism is necessary and sufficient for being named *n* (there are multiple counterexamples), and because assigning *n* to someone is not assigning *n* to the property of *being a bearer of n*.

Jeshion's target is the Uniformity Assumption, the predicativist's claim to be able to offer a unified account of predicative and referential uses of names. In particular, if a name *n* is true of *x* if and only if *x* is called *n*, there will, Jeshion argues, be many cases where the theory yields the wrong results, for instance when I truthfully say of my barber Joe Romanov that "Joe Romanov is not a Romanov" because he has no relation to the Russian dynasty. Fara, in her response, attributes to Jeshion the following argument: i) predicativists think all literal uses of names satisfy the *being-called* condition; ii) there are non-metaphorical uses of predicative proper names that do not satisfy the *being-called* condition; iii) so, there are literal uses of predicative proper names that do not satisfy the *being-called* condition and hence no unified analysis of literal uses of predicative proper names, which means that the predicativist is not better off than the referentialist. Fara rejects the assumption that all non-metaphorical uses are literal uses, and argues that most of Jeshion's examples concern non-metaphorical *and* non-literal uses. The Romanov cases remain problematic, however, and Fara argues that such examples involve proper nouns, not proper names. In her response, Jeshion points out that she didn't offer an argument *against* predicativism, but an argument showing that the predicativist's *own unification argument* doesn't hold up – in particular, it cannot be used to argue that predicativism is superior to referentialism. So for instance, while Fara can certainly argue that Romanov cases involve use of "Romanov" as a proper noun that doesn't need to satisfy the *being-called* condition, nothing stops the referentialist from saying the same about "there are three Alfreds in Princeton," and the referentialist's distinction seems no less arbitrary than the one Fara introduces. Hence, the predicativist's uniformity argument fails as an argument in favor of predicativism. Though the debate has become fairly complex, Jeshion's and Fara's contributions are among the most valuable in the volume, and should be read by anyone with an interest in contemporary discussions of the semantics of proper names.

Marco Santambrogio, in "Empty Names, Propositions, and Attitude Ascriptions", uses the problem of empty names as a frame for introducing a new theory of language-bound propositions (or the propositional contributions of names) that allows empty names to have expressive value even if they have no referent; the guiding idea being that direct reference doesn't require singular propositions. Though Santambrogio does a fair job of allaying certain worries, the view also requires more justification than space allows him to give it here; certain moves seem *ad hoc* and the results (fascinating but) somewhat baroque. In "Millianism, Relationism, and Attitude Ascriptions" Ángel Pinillos develops further his version of *semantic relationism*, based on Fine (2007), to circumvent certain objections raised by Soames (2010) related to certain versions Frege's puzzle (Pinillos also provides a lucid introduction to relationism). Relationism explains the difference in informativity between "Hesperus is Hesperus" and "Hesperus is Phosphorus" by the two occurrences of "Hesperus" in the former being *coordinated*, whereas the occurrences of "Hesperus" and "Phosphorus" in the latter are not (where *coordination* is a semantic property). Soames objects that coordination obviously cannot explain problematic *de dicto single* occurrences, such as "Lois Lane believes that Clark Kent can fly" uttered in isolation. Fine's response is to appeal to inter-discourse coordination (cf. Fine 2010). Pinillos argues that inter-discourse coordination fails; instead, the problematic *de dicto* belief ascriptions always implicitly involve other mental state ascriptions, which can then be used to facilitate appropriate coordination-based solutions.

Sam Cumming's "The Dilemma of Indefinites" is one of the most thought-provoking contributions to the volume. Cumming argues that there is good empirical evidence for a referential analysis of sentences of the form "an *F* is *G*" but also good evidence that such sentences have existential – i.e. not object-dependent – truth-conditions. His radical, but intriguing, response is to deny that the truth-value of an utterance is determined by its semantic content and circumstance of evaluation; if semantic content doesn't determine truth-conditions it can be consistently maintained that "an *F* is *G*" has both singular content and existential truth conditions. He goes on to sketch a novel view of the relation between semantic content and truth: it is possible to secure reference through *private commitment*, which we do when we use indefinites; *truth-conditions*, on the other hand, are a matter of *public commitment* and "an utterance is true if things are the way the speaker's utterance publicly commits to them being." When we use indefinites (rather than definites), then, our utterances, though they have singular contents, eschew such public commitments to reference and object-dependence; it is only for utterances that refer by way of public commitments (those involving definites) that semantic content and truth-conditions coincide.

Perhaps even more provocative is the account Joseph Almong, Paul Nichols and Jessica Pepp sketch in "A Unified Treatment of (Pro-)Nominals in Ordinary English". Rejecting a level of *logical form*, they argue that there is no difference between deictic, anaphoric and bound uses of pronouns like "she" – rather, "she" is always *referential*. And arguing that the "formalist program" in semantics is at odds with the referentialist, externalist tradition that emphasize causal-historical factors in determining linguistic meaning, they instead promote a semantic framework that incorporates such factors. So, for instance, "the contribution of a pronoun to the semantics of a complete utterance is never determined by the application of a semantic rule;" rather, it refers "in virtue of causal-historical connections, and pronoun interpretation is *a posteriori*." The difference between anaphoric and deictic pronouns is neither syntactic nor semantic, but a function of communicative situation and the aspects of context that makes pronoun application appropriate – the aspects audiences use to identify the referent. As the authors admit, the approach cannot be fully developed or justified in the context of the present article alone, and for readers it is perhaps a bit frustrating that they frequently refer to an upcoming "Part II". Theirs is an interesting take, though I will remain skeptical until I see how the gaps are filled in. For instance, if anaphoric pronouns are really referential, then it seems to me that *reference* must be a *brute* word–object relation and not a matter of satisfying certain criteria; but in that case the account does not obviously square with the spirit of the externalist, causal-historical approach emphasized elsewhere – the examples they discuss may fit the "perceptual-chain" model they appeal to, but are difficult to generalize to, say, pronouns in conditionals ("if the US ever gets a queen, she will be tall").

The final article, Edward Keenan's "Individuals Explained Away", is the most technical article in the volume, but it certainly rewards close study. It is also a defense of the formalist approach. In contrast to Almog et al., Keenan argues that natural semantics can do without recognizing individuals at all; indeed, we can do without propositions and possible worlds (to interpret non-intensional contexts). Without going into details of the formal apparatus, Keenan suggests generalizing standard extensional model theory, and replaces the universe of objects our naïve ontology may appeal to with a universe of atomic properties playing the roles that objects play in classical semantics. To do so, he first recasts standard extensional semantic – a booleanly structured set of truth-values, {T, F}, in which sentences are interpreted, and an unstructured universe U of individuals in which individual constants and predicates are interpreted – in purely Boolean terms, where U is eliminated in favor of a booleanly structured set of properties that provide interpretations for common nouns; proper name

interpretations are derivative, defined in terms of properties and truth values (a similar move was suggested in Lewis 1970), and *individuals* are nothing more than "homomorphisms from the property lattice to the truth value lattice." The approach is then generalized to evaluative adjectives, which exhibit properties that make them difficult to account for in standard model-theoretic semantics – successful treatment of these are accordingly justification for using an extended version of extensional model theory. The consequences of eliminating a universe of individuals that singular terms denote or refer to and variables range over are potentially far-reaching, but what Keenan *doesn't* really do is explore the potential impact on foundational discussions about the semantics of proper names. For instance, since the account gives priority to common noun interpretation over proper name interpretation, it might potentially be well-suited to predicativist views.

In conclusion, *On Reference* is a rich and far-reaching collection, and contains a good mix of provocative novel takes on old debates and refinements of familiar positions. And even if not every interesting topic relevant to *reference* is covered, or every article breaks new ground or offer entirely convincing defenses of the positions they seek to defend, it is an invaluable companion to anyone who wishes or needs to stay on top of current trends in discussions about reference, the semantics of proper names or philosophy of language in general.

Fredrik Haraldsen

References

FINE, K. (2007): *Semantic Relationism*. Oxford: Blackwell.

FINE, K. (2010): Comments on Scott Soames's 'Coordination Problems'. *Philosophy and Phenomenological Research* 81, 475-484.

KRIPKE, S. (1980): *Naming and Necessity*. Oxford: Blackwell.

LEWIS, D. (1970): General Semantics. *Synthese* 22, 18-67.

LUDWIG, K. (2007): The Epistemology of Thought Experiments: First-Person Approach vs. Third-Person Approach. *Midwest Studies in Philosophy* 31, 128-159.

MACHERY, E., MALLON, R., NICHOLS, S. & STICH, S. (2004): Semantics, Cross-Cultural Style. *Cognition* 92, B1-B12.

SOAMES, S. (2010): Coordination Problems. *Philosophy and Phenomenological Research* 81, 464-474.

Jakub Mácha: *Wittgenstein on Internal and External Relations: Tracing all the Connections*
Bloomsbury, 2015, 262 pages[1]

In philosophical texts addressing the philosophy of Ludwig Wittgenstein, we are usually confronted with a division of his philosophy to the early, (middle) and the latter phase. However, as an alternative, the authors of *The New Wittgenstein* have suggested that as far as the main (i.e. therapeutic) purpose of Wittgenstein's philosophy is concerned, his work is consistent (cf. Crary & Read 2000). The Czech philosopher Jakub Mácha presents a similar view in his monograph *Wittgenstein on Internal and External Relations: Tracing all Connections*. The author's strategy is to look at Wittgenstein's philosophy from a perspective of a distinction between internal and external relations. Wittgenstein's philosophy has been discussed to a significant extent also among Slovak and Czech analytic philosophers. Anyway, Mácha's monograph comes undoubtedly with some new insights. The book is presented as an 'album' of themes, notes, problems and issues that Macha chose from Wittgenstein's work. Nevertheless, it is not quite of an exegetical nature. In considering these issues, Mácha keeps his own stance toward Wittgenstein's ideas. In order to show the fundamental nature of the distinction between internal and external relations, author puts emphasis on the problems that may be conceived of as secondary to Wittgenstein's main focus: "I admit that I have tried to extract a workable philosophical view or, rather, a coherent set of views from Wittgenstein's *Nachlass*" (p. ix).

Against the so-called new Wittgensteinians who took seriously Wittgenstein's argument that his intention was not in any way to create a philosophical theory, Macha claims that, in Wittgenstein's philosophy, there has always remained something that can be attributed to theory: "I must insist that there still remains something in Wittgenstein's philosophy (the early as well as the later) that can be called a theory. This attempt at setting out a theory is neither about world nor about knowledge nor about language. It is a theory of how to analyse a philosophical text in order to get rid of any philosophical problems that emerge due to the unsurveyable character of natural languages" (p. ix.). According to Mácha, this theory is embedded in Wittgenstein's method of analysis, which binds his early and late philosophy together.

[1] ✉ Lenka Cibuľová
Dovalovo 113, 033 01 Liptovský Hrádok, Slovak Republic
e-mail: lacibulova@gmail.com

More substantially, Mácha suggests that "Wittgenstein's method of analysis rests on the distinction between internal and external relations" (p. x). As a preliminary definition of internal and external relations he appeals to Wittgenstein's quote from *Tractatus Logico-Philosophicus* with a slight terminological modification: "A relation is internal if it is unthinkable that its terms should not possess it, and it is external otherwise" (p. ix).[2] The justification of this claim is questionable, because Wittgenstein himself did not write much about internal and external relations and never ascribed them such a fundamental character as Mácha claims in his book. However, it is true that Wittgenstein often wrote the least about the most important themes in his thinking – e.g. ethics in *Tractatus* or forms of life in *Investigations*. Nevertheless, Mácha argues for the fundamental character of the distinction between internal and external relations in Wittgenstein's thought.

The book consists of twenty chapters grouped into five thematic units. These depict all sorts of topics: an introduction to logical analysis, the distinction between internal and external relations reflected in Wittgeinstein's early and late work, as well as Mácha's own conclusions.

In the first part of the book, *Introduction*, Mácha acquaints reader with the objectives and procedures of logical analysis, explaining how it relates to the distinction between internal and external relations, why they are important and what problems such a differentiation is associated with. Although Wittgenstein's idea about the form of logical analysis had changed during the thirties, the general idea remained the same: "Two forms of expression are identified that look the same in ordinary language. The aim of analysis is to show, however, that they are different" (p. 5). In order to be able to identify ambivalent uses of words and sentences in a language, there is a need for a generic logical distinction between internal and external relations. According to Mácha, Wittgenstein introduced this as a heuristic tool: "The general lesson I would like to draw is how a metaphysical distinction – far from being nonsensical – can be transformed into and employed as an analytical tool" (p. 5). This distinction is somehow present in whole Wittgenstein's philosophy; only its wording has changed. The concept of internal relations was replaced by the concept of grammatical or conceptual relations and the term 'external relation' was replaced by the term 'factual relation'.

The second part, *Prelude*, is devoted to the emergence of the internal/external relations in the context of philosophical thought in the early twentieth century. Mácha pursues the question whether all relations could be classified as internal or external. This issue had been already studied by analytic philosophers such as Moore or Russell on the one hand, and by the British idealists such as Francis

---

2 Wittgenstein originally did not mention relations but properties (TLP 4.123).

Bradley on the other hand. Mácha's intention is not only to interpret the various approaches that have become a background of Wittgenstein's reflections; he wants to show that Wittgenstein's conclusions are ultimately closer to Bradley's than to Russell's concept. Russell and Moore argued against internal relations between objects and elementary propositions. The view that all relations are internal was attributed to Bradley. Yet Mácha points out that the analysts interpreted Bradley incorrectly. From his ontological monism follows un-reality of all relations because Reality is only one. Relations belong only to Appearance and are partly internal and partly external.

The third part of the book is entitled *Wittgenstein's early writings.* Here Mácha is dealing with definitions of internal and external relations in Wittgenstein's early writings, especially in *Tractatus Logico-Philosophicus.* Concerning the distinction between internal and external relations, Mácha focuses on the problems of the doctrine of external relations, the nature of simple objects and Wittgenstein's picture theory. In Wittgenstein's early philosophy, the distinction between internal and external relation is associated with the difference between showing and saying. According to Mácha's elaboration, it can be said that while the internal relations are *shown* in a logically adequate language, external relations can be expressed by propositions and therefore can be *talked* about. However, the author points out that these two differences are not identical – all the internal relations are shown, but not all that it is shown should be regarded as an internal relation. By this Mácha justifies why he did not include an important area of Wittgenstein's thought – his ethics into this work: "it is not straightforwardly clear how to apply the internal/external distinction in ethics or to Wittgenstein's reflections about the sense of the world" (p. 42). By gradual examination of Wittgenstein's early texts, Mácha came to the following characteristics of internal relations:

1. Internal relations are such that it is unthinkable (or impossible) that their relata do not possess them.
2. Internal relations hold between concepts or universals.
3. Internal relations can be exhibited in tautologies.
4. The identification of a term of an internal relation is, *eo ipso*, the identification of all other terms. This characteristic, of course, does not apply to internal properties.
5. The external/internal distinction is an instance of the more general saying/showing distinction.
6. Internal relations can also be labelled as structural or formal relations. (cf. p. 48)

Wittgenstein agreed with Russell that all relations among elementary propositions are external, but he admitted relations between propositions and objects: "despite the doctrine of external relations, Wittgenstein conceives of logical entailment as being based on internal relations and, hence, as necessary" (p. 49). Wittgenstein is here close to Bradley since Wittgenstein had held that the relations do not constitute facts (and therefore the characters for them in a logically perfect language are superfluous), and hence they are unreal. If logic were about objects and the elementary propositions, it would be accidental. Logic is therefore, according to Wittgenstein, about complexes, which implies that the relation of logical entailment is a part/whole relation within a given complex (p. 54). Internal relations can be according to him built into logical notation, where they would *show* themselves: "all the relations that can be expressed in a proposition are indeed external, and internal relations can be shown in a logically adequate notation" (p. 55).

The eighth chapter, *The nature of simple objects*, deals with the question what exactly Wittgenstein's simple objects are and what is the nature of internal relations between them. In the ninth chapter, *The picture theory*, the focus is on the issue whether Wittgenstein intended his picture theory as a picture theory of sense, or whether he introduced it only as an analogy between a picture and language. Mácha provides us with this explanation: "The point of introducing the picture theory of representation and hence the internal relation of depicting is, on my understanding, to improve the analyzed language in the direction of a logically adequate language" (p. 68). The expressions of internal relations should serve as a practical order, as an imperative for correct use of language (expressions).

The fourth part of Mácha's book pursues the definitions of internal and external relations in Wittgenstein's late philosophy, i.e. from thirties onward. At the beginning of the tenth chapter, *Definitions of the internal/external distinction; the later writings*, Mácha again notes that Wittgenstein's understanding of internal relations did not change substantially during his philosophical production. There was a change in emphasis rather than content. The important distinction of the "early Wittgenstein" between *saying* and *showing* is replaced by the distinction between what is expressed by language and what is shown by the grammar of language. Thus, internal relations amount to grammatical relations. The distinction between internal and external relations should help us identify words and sentences that are used incorrectly. The sentence can describe the state of affairs in two ways: (i) it can deal with specific objects, their properties and relations between them, or (ii) it can deal with properties and relations between concepts. The first case is an expression of external relations, while the second case is an expression of internal

relations. Having examined Wittgenstein's late philosophy, Mácha comes with the following definitions of internal relations:

1. Internal relations hold only between concepts while external relations hold between objects and concepts.
2. Internal relations can be exhibited in grammatical propositions, which express either rules of a language game or general facts about our human form of life.
3. Propositions that express internal relations are timeless, whereas propositions that express external relations are temporal.
4. Internal relations relate their terms only in virtue of these very terms, not in virtue of other things or rules.
5. Internal relations allow no exception. (p. 102)

In the fourth part of the book, the author then examines various issues that could be resolved by applying the distinction between external and internal relations. These comprise problems of intentionality, the distinction between reason/motive and cause of an intentional act (with an emphasis on expectation and its fulfilment), the rules and their application, Wittgenstein's philosophy of mathematics, philosophy of colours, the problem with the "standard meter", the problem of seeing aspects, philosophy of psychology and, finally, Wittgenstein's reflections on aesthetics and arts. In each case, Mácha proceeds as follows: first, he introduces Wittgenstein's presentation of the problem, then he clarifies how this problem is specifically related to internal/external relations and, finally, he examines the reflexive use of internal relations in the given context.

In the fifth and last part of the book, *Conclusion*, Mácha sums up the main principles and insights resulting from the previous chapters, now interpreted in terms of the two methodological principles coming from *Tractatus Logico-Philosophicus* (4.122) and *Remarks on Colour* (first paragraph):

1. To insist on the distinction between internal and external relations in the depth grammar.
2. The reflexive cases of internal relations are in fact those cases of direct expression where no relation at all is expressed. (p. 199)

The last two chapters discuss the question of why exactly we are expressing internal relations. Propositions that express internal relations do not represent the state of affairs and thus they do not amount to "moves" in a language game. Then is it not the case that they are superfluous and meaningless? No, it is not. Internal relations tell us something about the logic, or grammar of our language – what they

are and what they should be. Their expression can function as an imperative: "Expressing an internal relation can function as a kind of reminder to someone who is not aware of the logic of our language or it can function as a stimulus to improve our logic or grammar. In short: expressing an internal relation has normative force and can also be taken as an imperative" (p. 201). Such statements have their positive role only until the philosophical confusions caused by incorrect use of language are removed. As they change the language in which they are expressed, they cannot be expressed in a modified language (p. 202).

The last chapter deals with *The maxim of no reflexive uses of internal relations*. This is Mácha's name for a methodological principle, which he finds in Wittgenstein's work. This maxim actually says that "a reflexive use of an internal relation might be a failed case of emphasis. One should consider whether straightening it out into an intransitive use (where no relation is expressed at all) would make the language-game more plausible" (p. 207). In practice, this means that instead of the expression "Now I see a knife as a knife" we say "I see a knife". Finally, Mácha summarizes the individual cases of reflexive use of internal relations. He concludes that the maxim in fact requires that there is some difference between the relata of the internal relation. That implies that there has to be some external relation that could explain this difference.

Mácha's *Wittgenstein on Internal and External Relations: Tracing all the Connections* is a result of thorough examination of Wittgenstein's lifelong work. The author suggests that the distinction between internal and external relations is one of Wittgenstein's most fundamental distinctions.

After many monographs and papers on Wittgenstein's theory of language and his logic, I consider Mácha's book very refreshing (along with Beran's "phenomenological" Wittgenstein – Beran 2013, or Glombíček's detailed study of *Tractatus* – see Glombíček 2016). Mácha's book is apparently not suitable for readers, who are interested mainly in ethical aspects of Wittgenstein's philosophy. Nevertheless, there are still some interesting and inspiring thoughts that can enlighten also that aspect. Problems, which are more challenging, are accompanied not only with Wittgenstein's account, but also with Mácha's examples and explanations. This book is not just another contribution to the debate on the consistency, the nature and purpose of Wittgenstein's work. It extends to a variety of topics and interesting issues, which may be of interest to all readers having a sympathy with analytic philosophy in general and Wittgenstein's philosophy in particular.

Lenka Cibuľová

References

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

GLOMBÍČEK, P. (2016). *Filosofie mladého Ludwiga Wittgensteina*. [*The Philosophy of Young Ludwig Wittgenstein*]. Červený Kostelec: Nakladatelství Pavel Mervart.

Organon F 24 (1) 2017: 135-137

The Emergence of Structuralism and Formalism: A Conference Report[1]

On June 24-26, 2016, Catholic Theological Faculty of the Charles University, Prague, and the Institute of Philosophy of the Czech Academy of Sciences co-hosted "The Emergence of Structuralism and Formalism" conference. The organizers succeeded in attracting four leading scholars of the field – Michael Detlefsen, Leon Horsten, Michael Resnik and Stewart Shapiro – as keynote speakers, with many other well-known figures participating in one of the six conference sessions during the three days of the event.

The topics discussed at the conference were the following (the order of presentations is retained). Opening the first session, L. Horsten considered the prospects of structuralism about set theory in his talk "Structuralism for Set Theory?". N. Tennant in "Structuralism about Truth Itself" explained why verification and falsification in a model are structural notions. V. Kolman's "Intuition and the End of all -isms" discussed implications of the tendency to stress the practical rather than the subjective dimension of intuition. C. Posy in "The Flight from Intuition Revisited" explained why modern mathematics, category theory notwithstanding, is still sensitive to intuition. M. Detlefsen's "The Elements of Formalism" aimed at identification and clarification of principle elements of mathematical formalism. M. Steiner considered Wittgenstein's readiness to employ mathematical systems without previous proof of their consistency in "Wittgenstein against Formalism". M. Gabbay in "Formalism and (set theoretic) truth" considered possibilities of infinitary logic utilization for overcoming the limitations of the problems raised by Gödel's theorem. D. Svoboda questioned the validity of the reasons that led formalists to regard mathematics as a contentless game in "The Emergence of Formalism and a new Conception of Science". C. Mayo-Wilson in "Formalization and Justification" argued that informal proofs often provide greater justification for believing a theorem than do formal derivations.

[1] ✉ Josef Menšík
Department of Economics, Faculty of Economics and Administration
Lipová 507/41a, 602 00 Brno, Czech Republic
e-mail: mensik@mail.muni.cz

Saturday programme was opened by O. Linnebo's talk "Structure Abstraction". He tried to revise his former position regarding pure structures understood as being abstracted from particular systems. J. Wigglesworth in "Non-eliminative Structuralism, Fregean Abstraction, and Non-Rigid Structures" addressed the problem of structures admitting non-trivial automorphisms. L. Kvasz in "Structuralism as a Philosophy of Mathematics – What it is about?" claimed that structuralism explains only some aspects of mathematics which he explicitly identified. J. Menšík's "Mathematical Structuralism: Internal and External" was concerned with a division of structuralists into two broad groups and offered some possibilities for their reconciliation. M. Resnik in "Non-Ontological Structuralism" explained how his approach evolved from sui generis structuralism to a non-ontological version that embraces Quine's doctrine of ontological relativity. P. Sousedík in "Ante-rem Structuralism and Identity" addressed the supposed non-relational properties of mathematical entities, the cross-structural identity in particular. J. Seldin in "Formalism and Structuralism, a Synthesis: the Philosophical Ideas of H. B. Curry" showed that while considering himself as a formalist, Curry should better be recognized as a kind of structuralist. G. Schiemer in "Klein's invariant-theoretic Structuralism" discussed Klein's group theoretical approach in geometry and analyzed its structuralist underpinnings.

Last day of the conference was opened by S. Shapiro, R. Samuels, E. Snyder who in "Neo-logicism, Structuralism and Frege Application Constraints" argued that both neo-logicism and structuralism meet (or fail to meet) Frege's application constraint – a condition to incorporate the applications of a mathematical theory into its very foundations – in a remarkably parallel manner. D. Macbeth in "A Non-structuralist Alternative to Formalism" drew attention to the idea of Freage and Peirce that deductive reason can be both constructive and extend our knowledge. A. Islami in "Formalism in the Face of Complex Numbers" showed that the process acceptance of complex numbers did not fit the formalist conception of mathematics as a purposeless introduction of concepts and their manipulations. F. Doherty in "The Structuralist Roots of Formalism: Hilbert's Early Views" claimed that Hilbert's early views were misunderstood and that he was actually a structuralist before becoming formalist. J. von Plato in "Formal Computation as Deduction" gave an account of how in 1930s steps of formal computation were identified with steps of formal deduction. M. Schirn's "On Hilbert's Formalist Approach before and after Gödel's Incompleteness Theorems" enquired into the evolution of Hilbert's formalism. V. Švejdar in "Modern Czech Logic: Vopěnka and Hájek, History and Background" introduced the Czech logicians Vopěnka and Hájek and discussed their work and their mutual interactions. The last talk also closed the programme of the last of the conference sessions.

All in all, the conference provided a well focused platform of just about the right size for a lively exchange of ideas and contacts, as well as a welcomed opportunity for the present leaders of the field to carry on with various ongoing discussions which started elsewhere. As the event drew to the end, contentment was registered all around with only one question being repeated all over again: when is the next Prague conference on the philosophy of mathematics going to take place?

Josef Menšík

Institute of Philosophy, Slovak Academy of Sciences

and

Slovak Philosophical Association

Modal Metaphysics: Issues on the (Im)Possible V

August 17-18, 2017
(Bratislava, Slovakia)

Keynote speakers

PHILLIP BRICKER (University of Massachussetts, Amherst)
SAEHWA KIM (Ewha Womans University, Seoul)

We invite submissions for a 30 minute presentation followed by 10 minute comments and a 15 minute discussion. Areas of interest might include any aspect of analytic metaphysics, epistemology and logic of modality.

A paper of approximately 3000 words should be prepared for blind review and include a cover page with the full name, title, institution and contact information.

Papers can be submitted in pdf or doc(x) and should be sent to **modalmetaphysics@gmail.com.**

Deadline for submission: **March 15, 2017**
Notification of acceptance: **April 30, 2017**

If you wish to submit a paper, comment on an accepted paper or would need any further details, please, email us to the above address, or visit the conference website **www.metaphysics.sk**

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

and

Institute of Philosophy, Slovak Academy of Sciences

Current Trends in Deontic Logic 2017

November 23-24, 2017
(Bratislava, Slovakia)

Keynote speakers

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

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

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

Organizing Committee

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

Filozofický ústav Slovenskej akadémie vied v Bratislave

Vás pozýva na

XXI. Česko-slovenské sympózium k analytickej filozofii

ktoré sa uskutoční

30. 8. – 2. 9. 2017

v Kongresovom centre SAV Academia v Starej Lesnej

V tejto súvislosti vyhlasujeme výzvu na zasielanie abstraktov k prednáškam.

Svoje abstrakty v rozsahu **300-500** slov zasielajte na mailovú adresu **cssba2017@gmail.com** do **15. 4. 2017**. Časový limit príspevku je 25 minút + 10 minút bude vyhradených na diskusiu. Programový výbor posúdi zaslané abstrakty a vyhradzuje si právo rozhodnúť o prijatí alebo neprijatí príspevku. Toto rozhodnutie sa bude zasielať do **31. 5. 2017**.

Programový výbor

Lukáš Bielik, Silvia Gáliková, Vladimír Havlík,
Jaroslav Peregrin, Marián Zouhar

Organizačný výbor

Lukáš Bielik, Dušan Gálik, Marián Zouhar