ACTA UNIVERSITATIS CAROLINAE
PHILOSOPHICA ET HISTORICA 2/2007

ACTA UNIVERSITATIS CAROLINAE – PHILOSOPHICA ET HISTORICA 2/2007

miscellanea logica (viii)

Foundations of Logic

Editor JAROSLAV PEREGRIN

CHARLES UNIVERSITY IN PRAGUE / KAROLINUM PRESS / 2010

Edited by: Prof. RNDr. Jaroslav Peregrin, CSc.

Reviewed by: Doc. PhDr. Vojtěch Kolman, Ph.D.
Doc. RNDr. Marie Duží, CSc.

ISBN 978-80-246-1815-9
ISSN 0567-8293

Table of Contents

2007 ACTA UNIVERSITATIS CAROLINAE
PHILOSOPHICA ET HISTORICA 2
miscellanea logica (viii)
PAG. 7–15

Logic and Values

VLADIMÍR SVOBODA*
Institute of Philosophy, Academy of Sciences of the Czech Republic, Prague

Though the word "logic" has been used in many different senses throughout the history of Western civilization, the present day logic does not seem to have any serious 'identity problem'. Seen from outside, logic is usually perceived as the discipline that employs various rather strange symbols to formulate (something like) judgments, arguments or proofs and for these purposes develops means used in bizarre (but probably respectable) 'calculations' that can perhaps convince those who understand them about something potentially important. (The respect often stems from the fact that the calculations are associated with mathematics or with computers, i.e. with something that has proved to be useful.) In general – people from outside usually do not question the status of logic as a well-designed discipline. The insiders have a rather different outlook but quite like the outsiders they rarely raise doubts concerning the boundaries of logic. The reason is not that from inside the realm of logic looks quite well delimited, it is rather that few insiders feel the urge to conceive of it as a kind of totality. Most of them are immersed in the particular problems they deal with, do not have any misgivings about their significance, and are content with the general definitions of their discipline that they acquired from common textbooks or from their teachers.

In this paper I want to turn attention to some very general questions that concern the nature of logic – it could be said that I will deal with the problem of the identity of logic. I have chosen the notion of *value* as one through which we can approach (as I believe) the elusive issues. In fact I will make use of the notion in two different ways. In the first part of this paper I will look at logic from outside and put forward the problem of the value of logic as a discipline. In other words – I will aim at answering the question what is (or could/should be) the benefit of exercising logic. In its second part I will focus on the question what role values play (or could/should play) 'inside' of logic. I will suggest that the elusive and general term "value" can perhaps turn out quite useful in the project of delineation of the realm of logic. The conclusion I want to defend could be formulated in the following provocative slogan: logic is a kind of general formal theory of values.

* Work on this paper was supported by grant project of Czech Science Foundation GA ČR no. 401/07/0904.

1. If we ask the question in what sense logic is *valuable* we consider value in a pragmatic sense. Something is recognized as valuable if it can help us in our (practical or theoretical) dealing with the world – something that can be in some sense useful. Of course, logic cannot directly meet our physical needs but it can still provide a kind of benefit. Put roughly – it can enhance our understanding, knowledge and certainty.

To give this bold statement a more concrete content I will formulate, in a naive form, two provisional theses concerning the value of logic:

T1 Logic is (particularly) valuable for dull beings.
T2 Logic is valuable (even) for bright beings.

What is meant by the first thesis? It suggests that exercising logic is useful especially for those who lack proper understanding of language (in particular sentences that play the role of premises of our arguments) and whose mental capacities are limited and their exercising is fallible. As we know in case of deductive arguments conclusion cannot contain (say) more than was already contained (said) in the premises. Thus for a being that would be smart enough to fully understand what sentences say and to fully apprehend their inferential relationships logic would not be of much use. Such supernatural being could not employ means of logic to acquire any new comprehension. It could perhaps use logical means as a tool of convincing some of its less brilliant fellows about the validity of some arguments or proofs, but this usefulness would be still based on the fact that logic is helpful for the limited ones.

If this is right then we could conclude that logic is a discipline that is deeply human. Godlike beings simply could not really profit from its entertaining. But is this right? I would say that only partly. It is quite clear that overcoming the limitations of human (and humanlike) beings is a substantial merit of logic. We often lack complete insight into the linguistically articulated information that we come across and thus feel uncertain what it actually amounts to. Logic may help us surmount the problems stemming from our limitations. No doubt – techniques of reliable extracting information contained in some statements that allow its presentation in explicit and determinate form can turn out very useful. But does the benefit of logic really consist exclusively in this? Let us consider whether logic could not be of some use even for really smart beings – beings whose mental capacities are unlimited.

How could logic be helpful for divinely bright beings? I think that we cannot answer this question unless we know what kind of language the godlike beings use for their communication. I suppose that if the perfect beings communicate by a perfect language then there would be no space for logic in their world. Expressions of perfect language express any idea in perfectly determinate and perspicuous way and for those who have perfect understanding and unlimited mental capacities nothing can stay hidden. The problem is that though we can dream about a perfect language, we can hardly imagine what it could look like. The idea that divine beings would communicate by something like a full-blooded language built on similar principles as those of, for example, first-order predicate logic somehow does not look plausible. Natural languages tend to be untidy, loose and vague (as well as functional, rich and charming) and though we try to develop specific discourses whose linguistic means would be neat

and possessing clear-cut contours, we cannot really reach any proper insight into the realm of perfect language (some say that the reason is that such realm cannot exist).

Thus we can perhaps leave the speculations about the perfect language and assume that the divine beings employ similarly imperfect language as we do. In such case, however, even the brightest mind sometimes inevitably gets embarrassed. Common languages are vague and indeterminate. The borderline between meaningful and meaningless expressions is loose and context dependent and though in some cases inferential relations among meaningful expressions are determinate and obvious, in other cases they are indeterminate in a way that can't be eliminated by application of the mightiest mental capacity. Let us consider some examples of arguments that may illustrate this observation:

Kohlrabies are fellows.
No fellow is a galloot.
Kohlrabies are not galloots.

I want a frankfurter with mustard and bread.
I want mustard.

It is necessary that it is possibly necessary that Hugo is a teacher.
It is possible that it is necessarily necessary that Hugo is a teacher.

Feed the goat!
Feed the sheep or the goat!

The premise of this argument is not true.
It is raining.

It is rather clear that there is hardly any firm matter of fact concerning the correctness of these arguments (or 'arguments'). Even the brightest and most competent speakers of English cannot be assumed to automatically achieve an agreement concerning their validity. Thus the only way is to work on means that would open a path to an agreement based on adopted conventions. And logic has a prominent place among means that allow reaching conventions of this kind.

If the conclusions we have reached so far are sound, it seems that the value of logic is closely bound to two imperfections – the imperfection of our language and the imperfection of our mental capacities. Logic offers ways how we can overcome (at least in some respects and for some purposes) these imperfections. The means developed within logic offer different ways of regulating language – starting from introducing minor, rather natural conventions to creating linguistic means whose character is quite different from the natural ones. The important thing is that the 'regulated' linguistic means are equipped with *meanings* that are (at least in some respect) more definite than in case of natural language expressions and their inferential relations are fixed in such a way that their existence is demonstrable (even from the perspective of a limited language user). Thus the value of logic consists (perhaps beside other things)

in its providing an explication of the pretheoretical notion of inference or entailment, i.e. in replacing the naive, indefinite notion of entailment by a notion fixed by a definition or a corpus of definitions.

Now we may ask an obvious question concerning the character of the definitions provided by logic. Are the definitions *analytic* or rather *synthetic*? The answer is not difficult – the definitions lay in between these two poles. They are synthetic in the sense that they introduce new conventions. But the conventions are not arbitrary as there certainly are respects in which they can be clearly *wrong* (typically too wide or too narrow). Thus we may conclude that those who work in the field of logic are in a somewhat unfortunate position – any definition of entailment they offer is in danger of being *wrong* (unacceptable, too far from intuition) but it cannot be *right* in the sense of being completely adequate.

This view that may appear somewhat pessimistic (at least *prima facie*) is certainly not shared by all logicians. Some of them are firmly convinced that it is only the peculiar system they employ or develop that captures the *real* inferential relations that are out there in the realm of abstract entities studied by logic (concepts, sets, propositions etc). But to stick to this conviction in the context of the varied world of present day logical theories appears to be more and more difficult – it requires that you take large number of your fellow logicians as fools whose work is misconceived and totally useless.

Another possible way of looking at the problem is that all or nearly all systems developed and used within present day logic are objectively correct or adequate. The realm of abstract entities (the Platonist universe conceived in the most liberal way) is so immense and fine grained that virtually any proposed language can be treated as one whose expressions point to some items existing out there and any system of relations fixed by its means (rules, axioms, semantic methods etc.) can be said to capture something real.

Independently of whether we appreciate the plurality of the notions of (logical) entailment that is characteristic of the present day logic we can conclude that having such notions at disposal can be useful for normal limited and fallible beings as well as for beings with unlimited mental capacities. We do not necessarily need to know what objectively follows from what but we certainly need to hold onto principles that allow mutual understanding and we are motivated to strive to formulate the principles in even more useful, efficient, elegant and economical ways.[1]

2. Let us now move to the role that values play 'inside' of logic. If we were to trace the word "value" in logical texts we would probably come to the conclusion that its appearance is nearly in all cases associated with the word "truth". The term "truth value" is one of the most frequent expressions of the logical jargon, at least if we focus on the relevant textbooks. But treating something as valuable is not always so closely associated with

[1] It should perhaps be noticed that though we have concluded that the need of logic stems from limitations of our language and our mental capacity, it would be wrong to think that the usefulness of logic grows directly proportionally to the degree of mental deficiency. We all know that the really dull ones can hardly be convinced even by the most elaborated and sophisticated logical argumentation.

truth. For example recognizing some argument as correct one amounts to treating it as in some sense valuable. Similarly if we say that some rule is valid we treat it as valuable in the sense that it is worth respect. Correctness and validity are (besides truth) qualities that are valued in the world of logic.

But let us first concentrate on truth-values or on truth as a value. According to the conception of logic explicitly articulated and promoted by Frege, the notion of truth occupies the central position in logic. In the famous passage from the article *Der Gedanke (The Thought)* Frege characterized logic as the theory that should discover the laws of truth:

> The word "true" indicates the aim of logic as does "beautiful" that of aesthetics or "good" that of ethics. All sciences have truth as their goal; but logic is also concerned with it in a quite different way from this. It has much the same relation to truth as physics has to weight or heat. To discover truths is the task of all sciences; it falls to logic to discern the laws of truth. (Frege 1956, p. 289)

Let us for now put aside the question which entities most rightly deserve to be ascribed truth values (and hence are the proper objects of logical inquiries) and let us assume that the laws of truth can be studied (more-or-less?) independently of the possible alternative answers. Let us call such entities *truth-bearers*.[2]

As we certainly need to employ as truth bearers (also) some entities that can be presented to the eye or ear of different language users we probably cannot but adopt as truth-bearers expressions of some language – i.e. symbols or sequences of symbols that will be composed in a systematic way. From this point of view language is any system of signs on which we can, in a systematic and non-trivial way, project (or into which we can translate) some part of our natural language. Thus language is any system of signs that either have meanings – in the full-blooded sense of the word – or can be equipped with meanings (can get suitably interpreted).

But since Frege's time there have appeared at least three important changes. The first change concerns the perception of logical principles such as axioms. While Frege and his contemporaries would conceive of axioms as of the most substantial and evident truths of the domain in question, the further development of the various alternative (and competing) systems of logic changed this view almost completely. Axioms are no longer conceived of as exemplifications of the most substantial truths concerning a certain area, but they are rather seen as a kind of more or less deliberate postulates. Thus they are not taken as substantial (in the sense of externalization a kind of substance of a matter) but as a kind of stipulations. They are not directly influenced by any external reality or convictions concerning facts but they are a result of decisions motivated primarily by the goals of the particular logical theory. If we adopt this conception of axioms, then it is doubtful whether they deserve to be regarded as non-trivially true.

The other two changes are even more closely connected with developments in non-classical logics. The first concerns the central concept of truth or truth-value. Many logi-

[2] We should keep in mind, though, that unless the issue of the primary truth-bearers is resolved the subject matter of logic is established only provisionally.

cians have not only admitted that we should count with truth-value gaps, but they have begun to employ strange values like "indeterminate", "nearly true", or 0.392. We can only speculate what would be Frege's view of the change, but it seems likely that he would welcome it adversely. The laws of truth he had in mind would probably not be the laws – or rather rules – invented for treatment of something so elusive as truth values like 0.392 or 0.76. But all the same it is hard to deny that we normally consider some sentences or claims as more (or less) true than others, and studying how the 'fuzzified' truth spreads across our arguments may turn out quite reasonable and useful. The strange intermediate truth values provide sentences with a status that seems cognitively important and worth pursuing.

The third development that Frege probably could not foresee is connected with a more substantial broadening of the scope of logic. Traditionally – for many hundreds of years – logic was understood as a discipline focused exclusively on statements, assertions or propositions. No matter how we understand the terms, it is clear that the explication of the concept of entailment was based on truth ascription: The conclusion of an argument was taken to be entailed by its premises if and only if it was impossible that all the premises were true and the conclusion false.

But during the 20th century some philosophers pointed out to cases of entailment that escape this explication. As an example of such entailment consider syllogisms (or 'syllogisms') like:

Send the appeal to all ministers of the Czech government!
Karel Schwarzenberg is a minister of the Czech government.
Send the appeal to Karel Schwarzenberg!

It seems intuitively rather clear that the prescription in the conclusion of the argument is entailed by its premises. It is, however, also clear that logic as traditionally conceived is not engaged in examination (and substantiation) of arguments of this kind as their conclusion and partly also their premises do not qualify as truth-bearers.

Jørgen Jørgensen was, as far as I know, the first one to clearly point out the problem:

> *According to a generally accepted definition of logical inferences only sentences which are capable of being true or false can function as premises or conclusion in a inference; nevertheless it seems evident that a conclusion in the imperative mood may be drawn from two premises one of which or both of which are in the imperative mood.* (Jørgensen 1937–8, p. 38)

The problem was summarized by A. Ross in the shape of the so-called Jørgensen's dilemma that can be put forward in the following three theses:

1. Imperative (or generally prescriptive) sentences are neither true nor false – they do not have truth values.
2. The relation of logical entailment is based on the concept of truth, hence arguments whose conclusion is neither true nor false cannot be logically correct.
3. Some arguments whose conclusion has the character of an imperative (prescription) are logically correct.

Each of the theses looks plausible but they can't hold together – so at least one of them must be rejected. But which one?

It would not make much sense to get into a detailed analysis of the proposed solutions of the dilemma and the related arguments here. Let us just briefly sum up the basic pros and cons of the three obvious solutions.

The approach that denies that imperative sentences lack truth values is not easy to defend but its proponents have put forward arguments that seem worth considering. We can, for example, claim that any imperative sentence can be paraphrased as a semantically equivalent indicative sentence. For example the sentence:

S1 *Send the appeal to all ministers of the Czech government!*

can be paraphrased as the sentence

S2 *I hereby command you to send the appeal to all ministers of the Czech government.*

This move, however, does not take us too far in the desired direction. We have got rid of the grammatical imperative but we ended with a sentence that is, as concerns truth ascription, similarly problematic – namely a performative sentence.[3] May be we could do better with some other paraphrases. Treating

S3 *I wish you to send the appeal to all ministers of the Czech government*

as synonymous with S1 points to the fact that common motivation for uttering an imperative sentence is a certain wish (or generally a volitive stance). But are really pronouncements of imperatives bound with wishes so directly? Can't we issue orders or advises hoping that they won't be fulfilled?

From the logical viewpoint, it is quite interesting to consider the proposal to take imperatives as equivalent with sentences describing alternative scenarios of future actions. Under this approach, S1 gets paraphrased by sentences like

S4 *Either you will send the appeal to all ministers of the Czech government, or you will be punished.*

But this strategy hinges on a rather implausible assumption that each pronouncement of an imperative sentence is in fact bound with a kind of threat (and if the threat is not fulfilled, the imperative turns out false).[4] It is needless to say that the considerations concerning the possibility of endowing imperatives (or in general prescriptions) with truth values are far more complex. But though the discussions are not closed, we can say that their proponents have not convinced a significant number of logicians and philosophers.

---

[3] Some philosophers, however, argue that performatives can be treated as having truth value – see for example Bach–Harnish (1979), or Searle (1989).

[4] This approach has been used in theories that try to reduce deontic logic to alethic modal logic; see e.g. Anderson (1958).

There are much more logicians who are willing to deny the third thesis i.e. to say that the arguments that look like instances of valid inferences only make a deceptive impression. These logicians would claim that either they are not cases of inferences at all or that they represent a kind of secondary (quasi)inferences whose apparent validity is parasitic on analogous classical inferences containing only indicative sentences (propositions). This approach is certainly legitimate. Its pros and cons can be discussed, but if the disputers won't agree on common criteria of logicality, their arguments are likely to bypass each other.

If we conclude that the most reasonable attitude to Jørgensen's dilemma is to reject its second thesis, we are obviously aiming at a kind of liberal conception of the notion of logical inference. Under this understanding we adopt the view that Fregean conception of logic as the science of truth is too narrow. Logical theories are not just formal studies of the principles governing the spreading of truth values among suitable truth-bearers. They comprise also other formal studies. In the case of logical theories focused on imperative inferences we are confronted with formal studies of the principles that govern (or could/should govern) 'spreading' of a specific normative validity (or 'demandatory force') among suitable meaningful expressions. We can nevertheless imagine that also other values that can be ascribed to expressions of a language might appear in focus of studies that could be called logical. We could think of values like *correctness*, *acceptability*, *answerability*, *justifiability*, *informativeness* etc.[5] Someone might suppose that any semantics employable in logic is in more or less direct way reducible to the fundamental or natural semantics based on the concept of truth. This may or may not be right, but we should not, in my view, overestimate the importance of this issue. In principle it is perhaps possible to take the truth value as the only primitive semantic value and to conceive of all the other values as derived. But on the other hand we can try to start with other concepts (e.g. with the concept of satisfaction) and use them to define the concept of truth.

We could also just decide to stretch the notion of truth so far that it will be attachable to items which normally do not seem suitable truth bearers.[6] This is basically what David Lewis proposes in his characteristically charmingly blunt manner. After equipping imperative sentences with truth value he says:

> I have no real dispute, however, with anyone who finds it intolerable to say that an imperative sentence, when used to command, has a truth value. […] … 'truth value' serves only as a mnemonic label for the values of the function introduced in (3) [the number refers to a paragraph introducing an interpretation function] by existential quantification. If you dislike the label – or any other – feel free to substitute the euphemism of your choice. (Lewis 1979, p. 169)

I suggest that this kind of conventionalist or instrumentalist attitude to these problems is preferable to any kind of foundationalism. Attempts at a rigorous foundationalist

[5] The steps following the specific rules of inference in this case would not be steps *salva veritate* but *salva dignitate* or *qualitate*.

[6] For example, we can come across rather unconvincing attempts at equipping imperatives with 'classical' truth values like Walter (1996).

delineation of the domain of logic and its semantic background are in my view futile (though they may yield interesting partial insights). We should rather allow for talking about degrees of logicality of particular theories or particular systems.

If we adopt this way of thinking and consider estimating the degree of 'logicality' of a logical system we should, in my view, consider primarily answers to the following (admittedly somewhat loose) questions:

1) To which extent does the language of the system in question really deserve being called (and treated like) a genuine *language*? Can the system of expressions be employed as a useful means of conveying information relevant for our practical decision making?
2) To which extent are the constants of the language purely structural – transcending all kinds of discourses?
3) To which extent does the system provide for a suitable and adequate explication of our pre-theoretical intuitions concerning relation(s) of entailment?
4) To which extent is the system built in a systematic way and defined and based on well justified and modest principles?
5) To which extent is the value transmitted across expressions of the language significant for our reasoning and communication?

Answers to the five questions are in fact in several respects interconnected. We should not expect any definite answers – the nature of the questions is rather heuristic. But the point of putting them forward is hopefully clear – in case of each of the questions we can claim that the greater our tendency to answer it in the affirmative, the more the system deserves to be classified as logical.

Though we cannot hope to achieve any firm results, it seems clear that we often come across systems that are clearly worthy of being called logical (e.g. the standard predicate calculus) but we as well encounter systems that will fulfill one or more of the criteria to such a small extent that it would be highly problematic or clearly inadequate to classify them as logical. Once again – we should not search for clear cut borderlines and certainty. This is nevertheless not a big problem. The family of disciplines whose borderlines are contingent and fuzzy has many respectable members.

References

Anderson, A. R. (1958): 'A reduction of deontic logic to alethic modal logic', *Mind* 67, 100–103.

Bach, K. – Harnish, R. M. (1979): *Linguistic communication and speech acts*, MIT Press, Cambridge.

Frege, G. (1956): 'The Thought: A Logical Enquiry', trans. by A. and M. Quinton, *Mind* 65, 289–311 (originally published in 1918).

Jørgensen, J. (1937–8): 'Imperatives and Logic', *Erkenntnis* 7, 288–296.

Lewis, D. (1979): 'A problem about permission', E. Saarinen, R. Hilpinen, I. Niiniluoto and M. Provence Hintikka (eds.): *Essays in Honour of Jaakko Hintikka*, D. Reidel, Dordrecht, 163–175.

Searle, J. R. (1989): 'How performatives work', *Linguistics and Philosophy* 12, 535–558.

Walter, R. (1996): 'Jörgensen's dilemma and how to face it', *Ratio Juris* 9/2, 168–171.

2007 ACTA UNIVERSITATIS CAROLINAE
PHILOSOPHICA ET HISTORICA 2
miscellanea logica (viii)
PAG. 17–22

Logic – Mathematics – Philosophy

PAVEL MATERNA*
Institute of Philosophy, Academy of Sciences of the Czech Republic, Prague

1. Logic: Motivation

The discipline of logic is a scientific theory, which – like any scientific discipline – develops. Therefore, a question arises, which invariant enables us to speak about *logic* when so many changes are observable during this development.

We can read (see Sochor, 2001. p. 7) that "what matters in logic is the way we derive (should rightly derive) consequences." Thus logic does not empirically investigate how people judge; instead it explores the objective criterion of *correctness* of our deriving consequences.

The origin of logic (usually identified with the origin of Aristotle's *Organon*) is indeed connected with this motivation: When do we *correctly* arrive at some claim?

How should we, however, understand the word *correctly*?

Setting aside the 'normative conception of logic', which only postpones the genuine solution of the problem, we come to the following informal characteristic, which captures the specific character of logic: *We arrive at a claim A (logically) correctly only if no empirical information (i.e., information about the state of the world) is needed on our way to A.*

We can illustrate this conception even by as simple examples as the arguments based on logical square. For example, if we deny (negate) a universal positive claim then we arrive, correctly, at a partial negative claim: It will be so always, independently of what the universal positive claim is about.

Thus what is logic interested in from the very beginning is the study of objective conditions of the correct way of *derivation* of a claim. If, of course, there is an objective criterion of correctness of a derivation then it must be some *objective relation that connects the elements of the given way with the goal, the conclusion*. This relation has got the name *entailment* (or *consequence relation*).

Thus we have got two groups of basic concepts of logic: one of them contains concepts concerning the procedures of derivation (*proof*, *theorem* etc.), the other contains concepts connected with entailment like *tautology*, *contradiction*, *satisfiability*, *model* etc. Some key notions connect these two groups: *semantic consistency*, *semantic completeness*.

* Work on this paper was supported by the grant no. 401/07/0904 of the Czech Science Foundation.

2. Logic and Philosophy

We have seen that what is invariant in the development of the discipline called *logic* is the study of entailment, of derivation (proving) and of the relation between them. Then, however, it is obvious that there is a wider discipline, which is concerned with the ways in which we arrive at our claims, and hence the ways of our knowledge: *epistemology* as a part of philosophy. From this viewpoint, logic is a part of philosophy!

Tichý says in a dialogue with Cmorej (see Cmorej *et al.*, 2007, p. 179):

> In my opinion logic is a theory of arguments and as such it is a part of epistemology and, therefore, of philosophy. … [Logic] … is not a formal discipline as for instance algebra is. It investigates a quite specific relation, the relation of entailment, one of the central problems of the theory of knowledge.

This conception runs against principal objections. Among the opponents, we can find especially the neopositivists (in particular the members of the Vienna Circle). The fact that logic is concerned with the 'formal' features of our knowledge has not been accepted by them as an argument for the claim that logic is a part of philosophy. It was rather interpreted as a support of anti-metaphysical declarations demanding a strict separation of logic from (traditional, "metaphysical") philosophy: the latter being taken by the neopositivists as nothing but a 'bad ersatz for music'.

(A historical remark *pro domo nostra*: At the time when being a part of philosophy meant being a part of 'dialectical materialism' the neopositivist argumentation has been accepted – sometimes sincerely, sometimes pragmatically – by many logicians in the satellite countries of the USSR.)

On the other hand, we must admit that the analytic philosophy, which is often taken to be a heir of the Vienna Circle, has lost much of its original anti-metaphysical ethos. In particular, in connection with Carnap's conversion to studying semantic problems (see, e.g., his *Meaning and Necessity*, 1947) we can detect a growing interest in revising the original radical syntacticism.

(The seeming paradox consisting in the fact that the Vienna Circle is usually taken as a *philosophical* school cannot be explained unless we are content with some global characteristics. Going *ad fontes* we can see that not even Carnap's *Logische Syntax der Sprache*, 1934, is as formal, 'syntactic', as it could seem. See, e.g., Procházka, 2006.)

Presently we can observe a rather opposite trend espoused by some influential logicians: an effort to expand the confines within which logic is limited as a part of philosophy. This trend can lead to weakening, or making less distinct, that characteristic of logic which we have stressed in **1.**, *viz.* the absence of any empirical information on the way to the claim. A representative of this claim is in particular Johan van Benthem, who means that logic is concerned with 'information flow in general, not just deductive elucidation' (see Benthem, 2005). This last characteristic is an excellent definition of modern epistemology, but we cannot ignore the fact that a relatively autonomous part of the discipline characterized in this way is just the deductive, non-empirical component, which is traditionally known and cultivated as logic. I do not think that we would gain something if we accepted the proposal to broaden the competence of logic. The collabo-

ration and mutual affecting of the deductive and non-deductive parts of epistemology can be fruitful only if we don't go the way of mixing up both: this would make methods of our work indefinite.

The relation between logic and philosophy is not, however, exhausted by the relation of *part – whole*. Remarkably, the philosophical starting points influence logic itself. Logic remains logic in the sense of our characteristic in Section 1 even when it offers different outcomes of procedures, where different outcomes are given by different philosophies. Here I have in mind first of all the difference between the classical an the intuitionist logic. Let us consider the procedure represented by the scheme $\neg\neg A \supset A$. Classically, but not intuitionistically the procedure is a tautology for any claim A. The distinction is yielded by the difference of philosophical starting points.

But this example (as well as other similar ones) is to be interpreted not in the sense that for the same procedure (involving procedures connected with $\neg$ and $\supset$) the two versions arrive at different outcomes. Actually, the difference of philosophies implies that the procedures connected with the intuitionist 'negation' and 'implication' differ from those ones connected with classical negation and implication. Otherwise we would have to tolerate what cannot be tolerated: inconsistence. Everything would be immediately clear if different symbols for negation and implication were used in classical and in intuitionist logic. The bad habit of using the same symbol for distinct procedures (cf. the 'box' in modal logics) persists in logic up to now.

3. Mathematics

To rightly evaluate the relationship between mathematics and logic we have to be aware of a general characteristic of mathematics. Without understanding this characteristic any considerations concerning the relation between (philosophical) logic, mathematical logic and mathematics would be only speculative.

Let us begin with simplest examples. What does a pupil of the first class of an elementary school learn when taught that $3 + 4 = 7$? (S)he certainly does not learn that $7 = 7$. What (s)he does learn are (computational) *procedures*. Abstract procedures are not set-theoretical objects. (As for detailed arguments for distinguishing between set-theoretical, i.e., simple objects and procedures as complexes see Tichý, 1995, where the author has, among other things, solved Russell's problem concerning what does stick together the components of a structured 'fact'.)

In our example the characteristic function of identity is *applied* to a pair of numbers the first of which is given by the procedure of *applying* the function of adding to the pair of numbers given respectively as 3 and 4.

Another example: Let a, b be any numbers. The equality $a * b = b * a$ is surely interesting. Again a case of equality of outcomes of two procedures.

Another basic procedure is *abstraction*, which makes it possible to construct a function in terms of variables and procedures. This procedure constructs, e.g., a function that associates with any real number r, the set of naturals less than r. For $x \rightarrow \mathbf{R}$ we have $\lambda x \lambda y\, [[\mathbf{N}y] \wedge [< y\, x]]$. If classes and relations are construed as the respective characteristic functions then such procedures as the above one make up the core of mathematics.

What is essential from the viewpoint of the relation between logic and mathematics:

A. *In mathematics the game of procedures and functions is indifferent to possible applications to objects of a certain kind.*

B. *The game ignores the state of the world.*

Ad **A.** In one of his essays, St. Lem compares mathematics to a mad tailor. The latter fabricates all kinds of products without any strategy and in this way builds up a rapidly growing storage. Sometimes a customer comes and calls for a commodity. Mostly he gets it from the storage.

Surely, this image of pure mathematics is not accurate. Lem does not take into account the possible influence of the mental environment of mathematicians. After all, the stimuli, which a mathematician accepts, are of two kinds: First, solving a problem is connected with quite a lot of further problems (we speak about 'autonomous development given by internal needs'), second, some general problems from other disciplines can hold the mathematician's interest: (s)he finds a mathematical aspect of such problems and develops a certain theory without becoming an applied mathematician. At large, however, Lem's metaphor is cogent.

Ad **B.** Be theoretical physics as theoretical as possible, its claims are always ultimately dependent on the state of the world, at least in that it works with dependencies that are indeed in a way necessary (unlike the contingent relations characteristic of an instantaneous state of the world) but not logically necessary. The so-called laws of physics are no tautologies, they could hold with other physical constants etc. Nothing like that can be claimed about mathematics. Verifying mathematical claims we do not confront them with empirical facts, we prove them. (Proofs are certain kinds of procedures.)

An interesting fact is seemingly incompatible with the point **A**. Similarly as logic, mathematics is not 'immune from' philosophy. Now I do not mean the discipline called *philosophy of mathematics* or *foundations*, but mathematics proper. It is well-known that the classical mathematics differs from the intuitionist, constructivistic mathematics due to a distinct philosophy: the former is mostly realistic while its alternatives are based on anti-realistic positions. In itself, this statement is not too interesting: people use to maintain different views, let them work in any discipline. What is interesting is that the philosophical difference has an impact on the discipline itself. A mathematical theorem that has been proved via an indirect proof may not be a theorem of the intuitionist mathematics. In other words, a mathematical procedure – *although indifferent to possible applications to objects of a certain kind* – leads in this case to different conclusions in dependence on philosophical starting points. This happens of course – as in the case of logic – because distinct procedures are associated with the same symbols.

4. Logic and mathematics

Let us return to Tichý's view:

> … [Logic] … is not a formal discipline as for instance algebra is. It investigates a quite specific relation, the relation of entailment, one of the central problems of the theory of knowledge.

This will be our basic guide as concerns the characteristic of the relation between logic and mathematics. In confrontation with the point A we can say: Logic is not *indifferent to possible applications to objects of a certain kind*. A selection of relevant objects is suggested in an only seemingly circular characteristic of logic in Tichý, 1978:

> Logic is the study of logical objects (individuals, truth-values, possible worlds, propositions, classes, properties, relations, and the like) and of ways such objects can be constructed from other such objects.

The criterion of this selection is clear: As a part of epistemology logic is connected with such categories that are needed when entailment and related classes and relations are analyzed. Whereas logic comports with mathematics in the point **B**, it differs from it in the point **A**. In logic, the indifference of mathematics to possible applications is replaced by fixing objects of one kind – those that are needed when entailment is analyzed.

Does this perhaps mean that on such a 'loss of innocence' the procedures investigated by logic are infected by an empirical factor?

We should be aware of the fact that by 'experience' we mean an *extra-linguistic* experience. Every logician understands at least the expressions of his/her native language. The Czechs, Germans, Americans etc. know, e.g., that the meaning of adjectives consists in constructing a function the value of which on some property is another property (so adjectives denote an extra-linguistic modification of a property). This piece of knowledge, which is given *a priori* (otherwise there would be no *understanding* present) makes the logician realize respective logical steps, whose outcome is, e.g., the solution of some seemingly paradox phenomenon ('a small elephant' vs. 'a great mouse' etc.). On the linguistic inputs, which are simply given, logic builds up its procedures without checking whether some dependence on the state of the world is present. So our answer to our question is negative.

5. Philosophical and mathematical logic

In virtue of his/her indifference to possible applications (point **A**) the mad tailor produces *tools* employable possibly in all kinds of disciplines. In the 19th and the beginning of the 20th century logicians (Boole *et al.*, Frege, Russell, …) arrived at the conclusion that the mathematical tools make it possible to let logic move and show that Kant was deeply mistaken when he considered logic to be a completed discipline, where nothing new could be expected. *Mathematical logic* arises as an application of procedures investigated by mathematics to solving problems connected with entailment and proving.

As an alternative name of mathematical logic one began to use (and we use it till now) the name *symbolic logic*. This choice is justified as follows: Mathematics, as a discipline concerning abstract procedures, is bound – because of its extreme abstraction – to expand the means of expression used by the 'normal' natural language by some special artificial symbols. It proceeded in this way from the very beginning, and the conveniences of such a completing of the language by special symbols are obvious. (To illustrate: How would we calculate *two time three times two times three times two times three*? And even

if we disambiguated this entry by means of parentheses, e.g. *(two times three) times (two times three) times (two times three)*, we would surely appreciate the importance of using symbols: $(2 \cdot 3)^3$.)

On the other hand we should be aware of the fact that manipulating symbols can raise a temptation to think that the subject matter of mathematics are just symbols, i.e., not take into account that symbols are symbols *of something* and instead of it be content with handling them as the proper subject matter of mathematical investigation.

The orientation of pure mathematics to handling symbols is natural in the sense that the mathematical symbols can be read as instructions to procedures; since they have been created artificially they do not suffer from the complexity and ambiguity of the way in which the world of procedures is encoded in the natural languages.

In mathematical logic we should be constrained by the view to its pretheoretical subject matter, which is entailment and so truthfulness. This view comes to light – similarly as for Frege (see his correspondence with Hilbert) – as follows: *Prior to* any axiomatization we have to know, which object(s) should satisfy the axioms. If this demand is not fulfilled then the question of truthfulness is dropped or reduced to the question of consistence. Thus the notion of truth(fullness) is either rejected or relinquished to particular sciences or such unscientific subject fields like philosophy, or, finally, extensionally identified with consistence, so that there would be no consistent claim that would not be true.

References

Carnap, R. (1934): *Logische Syntax der Sprache*, Springer, Vienna.

Carnap, R. (1947): *Meaning and Necessity*, University of Chicago Press, Chicago.

Cmorej et al. (2007): *Filosofické dialogy*, Filozofický ústav SAV, Bratislava.

Procházka, K. (2006): 'Consequence and Semantics in Carnap's Syntax', in: V. Kolman (ed.): *Miscellanea Logica* VI: *From Truth to Proof*, FF UK, Praha, 77–113.

Sochor, A. (2001): *Klasická matematická logika*, Karolinum, Praha.

Tichý, P. (1978): 'Questions, Answers, and Logic', *American Quarterly* 15, 275–284; reprinted in Tichý (2004), 295–304.

Tichý, P. (1995): 'Constructions as the Subject Matter of Mathematics', in W. Depauli-Schimanovich, E. Kőhler and Fr. Stadler (eds.): *The Foundational Debate (Complexity and Constructivity in Mathematics and Physics)*, Kluwer, Dodrecht, 175–185; reprinted in Tichý (2004), 873–885.

Tichý, P. (2004): *Collected Papers in Logic and Philosophy* (ed. by V. Svoboda, B. Jespersen & C. Cheyne), Filosofia, Prague & University of Otago, Dunedin.

van Benthem, J. (2005): 'Logic, Rational Agency, and Intelligent Interaction', unpublished lecture, available from http://dare.uva.nl/document/93913.

2007 ACTA UNIVERSITATIS CAROLINAE
PHILOSOPHICA ET HISTORICA 2
miscellanea logica (viii)
PAG. 23–39

Two Concepts of Validity and Completeness

JAROSLAV PEREGRIN*
Institute of Philosophy, Academy of Sciences of the Czech Republic, Prague

Abstract. A formula is (*materially*) *valid* iff all its instances are true sentences; and an axiomatic system is called (*materially*) *sound and complete* iff it proves all and only valid formulas. These are 'natural' concepts of validity and completeness, which were, however, in the course of the history of modern logic, stealthily replaced by their formal descendants: formal validity and completeness. A formula is *formally valid* iff it is true under all interpretations in all universes; and an axiomatic system is called *formally sound and complete* iff it proves all and only formulas valid in this sense. Though the step from material to formal validity and completeness may seem to be merely an unproblematic case of explication, I argue that it is not; and that mistaking the latter concepts for the former ones may lead to serious conceptual confusions.

1. Regimentation and its completeness

To start with, let us summarize some obvious facts about common logical calculi. The passage from a natural language sentence, such as

(1) *Mickey is a mouse and Donald is a duck*

to the corresponding formula of such a calculus, e.g.

(1') $P_1(T_1) \wedge P_2(T_2)$,

can be analyzed as proceeding along two different dimensions, which we can call *regimentation* and *abstraction*. The logical vocabulary of natural language becomes regimented: disambiguated, unified and standardized – a cluster of 'improvements' for which we will use the umbrella term *idealization*. In this way, for example, the natural

* Work on this paper has been supported by the research grant no. 401/06/0387 of the Czech Science Foundation.

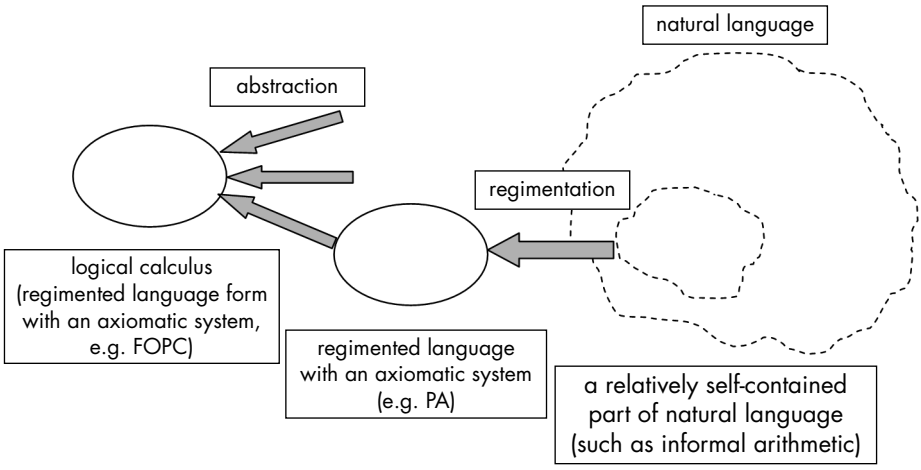
language connective "and" mutates into the operator "∧" with its precisely defined function, which largely, but not wholly, coincides with that of "and" (if only because natural language expressions do not have *precisely delimited* functions). The extralogical vocabulary becomes *abstracted* (*away*): we are not interested in concrete terms, predicates etc., for the principles we are after are those which are invariant across their variety; hence we replace them by formal parameters.

For the sake of clarity, we can decompose this step into two: first, we can imagine, we get from the natural language sentence to its regimented variant, replacing all expressions by (logical or extralogical) constants; and, second, we replace the extralogical constants by parameters, getting the *logical forms* of the original natural language statements. Hence we can imagine that in between (1) and (1') there is

(1'') **Mo**(**Mi**) ∧ **Du**(**Do**),

which arises out of (1) by mere regimentation and out of which (1') arises by mere abstraction.

Whereas regimentation can be seen as in principle one-to-one (which is of course not literally true, for the idealization it effects erases the distinctions between some natural language formulations), abstraction is notoriously many-to-one, it articulates the common structure of many different sentences. As an example of a merely regimented language we can consider the language of first-order Peano arithmetic (PA), whereas the language which can be seen as articulating the logical structure common to PA and any other first-order theory is that of the first-order predicate calculus (FOPC). Hence the situation can be pictured as follows:



What principles govern the process of regimentation and how can we assess whether it succeeds or fails? It seems that two conditions should be fulfilled. First, in order to be nontrivial, the regimented version of a natural language sentence must have greater perspicuity than the original – it must wear the 'logical properties' of the original that we

are interested in somewhat more on its sleeve. Second, in order to count as a *regimentation* of the original, it must preserve the 'logical properties', throwing by the board only the extralogical ones.

What exactly are the 'logical properties' to be preserved and made more conspicuous, and how can they be made so? I think there is little doubt that they are basically the *inferential* properties of sentences: they consist in what each statement can be correctly inferred from and what can be correctly inferred from it. What does it mean to make these properties more conspicuous? I suggest that this means to bring the statement into a form on which we can easily base explicit formulation of inferential rules. Given this, logical form turns out to be, as Quine puts it (1980, p. 21), "what grammatical form becomes when grammar is revised so as to make for efficient general methods of exploring the interdependence of sentences in respect of their truth values."

To avoid a possible misunderstanding which may arise due to the employment of the word "inferential": I am not suggesting that natural language itself is based on *explicit* inferential rules. Some of its sentences *entail* others – and hence the latter are (*correctly*) *inferable* from the former. This notion of inferability, unlike its cousin born within the context of formal calculi, does not make for the cleft between inference and entailment. Hence I will use the terms "inference" and "entailment" largely interchangeably.

However, is there not a viable, and, as some logicians would perhaps suggest, superior, alternative to this inferential construal of 'logical properties'? Should we not say that what is to be preserved and made conspicuous is simply *meaning* – or at least some 'logical part of meaning'? Many logicians would claim that an expression's having any inferential properties cannot but be derivative to its having meaning (see esp. Prior, 1964, for a paradigmatic declaration). So should we see logic as focusing primarily on (certain aspects of) meaning[1]?

To reply to this, it is important firstly to stress that if "inference" is construed in the sense clarified above, inferential properties are definitely *semantic* properties, at least on the relevant sense of "semantics". (We should not let ourselves be misled by the fact that some logic textbooks classify even this kind of inference as a syntactic matter – the fact that it is correct to infer *Mickey is a mouse* from *Mickey is a mouse and Donald is a duck* or *Mickey is a mammal* from *Mickey is a mouse* is clearly a matter of the *semantics* of the words involved – of *and* in the former case and of *mouse* and *mammal* in the latter. This relation is 'syntacticized' only later, within logical calculi.)

Secondly, what is it to preserve meaning? The meaning of a natural language statement is a rather blurry thing, whereas if a sentence of a regimented language has anything like meaning, then it cannot but be our explicit creation, couched within set-theoretical (or similar) terms. Hence it makes little sense to imagine the relation between the former and the latter as an *identity*, it is rather what Carnap and Quine called *explication* – and given this, we face the question what makes this explication a correct one, or a good one, or an adequate one. In other words, preserving meaning cannot serve as a criterion of the

[1] Tichý (1978), for example, claims that "logic is the study of logical objects (individuals, truth-values, possible worlds, propositions, classes, properties, relations, and the like) and of ways such objects can be constructed from other such objects."

adequacy of regimentation, for it is itself in need of an adequacy criterion. And it is hard to see what else could serve as an adequacy criterion save some intersentential inferential links – it is the presence of these links which is straightforwardly testable (by inspecting speakers' overt linguistic behavior)[2].

Take one of the usual ways of explicating the meaning of a sentence, namely as a set of possible worlds, common in the context of modal and intensional logics. How do we tell that the set of possible worlds assigned to a regimentation of a natural language sentence, such as (1"), really is an adequate explication of the meaning of the original sentence, such as (1)? One answer would be that this is guaranteed by the fact that we delimit the set simply as the set of those worlds in which Mickey is a mouse and Donald is a duck, i.e. in which (1) holds. But if there were nothing more to possible world semantics than this, its achievements could not but be trivial. It must, for example, hold that this set is a subset of the set assigned to *Mickey is a mouse*. Why? Because every world in which Mickey is a mouse and Donald is a duck is also a world in which Mickey is a mouse? But how do we know this? Have we visited all the worlds and seen that this is indeed the case? Surely not – it is because *Mickey is a mouse* is entailed by (1).

Hence the adequacy of a regimentation is to be measured by the preservation of the inferential structure (*modulo* the idealization underlying – and constituting the point of – the regimentation). However, as we know that as soon as we have a connective of the kind of *if … then …* of English or of the material implication of standard logic, instances of correct inference become interdefinable with necessary truths (for *B* is correctly inferable from *A* iff the sentence *if A, then B* is necessarily true; and *A* is necessarily true iff it is correctly inferable from the empty set of premises). Hence we can assume that what is to be preserved is *necessary truth*; and as we will, for the sake of simplicity, refrain from considering empirical sentences or theories, we may talk simply about *preservation of truth*. Hence the adequacy of the regimentation turns on what we will call the *material soundness and completeness* – on the extent of rendering of truth within the regimented language in terms of its formal reconstruction, typically provability, within the regimenting one[3].

Material soundness is usually not difficult to check and it appears to be a *conditio sine qua non* of a reasonable regimentation. Material completeness, on the other hand, need not be required – we may want to stay on some level of analysis which does not account for some kinds of truths. (Thus using propositional calculus as the framework

---

[2] This is not to say that inference is a matter of mere regularities – it is a matter of rules which presupposes not only regularity, but also the institution of a notion of *rightness* and *wrongness* (See Peregrin 2001, § 7.5).

[3] The last few sentences have admittedly glossed over a number of issues which are not insubstantial. First, we assumed that what we call correct inference is always context-independent ('necessary' or 'analytic') and that we can separate it from other, context-dependent varieties of inference (kinds such as the inference of *John is in France* from *John is in Paris*). To assume this about natural languages is clearly unwarranted, but as we have already pointed out, logic is based on a purposeful idealization of natural language, which requires the replacing of fuzzy boundaries with sharp ones. Second, the precise features of *if … then …* are also fuzzy and to say that it provides for the reduction of correct inference to necessary truth might require some comments. Third, the term 'necessary' itself is not wholly clear. However, I believe that we can afford not to burden the exposition with the discussion of these themes which, from the viewpoint of the present article, are side issues.

of the regimenting language, we deliberately resign on capturing the kinds of truths which hold in virtue of the words such as *some* or *all*.)

What is essentially important is that unlike in the case of *formal* completeness, which we can encounter in logic textbooks and which we will discuss later, there is no way to decisively check for material completeness, let alone *prove* it. The reason is that the 'naturalness' of the natural language consists precisely in the fact that it is in no respect 'criterially' delimited, and hence we cannot compare the delimitation of the provability within our regimented language with the truth within natural language. What we do is precisely a 'criterial reconstruction' of a natural, and hence 'uncriterial' notion (see Peregrin, 1995).

2. Abstraction and validity

Let us now turn our attention to the step of logical analysis which follows regimentation and which leads from (regimented) sentences to their forms (embodied in *formulas* – sentences with free *parameters*). This step may appear to be almost mechanical: after we set up a boundary between that part of the vocabulary of the regimented, and consequently the regimenting, language that will be abstracted away and that which will survive (hence the 'extralogical' and the 'logical' words), we simply replace the former with parameters.

Now, of course, we also have to switch from truth to *validity*: a formula is, in itself, neither true, nor false – it can only be *valid*, i.e. true for all values of its parameters. Hence we can say that a formula is *materially valid* iff all its instances are true. (Which instances? Primarily the regimented ones, but via them also the corresponding unregimented natural language sentences.) A valid form is supposed to represent a form of 'inherently' true sentences, i.e. of sentences which are valid in force of having this form. In our case, where the form is a *logical* one, it can represent the form of *logically true* sentences. It follows that logical truth is tied to the truth of all instances of the logical form.

This approach to logical truth was foreshadowed by Bernard Bolzano's (1837) approach to analyticity; but Bolzano himself realized its serious shortcoming: it makes analyticity (and in our case logical truth) depend on the contingent fact of the richness of the language in question (a form valid according to this definition might become invalid by the introduction of a new expression enabling the articulation of a counterexample). We can call this the problem of the (possible) "poverty of language". Bolzano avoided it by basing his definition on an 'ideal' language, language *per se*, which as such cannot lack anything.

Bolzano's modem successors, especially Alfred Tarski (from Tarski, 1936, to Tarski and Vaught, 1957) avoided this recourse to an ideal language and coped with the problem in a different way. Remember that a formula is valid iff it has only true instances; in other words if each its 'instantiation', effected by a replacement of its parameters by constants, yields a true sentence. The point of replacing the meaningless parameters of a formula by meaningful constants is in gaining a sentence (composed of meaningful words); and clearly the same thing could be effected by attaching the relevant meanings directly to the parameters, without the roundabout through constants. Hence we could

replace the instantiations of the formula by its *interpretations*: assignments of appropriate kinds of objects to its parameters as their denotations. This has the advantage that we solve the "poverty of language" problem without having to presuppose some such entity as an ideal language *per se*.

However, we have already pointed out that the meaning of a natural language expression is rather elusive and it may be difficult to get a grip on it – so is the Tarskian way not too thorny? The hope of Tarski and other semantically-minded logicians (notably Frege before him, and many others after him) was that what is relevant for logic is perhaps only some such feature or aspect of meanings which is sizeable in some explicit and relatively simple way. Frege, for example, hoped that the only semantic feature of a sentence which should interest a logician is its truth/falsity; and hence explicated its meaning (*Bedeutung*) as its truth value.

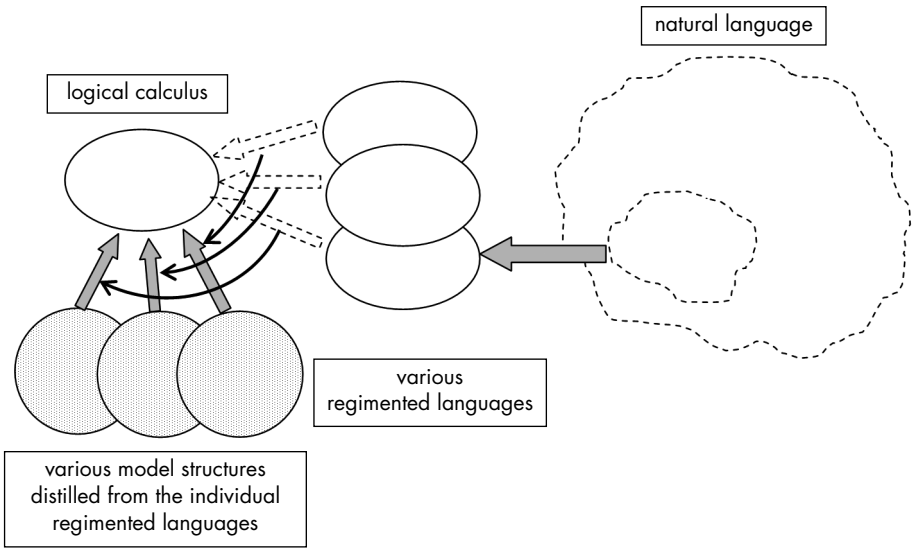
We must keep in mind that the ensuing explication of meaning (we will use the term *denotation* for the corresponding *explicatum*) is deliberately reductive. Nobody (surely not Frege) would want to claim that all true sentences have the same meaning in the intuitive sense of "meaning". Moreover, this kind of explication resembles the digitalization of an analogue signal: the semantic space in which expressions are located and which often resembles a kind of continuum is divided into discrete slots, which can then be treated as separate values to be assigned to expressions.

Given this, we need not be interested in the sentences displaying the form embodied by a formula, but only by the possible denotations of the parameters of the formula. Moreover, it is often assumed that every interpretation we must consider is uniquely determined in some finite way – typically by an assignment of values to elementary expressions and by recursively applicable rules extending the denotations from components to compounds. (This, of course, is the much discussed requirement of *compositionality* of denotations – see Werning et al., 2005.) This makes the space of interpretations particularly perspicuous and it enables us to be sure that we take all of them into account with no omissions.

Hence we can say that a formula is valid iff it is rendered true (or *satisfied*) by every possible interpretation (denotation-assignment). The problem of the "poverty of language" is solved by the successful 'digitalization' of semantics, which makes it possible to consider all denotations, irrespectively of whether they are or are not denoted by expressions of a language.

So provided our digitalization is indeed successful (in the sense that it incorporates all the semantic features of expressions to which our logical constants, now including quantifiers, are sensitive), we can replace any instance of a logical form by an interpretation which verifies the form just in case the instance is true; and moreover, we may have interpretations for which the corresponding instances are lacking because of restrictions of the expressive means of the language in question.

The following diagram illustrates what happened: the links of the previous diagram, connecting the language form, such as the FOPC, with its various instances (such as PA) mutate into the links which now connect the language form with various model structures; as indicated by the thin bold arrows:



Besides the problem of validity brought about by the process of abstraction, logical calculi also inherit the problem of soundness and completeness introduced by the process of regimentation. These concepts, however, change: a logical calculus is materially sound and complete if it proves all and only such formulas which are materially valid. The formal counterparts of these concepts then do not match provability with the material, but rather with the formal counterpart: the calculus is formally sound and complete iff it proves all and only formally valid formulas.

In this way the process of the transformation of material validity and material soundness & completeness into their formal successors appears to be successfully completed. Now, it would seem, we may forget about the former and acquiesce in the latter. But may we indeed? Is it merely another unproblematic case of explication – of a desirable, gradual movement from a fuzzy and loosely delimited concept to a clearer, sharply delimited one?

The trouble is that the existence of a gap between an *explicatum* and its *explicandum* is a fact which, though it can often be disregarded in taking the former directly as a proxy for the latter, can never be entirely *erased*. And taking its existence into account is crucial for the ability of assessing the adequacy of the explication. Otherwise we are in continual danger of letting the *explicatum* play the role of the *explicandum* even in contexts where this is impossible, i.e. where we rely on some properties of the *explicandum* which get abstracted away within the process of explication. Moreover, we may lose the ability to correct errors or infidelities possibly committed during the process of explication – cases where we abstracted away something that was relevant. And in the extreme case, this may even lead to 'arguments' that the *explicandum* 'in fact' (= as 'proved' by the explication) does not have the properties at all.

Elsewhere (Peregrin, 2001, Chapter 9) I portrayed the situation on the background of distinguishing between the 'realm of the natural' (i.e. of those entities which we

encounter and which we can know only *empirically* and hence always *uncertainly* and *incompletely*), and the 'realm of the formal' (i.e. of those which we *define* and hence can know *non-empirically*, *certainly* and *completely*). Explication is a matter of bridging the gap between the two: of offering an item from the realm of the formal as a handy proxy for an item from the realm of the natural. The adequacy of the explication is a matter of the faithfulness of the proxy. And it is an illusion to think that we could explicate the relation of explication *itself* and thus get rid of the adequacy problem – for to be explicated means to be brought into the realm of the formal, which inevitably *creates* a problem of adequacy.

Let us make a historical remark: it was only the 'digitalization', and consequent formalization, of semantics described in this section that retroactively explicitly vindicated the *de facto* existing alliance between the logicians of the Frege-Peano-Russell school and those of the Boole-Peirce-Schröder school (see, e.g. Peckhaus, 2004). The former school was in pursuit of general linguistic forms underlying the most basic general truths and inferential patterns needed for our enterprise of proving, justifying and reasoning; the latter was after certain structural properties of sets of individuals. To exaggerate, we may say that while the former pursued generally valid linguistic forms, the latter engaged in a kind of set theory. What makes the partisans of these two quite different enterprises allies?

After the formalization of semantics took place, we can say that their alliance lies in the fact that investigating the general validity of formulas can be explicated as studying the regularities of the universes from which the constituents of the formulas draw their denotations. Of course, this was intuitively taken for granted by the adherents of both the camps much earlier than it was explicitly articulated and this was why they all felt they belonged under the common umbrella of "logic". However, in this paper we want to point out that the stealthy replacement of the concepts of material validity and completeness by their formal simulacra should not stay unnoticed.

3. The limits of formalization

What exactly must we presuppose to be able to take the proof of formal soundness & completeness as proving also the material one? As we saw, we must take for granted that the 'digitalization' of semantics underlying the Tarskian model theory is adequate in the sense that the 'digitalized' denotation of every complex expression depends only on the denotations of its parts; and that these values can be taken as responsible for the logical (i.e. inferential) properties of statements.

But is the assumption of compositionality of denotations always reasonable? For example, can we assume that truth values, which act as denotations within classical logic, are compositional? Surely not in general: it is obvious that the truth values of a sentence such as *A because B* (*The streets are wet because it rains*) is *not* uniquely determined by those of *A* and *B*. Frege obviously believed that there is some important 'core' of language where such truth-functionality *can* be assumed; and this assumption has been accepted by the mainstream of his followers. Hence what is nowadays taken to be 'standard' logic is truth-functional (in the case of propositional calculus literally; in the case of predicate, with some provisos).

How do we know that the operators of standard logic are truth-functional? Of course, because we have *defined* them to be such. This makes it possible to prove the formal completeness of standard logic: due to the guaranteed truth-functionality of the operators we have a sharply delimited range of manageable interpretations and hence it is uniquely determined when a formula is valid – and we can investigate in how far the set of valid formulas coincide with the set of theorems of some calculus.

Are the new concepts of soundness and completeness a natural explication of the concepts of material soundness and completeness dealt with above? Can the talk about these formal concepts legitimately replace the talk about the original, natural ones? Well, it is clear that this general adequacy depends on the adequacy of the explication of logical operators, which, we saw, were explicated as insensitive to any other feature of meaning save those resulting from the underlying 'digitalization' (in the particular case of standard logic as truth-functional). Can this adequacy be simply taken for granted? This is largely dubious – material implication is by no means a faithful rendering of anything within natural language; and the situation is similar with ordinary negation (if only because natural language contains no single element which could be reasonably held for the counterpart of logical negation – negating in natural language is a very complicated phenomenon which takes various guises).

This is even more true of the surplus operators of predicate logic – there is nothing in natural language which straightforwardly corresponds to either of the quantifiers. Moreover, here we also need a suitable 'digitalization' of the meanings of terms and predicates. The standard method is to take terms to denote individuals (which is not very restrictive, for an individual can be anything) and predicates are taken to denote sets of individuals (which is much more restrictive). Moreover, we also need to stipulate which sets of individuals are available (only some well-behaved – e.g. finitely specifiable ones? Or *all* of them? But what does *all sets* exactly mean in the infinite case?).

Hence the proof of formal completeness builds on the assumption of the truth-functionality of the operators; and if it is to be related to the material completeness of the calculus, we must presuppose that the truth-functional operators constitute a reasonable explication of the corresponding expressions of natural language. But it is not easy to say what exactly the quantifiers explicate. Logical analysts of language, from Russell (1905) to Quine (1980), made it plain that the relationship between quantifiers and natural language structures is very complex and that consequently the logical analysis of language is an extremely nontrivial enterprise. It seems that the ambitions of FOPC w.r.t. natural language are rather holistic: we take it that the quantifiers suffice to regiment – sometimes in a rather intricate way – a great deal of locutions of natural language. Our reasons for believing this are mostly 'inductive' – the practitioners of logical analysis usually find some way to regiment a given locution. This assumption has not been – and cannot be – *proved*, which makes it impossible to consider the proof of the formal completeness as emulating that of material completeness.

But, someone might object, why worry about natural language at all? We are replacing its cumbersome means by the streamlined means of a formal language. Why not simply leave the original means behind and never look back? Well, though there are situations where scientists, especially mathematicians, resort to a language of the form of FOPC, if logic is to be after the generally valid structures of discourse, or at least scientific dis-

course, then there is no way to neglect everything which is not couched in FOPC – it is precisely these theories which logic was called upon to study. Hence the question of the relationship of the formal calculus to natural language remains vital[4].

4. Hilbert & Ackerman on completeness

It may be helpful to consider some historical examples. In their *Grundzüge der theoretischen Logic*, Hilbert and Ackermann (1928) write:

> Die Vollständigkeit eines Axiomensystems lässt sich in zweierlei Weise definieren. Einmal kann man darunter verstehen, dass sich aus dem Axiomensystem alle richtigen Formeln eines gewissen, inhaltlich zu charakterisierenden Gebietes gewinnen lassen. Man kann aber auch den Begriff der Vollständigkeit schärfer fassen, so dass ein Axiomensystem nur dann vollständig heisst, wenn durch die Hinzufügung einer bisher nicht ableitbaren Formel zu dem System der Grundformeln stets ein Widerspruch entsteht.[5]

It would seem that at least the first of these two senses of "completeness" is our material completeness – it amounts to capturing all the truths of some antecedently given range. What Hilbert and Ackerman call the "stricter" delimitation of completeness is then based on the assumption that a theory complete in this sense is a theory which does not allow for a consistent extension. And as consistency of a set of statements amounts to the fact that no contradiction is inferable from the set, T is proclaimed consistent (CON(T)) iff there is no A such that $T \vdash A$ and $T \vdash \neg A$; and it is proclaimed complete iff there is no T^* such that CON(T^*) and $T \subsetneq T^*$. This is not, of course, the Tarskian explication in terms of models, but it is, nevertheless, a delimitation which allows Hilbert and Ackerman to *prove*, in the distinctively mathematical manner, that the propositional calculus is complete. Hence it must amount to a *formal* completeness.

Where did Hilbert and Ackerman move from the material to formal completeness? Was it on the way from their original to their stricter concept? Or did the original, despite appearances, already amount to formal completeness? The crucial question is how we construe their term "inhaltlich zu charakterisierenden Gebietes". What seems to come naturally is to understand "inhaltlich" as "informal" and "Gebiet" as a range of human knowledge and consequently of sentences articulating it, such as arithmetic. However, it is not easy to reconcile this reading with other things which the authors say.

---

[4] Moreover, I do not think that we should embrace the idea of some of the pioneers of logical analysis in that natural language is inherently imperfect and that replacing it by an artificially streamlined language of logic would be a step forward. On the contrary, I suspect that the millennia of development of natural language under the pressure of natural selection have been perfecting it as a tool of communication, and that even the features which have plagued the logical analysts, such as the all-pervasive context dependence, have an important role to play.

[5] "The completeness of a system of axioms may be defined in two ways. In the first case it is possible to construe it in such a way that the system allows for the establishment of all correct formulas of an informally characterized range. However, one can also grasp the concept of completeness in a stricter way, so that a system of axioms is complete only when the addition of a so far non-inferable formula to the system of axioms leads to a contradiction."

From the context of the book it is beyond doubt that the term "Formel" refers to a formula in our sense, i.e. to a statement form, containing parameters, of the kind of our (1'). However, it seems that the talk of a range of human knowledge would require reading the term "Formel" as merely a regimented statement, consisting of constants, like (1'') above – it seems that when talking about a specific range we have in mind some truths specific to it, not *logical* truths, which are common to all ranges.

Moreover, Hilbert and Ackerman continue:

> Die Vollständigkeit in erstem Sinne würde hier besagen, dass man aus den Axiomen … aller immer richtige Aussagenformeln ableiten kann[6].

where the "immer richtige" formulas are just tautologies – formulas which are true independently of the truth values of their component formulas. This would indicate that this kind of completeness is already the *formal* completeness.

I think that what is going on here is a tacit fluctuation between the two different notions of completeness we separated above – between a calculus being complete in the sense of capturing all those forms of the given range that have only true instances, and completeness in the sense of capturing all formally defined tautologies. Introducing the logical connectives, the authors instruct the readers to read $A \to B$ as "if A, then B". This totally obscures the fact that we should ask in how far we are substantiated in replacing the English "if A, then B" by $A \to B$. (And even if this were an excessive pedantry in the case of $\to$, in the case of quantifiers, it is definitely not.)

It is clear that once we replace "if A, then B" by $A \to B$ (and similarly for other logical connectives), completeness is forthcoming; it is in fact trivial. The crucial step from the material to the formal is made by the replacement; and taking this step for so unproblematic as Hilbert & Ackerman do means to conceal the very existence of the gap between the material and the formal notions. This creates the illusion that what the authors proved was *material* completeness; but we can see that this is possible only thanks to the fact that they simply identified the natural language connectives with the artificial truth-functional ones.

5. "Representational semantics" instead of model theory?

Georg Kreisel (1967) starts his texts included in the renown compilation of Hintikka with distinguishing between *formal* and *informal rigour*, the former concerning the set up of an artificial rule-based system, the latter the relationship of such system to pre-formal concepts. Hence it seems that Kreisel urges precisely what was pointed out above: that it is one thing to study relationships internal to a formal system and another things to study the relationship of this system to the preformal reality it was devised to capture.

Indeed Kreisel addresses, among others, the very concept of "intuitive logical validity" and that of the "truth in all set-theoretic structures" – hence he would seem to be addressing the central theme of this paper. However, what he says, from the viewpoint

[6] "Here the completeness in the first sense would say that it is possible to infer all the always correct propositional formulas from the axioms."

entertained here, is rather embarrassing: the only distinction between the two kinds of validity, according to him, is that a formula is intuitively logically valid iff it is true in all structures whatsoever, whereas it is true in all set-theoretic structures iff it is true "in all structures in the cumulative hierarchy". The confrontation of what seemed to be intuitive and formal validity turns out to be the confrontation merely of two versions of the model-theoretical explication of validity.

I think we have two options: either we read Kreisel as talking about a specific kind of language which is from the beginning devised to talk about set-theoretical structures, in which case it is unclear what he means by "intuitive" (for what could be meant by this word would have to be merely an intended interpretation); or we construe him as addressing the truly intuitive logical validity, i.e. truth in force of logical vocabulary alone, in which case his replacement of this concept with the truth in all structures would seem unwarranted.

This problem was noted by Etchemendy (1990, 147–8), who saw the deficiency of Kreisel exposition quite clearly:

> The problem is that Kreisel simply identifies, without argument, the intuitive notion with the model-theoretic notion of truth in all structures. … What [Kreisel's argument] shows is that, in the first-order case, truth in all structures is equivalent to truth in a restricted collection of structures. … What Kreisel's argument does not show, however, is that this extension coincides with the set of logical truths of any given first-order language – say, the logical truths of the language of elementary arithmetic.

Unfortunately, the remedy Etchemendy proposes is also not sufficient. He proposes to replace the Tarskian explication of consequence by means of one based on "representational semantics" – a theory which Etchemendy declares to be "superficially close" to Tarski's model theory, but differing in that the model structures are required to represent the possible states of the world.

The trouble is that what Etchemendy proposes instead of filling the gap within Kreisel's proposal opens up an additional one. Instead of concluding that the step from the intuitive, material concept of validity to the model-theoretic one involves some substantial and arbitrary commitments (which may, in some contexts, be quite in order, *if we are clear about them*), he tries to save the situation by means of adding *one more* commitment, namely that model structures represent possible states of affairs. And it is a commitment even more problematic than the previous ones.

As I argued at length elsewhere (see Peregrin, 1995, § 4.7), it makes little sense to imagine the space of model structures as constituted by modeling, one by one, all the possible states of the world; rather what is constitutive to this space is logical truth. The fact that we have no structure in which an individual has and at the same time does not have a property does not follow from the fact that during our mental journey through the possible states of the world we never encountered anything like this, but rather it follows from the fact that we use our language in such a way that the truth of a statement of the form *X is P and is not P* would not make sense. Possible worlds arise out of the explication of logical truth, not vice versa.

Arguments in favor of the model-theoretic approach to validity similar to those of Etchemendy were given by Priest (1999). He argues, in effect, that the model-theoretic account of consequence and validity is correct to the extent to which the model structures taken into account represent possible situations. However, unlike Etchemendy Priest does see the problem as one of failing to tackle the question of checking the 'representational adequacy', i.e. of inspecting the relationship between the model structures on which the formal application is based and the states-of-the-worlds which are alluded to by the intuitive conception. (And that this is not a trivial problem can be indicated with the help of even the simplest of English sentences, such as "It rains" – for can we really see the part or an aspect of the world which is relevant for the truth/falsity of the sentence as a first- (or, for that matter, higher-) order structure, i.e. as a couple of relations distributed among discrete individuals?)

Priest (*ibid*., p. 189) thus concludes his model-theoretical account for validity with the following disclaimer:

> Strictly speaking, then, we have not given a final account of what it is for an inference to be valid; we have reduced the matter to that of the truth of a certain mathematical sentence. We may well ask the question of what it is for such a statement – or any mathematical statement – to be true. This is a profound question, but is far too hard to address here. One problem at a time!

However, even if we do disregard the general question of mathematical truth (which I agree with Priest is better left for another occasion), we *must* face a more down-to-earth question, namely what would make an answer to the mathematical question helpful for answering the original one? And this question is inevitable, because if the answer is *nothing*, then wrestling with the mathematical question would be a waste of time.

To repeat: I do think that dealing with the mathematized version of the problem *is* helpful; but I also think that silence about the question why and how it is helpful is precarious.

6. Validity of arguments

Finally, let us consider an example of a contemporary logic textbook: I take one which I hold for one of the very best, namely *Logic, language and meaning* by the virtual L.T. F. Gamut[7]. In the introduction of the book, the authors, delineating the subject matter of logic, claim that logic addresses argument schemata, i.e. schemata which arise from taking real arguments and abstracting away some of its component (those which we take to be 'extralogical'). Hence an argument schema (precisely in the way outlined in the beginning of this paper) is a type whose tokens arise out of replacement of the parameters of the schema by certain concrete expressions. The authors give the example

[7] A collective pseudonym of the Dutch logicians J. van Benthem, J. Groenendijk, M. Stokhof, D. de Jong and H. Verkuyl.

(11) *A* or *B*
Not *A*
B

and claim:

> The letters *A* and *B* stand for arbitrary sentences. Filling in actual sentences for them, we obtain an actual argument. Any such substitution into schema (11) results in a valid argument, which is why (11) is said to be a *valid argument schema*.

This leads them to the following conclusion w.r.t. the subject matter of logic:

> Logic, as the science of reasoning, investigates the validity of arguments by investigating the validity of argument schemata. For argument schemata are abstractions which remove all those elements of concrete arguments which have no bearing on their validity. As we have seen, argument schemata can be formed from a variety of expressions and syntactic constructions. Usually they are not all considered together but are taken in groups. So, for example, we can concentrate on those argument schemata which can be formed solely from sentences, grammatical conjunctions, like *or* and *if … then*, and negation. Or we can single out arguments containing quantifying expressions.

I think that little can be objected to this; but surely the reader would expect that after the authors introduce the machinery of modern formal logic, they would return to this and show how the new machinery can be put into the services of characterizing valid arguments. Armed with the apparatus of standard logic, they indeed return to the topic in Chapter 4, where they give a modified explanation of the concept of validity of an argument:

> Translating the assumptions of a given argument into predicate logic as the sentences $\phi_1, \ldots, \phi_n$ and its conclusion as the sentence ψ, we obtain an argument schema $\phi_1, \ldots, \phi_n / \psi$. It has $\phi_1, \ldots, \phi_n$ as its premises and ψ as its conclusion. If accepting $\phi_1, \ldots, \phi_n$, commits one to accepting ψ, then this argument schema is said to be valid, and ψ, is said to be a logical consequence of $\phi_1, \ldots, \phi_n$. An informal argument is also said to be valid if it can be translated into a valid argument schema.

Now this starts to be slightly confusing. As $\phi_1, \ldots, \phi_n / \psi$ is said to be an argument *schema*, the "translation" mentioned must replace extralogical symbols by what are, in effect, *parameters* and hence must amount to what we called *abstraction* above. However, what then does it mean that "accepting $\phi_1, \ldots, \phi_n$, commits one to accepting ψ"? As it seems hard to construe 'accepting a scheme' otherwise than as a shorthand for accepting all instances of the scheme, we presume that what the authors claim is that accepting *any instance* of $\phi_1, \ldots, \phi_n$ commits one to accepting the *corresponding instance* of ψ. But what instances are to be considered here?

From what the authors claimed at the beginning of the book, it would seem that we are to consider the *natural language* instances – those from which the scheme was originally abstracted (in the authors' terminology, those which are "translated" by the schema). Only thus is this definition of validity in accordance with the one given at the beginning of the book. However, then it is not clear what the connection is between this pre-formal notion and the formal notions of semantic and syntactic validity to which the authors restrict their attention from that point on. Are the formal notions explications of the pre-formal one? Obviously, they are meant to be, but no argument whatsoever is given for the claim that they are.

Or should we consider the instances referred to as instances within the language of FOPC? In such a case, the semantic and syntactic validity would clearly capture this notion of validity (the syntactic more, while the semantic one less) directly, but as a consequence it would be a formal concept, not obviously related to the preformal one. Again, a discussion of the relationship between the former and the latter is missing. Although it is often legitimate to work with various kinds of simplifications, idealizations and explications, the replacement of the explicandum with the explicans should not go without saying.

7. The 'digitalization of thought'

The often unreflected step from material validity and completeness to their formal counterparts couched in terms of the standard, Tarskian semantics consists in the 'digitalization' which renders any simple contentful utterance as a statement of a relationship between objects. This, of course, is very natural; but the ventures of those who have tried to develop a picture of the world based exclusively on this kind of 'object-based ontology' (*viz*. Russell, 1914, or Wittgenstein, 1922; or later Barwise and Perry, 1983) indicate that it is not problem free.

In an unpublished dissertation, John MacFarlane (ms.) distinguishes three concepts of formality which usually get conflated during discussions about the formality of logic. First, formality may amount to constitutivity w.r.t. ('giving form to') thought as such. Second, formality may amount to neutrality to kinds of objects. Third, formality can mean abstraction of all kinds of meaning. Let us call these three concepts formality$_1$, formality$_2$ and formality$_3$.

MacFarlane points out that while it was mostly formality1 that was historically seen as the distinctive feature of logic, in the twentieth century many logicians and philosophers of logic have proposed the delimitation of logic based on formality$_2$. (I think this idea originated with Tarski, 1986; recently it has been elaborated by several authors – see, e.g., Sher, 1991.) And as MacFarlane duly points out, it is strange that the proponents of this notion of formality of logic do not discuss its relationship to the more traditional one.

In fact the step from formality$_1$ to formality$_2$ is in some respects parallel to the step from material to formal validity – in particular it presupposes the 'digitalization' of thought. To assume that formality1 is nothing over and above formality$_2$ (or that the latter is a straightforward 'explication' of the former) is to assume that any thought is a thought essentially about objects – or, expressed in a post-linguistic-term idiom, that

every significant statement is a statement about an object or objects. I do not think this is viable. (Consider the example already mentioned: the sentence "It rains" is not a claim about an object or objects – at least if we do not stretch the concept of object to the point of vacuousness.)

In addition, it presupposes something *more*, namely that the semantic content of all expressions which are not directly means of referring to objects (esp. the logical ones) are sensitive to nothing more than to the identity of the objects referred to by the basic ones – in other words that the 'digitalization' underlying the 'object-based ontology' is adequate. And this is something which is far from obvious.

8. Conclusion

The concepts of material validity and completeness are different from those of formal validity and completeness. While the former amount to the confrontation of a formal system with its preformal prototype, the latter is a matter internal to the formal system. And whereas it is often useful and legitimate to replace an informal entity by its formal explication, the attempt to formally explicate the very relationship between an explicatum and an explicandum portends a vicious circle. I am afraid that many writers on logic overlook this danger; and I have tried to indicate what kinds of confusions this may give rise to.

References

Barwise, J. & Perry, J. (1983): *Situations and Attitudes*, MIT Press, Cambridge (Mass.).

Bolzano, B. (1837): *Wissenschaftslehre*, Seidel, Sulzbach.

Etchemendy, J. (1990): *The Concept of Logical Consequence*, Harvard University Press, Cambridge (Mass.).

Hilbert, D. & Ackermann, W. (1928): *Grundzüge der theoretischen Logik*, Springer, Berlin.

Hintikka, J., ed. (1969): *The Philosophy of Mathematics*, Oxford University Press, Oxford.

Kreisel, G. (1967): 'Informal Rigour and Completeness Proofs', I. Lakatos (ed.): *Problems in the Philosophy of Mathematics*, North-Holland, Amsterdam; reprinted in Hintikka (1969), pp. 78–94.

MacFarlane, J. G. (ms): *What does it Mean to Say that Logic is Formal?*, dissertation, University of Pittsburgh.

Peckhaus, V. (2004): 'Calculus ratiocinator versus characteristica universalis? The two traditions in logic, revisited', *History and Philosophy of Logic* 25, 3–14.

Peregrin, J. (1995): *Doing Worlds with Words*, Kluwer, Dordrecht.

Peregrin, J. (2001): *Meaning and Structure*, Aldershot, Ashgate.

Priest, G. (1999): 'Validity', A. C. Varzi (ed.): *The Nature of Logic (European Review of Philosophy, vol. 4)*, CSLI, Stanford, 183–206.

Prior, A. N (1964): 'Conjunction and Contonktion Revisited', *Analysis* 24, 191–195.

Quine, W.V.O. (1980): 'Grammar, Truth and Logic', Kanger, S. & Oehman, S. (eds.): *Philosophy and Grammar*, Reidel, Dordrecht, 17–28.

Russell, B. (1905): 'On denoting', *Mind* 14, 479–493.

Russell, B. (1914): *Our Knowledge of the External World*, Allen and Unwin, London.

Sher, G. (1991): *The Bounds of Logic*, MIT Press, Cambridge (Mass.).

Tarski, A. (1936): 'Über den Begriff der logischen Folgerung', *Actes du Congrès International de Philosophique Scientifique* 7, 1–11; English translation 'On the Concept of Logical Consequence' in Tarski (1956), pp. 409–420 (see also the English translation of Tarski's Polish variant of the text published as 'On the Concept of Following Logically', History and Philosophy of Logic 23, 2002, 155–196).

Tarski, A. (1956): *Logic, Semantics, Metamathematics*, Clarendon Press, Oxford.

Tarski, A. (1986): 'What are logical notions?', *History and Philosophy of Logic* 7, 143–154.

Tarski, A. and R. Vaught (1957): 'Arithmetical Extensions of Relational Systems', *Compositio mathematica* 13, 81–102.

Tichý, P. (1978): 'Questions, Answers, and Logic', *American Philosophical Quarterly* 15, 275–284.

Werning, M., E. Machery & G. Schurz, eds. (2005): *The Compositionality of Concepts and Meanings*, Ontos, Frankfurt.

Wittgenstein, L. (1922): *Tractatus Logico-Philosophicus*, Routledge, London; English translation Routledge, London, 1961.

2007 ACTA UNIVERSITATIS CAROLINAE
PHILOSOPHICA ET HISTORICA 2
miscellanea logica (viii)
PAG. 41–70

Logic and Argumentation

SVATOPLUK NEVRKLA*
Faculty of Arts, Charles University in Prague

1. Logic as the science of arguments and two paradigms of argumentation

It is a commonplace opinion that logic and argumentation are somehow interrelated: we are repeatedly told that the knowledge of logic is essential for understanding and mastering such important human endeavors as rational argumentation.[1] Argumentation, in turn, is used in scientific, legal, religious, ethical, or practical discourse alike.

It is obvious that argumentation skills are essential for any cognitive activity, in which information is gathered, sorted and evaluated. Science of argumentation should tell us how to generate and evaluate arguments for or against some claims in a given context of already presented arguments and logic should be important prerequisite for such a science of argumentation.

We will now proceed with a preliminary definition of argumentation. Under *argumentation* we will understand a certain kind of dialogue[2] in which arguments are exchanged between two intelligent agents. But not all kinds of exchanges of arguments are cases of argumentation, as arguments are also used in other kinds of dialogues. So what distinguishes argumentation from the other kinds of dialogue? The answer is that it is the relation of contrariness between arguments.

But how do we define an argument in the first place? Roughly speaking, an argument is any sentence that is used in order to refute or defend some other sentences. This sentence must not always be declarative, exclamation of threats and suggestive questions can be used as arguments as well. It is also noteworthy that given several different tokens of sentences of the same type (either in form of verbal utterances or in form of a written text), some of those tokens may be arguments and some not. So which particular sentence is an argument can be established only on the basis of its

* Writing of this article was supported from grant GA ČR no. 401/09/H007 'Logical foundations of semantics'.

1 As we are for example told in the introduction to Peregrin's book (2004, p. 4) that "[…] It seems nonetheless, that there is a thing, everybody, who practiced logic from antique to nowadays, could agree upon: logic somehow helps us to determine which reasoning, which kinds of argumentation, which proofs are acceptable and which are not."

2 Walton (2004) mentions another forms of dialogue: persuasion, explanation, negotiation, etc.

actual use within a certain form of dialogue, namely in an argumentation. So in order to identify some sentence as an argument, we must identify its occurrence in some case of argumentation.

To avoid circular definitions we need some primitive notions and for the theory of argumentation, it is sufficient to treat the concept of argument as one of such primitive notions. It is more feasible to describe certain kinds of dialogue first and then define arguments as sentences used in those kinds of dialogue. So for now let an argument be defined as an element or a building block of an argumentation, which will be defined in terms of a relation between arguments.

We will therefore treat the concept of argument as a primitive concept for now and presuppose just that arguments can be presented and that there is the relation of contrariness between some pairs of them. This relation of contrariness or refuting will also be used as a primitive notion.

Provided we are somehow able to determine which argument presents a counterargument to another, we can define argumentation as a process of evaluation of an argument in light of the existing arguments for and against this claim and arguments for and against those arguments etc. This process follows certain procedural rules: Proponent of an argument presents it and then it is up to the opponent to present a counterargument to it and then again the proponent presents counterargument to this counterargument and so on, until one side has no more counterarguments and the other side wins[3].

An example of a formal approach to argumentation is introduced in Dung's (1995) landmark paper. Structure of particular arguments is unspecified in this approach; Dung describes argumentation on the basis of a relation among arguments only. He defines abstract argumentation frameworks as a pair constituted by a set of arguments and a binary relation of defeating on this set. There are no limitations on the cardinality of the set and the properties of the relation[4]. Dung then defines several semantics and procedures for evaluation of arguments that are acceptable at the given framework.

One of the nice features of this approach is that semantics of certain well-behaved frameworks can be construed using dialectical proof procedures, which correspond to the mentioned kinds of dialogue games, as exercised in natural language. Dung's semantics therefore have quite intuitive justification. For a short overview of Dung's semantics and dialectical procedures see Caminada (2008).

But it is still unspecific in what natural language arguments are couched and how do we know that one argument defeats another. The relation of formal abstract frameworks to natural language argumentation is unclear without further specifications of arguments[5].

---

[3] So far we have defined argumentation as dialogical exchange of arguments of a specific kind. However dialogue implies interaction between two agents and argumentation can be performed by single reasoner as well; it can be factual, polemical, or speculational etc. See chapter 4 of Besnard's and Hunter's (2008) monograph for more details.

[4] The relation of defeating can be for example reflexive on some arguments. As we have not defined an argument this should not surprise us. Examples of self-defeating arguments in natural language can be the 'liar sentences' or arguments with unsatisfiable set of premises and claim that is the negation of one of this premises etc.

[5] See chapter on Dung's frameworks in Besnard's and Hunter's (2008) book for more details.

So perhaps it would be easier to start with a definition of an argument that would abstract from peculiarities of natural language arguments mentioned above after all. And that is what logic should do. Logic should be a science of recognizing and evaluating arguments – it should help us recognize well-formed and rational arguments, distinguish them from pseudoarguments and fallacies, reconstruct them and evaluate their properties and mutual relations.

But if recognition of arguments is something that depends so much on pragmatic context, how could such science be possible?

The trick is that only certain well-formed kinds of arguments are considered as candidates for becoming the subject matter of logic. Logic is therefore a science that helps us recognize and evaluate only arguments of a specific sort. Arguments which logic studies can usually be expressed by declarative sentences[6], such that two parts of those sentences can be identified: premises and a claim (which is also sometimes called conclusion, but I use this concept in another sense).

I will call arguments with identifiable premises and claim *well-formed*. A well-formed argument can be defined as a set of sentences with one of them designated as the claim and the remaining serving as premises. Typically premises and the claim of an argument are connected in a manner suggesting that the premises should establish a reason for the claim of the argument, for example connectives or phrases such as 'therefore', 'because', 'so', etc. are used.

Let us note here that there is no need to limit our attention to sentences which have been put together with the help of some sentential connective or comma. Premises and the claim can be expressed by several sentences. As long as there is an indicator of which sentences serve as premises for which claim, we still can speak of a well-formed argument. Also some of the premises of an argument need not be stated explicitly in the argumentation at all, either because they are already implicit in the context, or because the agent considers them already well-established, generally accepted and unquestionable.

Those arguments where either premises or the claim cannot be identified will be called *pseudoarguments*. A pseudoargument is an argument (a sentence uttered with the intention to attack other arguments) in which it is unclear what is its claim and what are its premises.

Whether an argument is a well-formed argument or a pseudoargument depends on a context of its use. Often some premises of the well-formed argument are left implicit in a process of argumentation. It is up to the interpreter of the argumentation to decide whether an agent using such incomplete argument is omitting those premises (because he considers them implicit in the context or being out of question), or if he is just using a pseudoargument. In an extreme case the same sentence can serve both as a single premise and a claim of some argument. Such arguments are often ineffective, but still are well-formed.

Therefore, before using logic to analyze certain argumentation we must do some preliminary analysis, called argument reconstruction. During an argument reconstruction we analyze a text as a token of argumentation. On the basis of our understanding of the purpose, genre, author, intended audience and other features of the text we decide what arguments are used in it and what are their premises and their claims.

[6] For example in erotetic logic a role of questions is also examined.

Argument reconstruction is therefore a matter of linguistic theories and theories of literature rather than a matter of logic and presuppose understanding of the pragmatic features of the text. Logic can help us reconstruct arguments over and above providing criteria for well-formed arguments and thus determining the goal of argument reconstruction.

Logic therefore cannot be bothered with classification and description of pseudoarguments, it is sufficient if it provides criteria for well-formed arguments. From now on whenever I will use term 'argument' I will automatically mean a well-formed one and disregard pseudoarguments.

Logic, besides providing criteria for well-formed arguments, should be able to divide well-formed arguments into two groups, those that are rational and those which are not (for brevity I will call them irrational). This is often called the *argument evaluation* and it is the most exact description of the subject matter of logic. We can define logic as a science of evaluation of well-formed arguments.

In some books on logic, many pages are devoted to the problems connected with fallacies. Fallacies are those arguments which are often used because they are persuasive, but are still irrational. To determine which arguments are persuasive and which not is a business of rhetoric. So with its aid only limited number of classical fallacies can be identified and listed.

While it is a worthy and a useful achievement, I believe that describing and listing fallacies should not be a matter of logic. Again, logic itself is useful for the theory of fallacies in a negative sense only, as it provides clear criteria for demarcating rational arguments just as it provided clear criteria for demarcating of well-formed arguments.

As fallacies are irrational arguments, they lie out of the scope of logic and should be therefore left to other sciences. So logic should be concerned with the distinction of rational arguments from those that are not, no matter whether they are rhetorically persuasive or not. It is also not the goal of logic to answer the ethical question of whether fallacies should be used, but neither should it answer questions what arguments should be used and what strategies are the best in argumentation.

The sole subject matter of logic is the recognition of rational arguments, which presupposes our ability to reconstruct or construct well-formed arguments and is in turn a perquisite of the theory of argumentation, as already mentioned, in which rational arguments play an important role.

So now that we have finally arrived at the definition of the subject of logic, it seems there should be no more doubts what logic is. However, this task of logic, i.e. evaluating of arguments and distinguishing rational arguments from irrational ones, can be carried out in so many different ways that many new questions may arise, leading to doubts and suspicions as to whether such or another account of a rational argument still can be counted as logic.

The central questions 'How are arguments evaluated?' and 'What is a rational argument?' will occupy us for the most of the remaining part of the article.

The different opinions on what logic is or should be and what is its topic result from the different understanding of what constitutes rationality of arguments and therefore different opinions on what properties of arguments we want logic to describe and identify. In this article I will address what I see as two major paradigms of arguments and argumentation.

The first paradigm, which I will call *traditional*, was the dominant paradigm of logic in the past two millennia and is still dominant now. In this paradigm, the structure and content of particular arguments is studied in isolation. The focus is on the question of which arguments are admissible; their relations are not studied. This traditional paradigm had its highlight period during the late medieval age, when it served as the foundation of scholastic philosophy and later, in the first half of the nineteenth century, when it greatly benefited from its symbiosis with mathematics on the one hand and analytical philosophy on the other.

The second paradigm of argumentation, which began to emerge only recently, has its origins in the study of practical problems of application of logic in artificial intelligence, but does have some interesting connections to post-analytical philosophy and its criticism of the standpoints and conclusions of 'classical' analytical philosophers as well[7]. This new paradigm shift brought to attention relation among arguments, rather than their evaluation in isolation.

Let us look at the traditional paradigm in detail first. An example of a textbook on logical evaluation of arguments based on this paradigm is that of Feldman (1999).

The context in which the argument was presented is not considered, as well as the relations among arguments within the whole process of argumentation, the topic of the argumentation, attitudes and background knowledge of agents involved as well. All these aspects are irrelevant for logical considerations within this paradigm. Rational arguments, we are told, are those, in which the conclusion is somehow justified by its premises alone. An argument is said to be rational if its premises present some reason for accepting the claim.

An argument is sound, if it is rational and all its premises are justified. In that case the conclusion of the argument is also justified.

Recognition of sound arguments is therefore something that cannot be done by logical considerations alone, because the problem of deciding whether a proposition is justified lies outside the domain of logic. The logic, as stated before, should be a science of rational arguments, not sound ones.

Unlike the soundness of arguments, which must be evaluated within a specific domain of some particular science, the question whether an argument is rational must be answered without reference to methods and discoveries of other sciences and must be delegated to logic.

Rational arguments further divide into *valid* and *cogent* arguments. An argument is valid if it is impossible for all the premises to be true together with the claim being false[8]. Cogent arguments are those rational arguments, which are not valid. So their premises present some good, although not quite conclusive reasons for accepting the claim.

---

[7] For an overview of the origins of this approach to argumentation and its application in artificial intelligence see Bench-Capon's and Dunne's on-line tutorial (2005); for short introduction to its main results see their introduction to a special issue of Artificial Intelligence Journal (2007).

[8] Someone may feel uneasy about calling argument valid just because it has premises that can never be true. However that is just a matter of terminology. By identifying such argument as rational I do not mean to imply that it is a respectable argument. By definition this argument will be also unsound and therefore quite useless in a rational argumentation. However soundness of arguments, no matter how important for the science of argumentation, is not of concern to logic in this paradigm, even in such an extreme case when unsoundness of an argument is revealed already on level of logical considerations.

An example of a valid argument would be a following argument:

Anne likes Bernard, therefore **Anne likes someone**.

An example of cogent argument would be:

Anne likes Bernard, therefore **Anne treats Bernard well**.

It is usual that we treat well people we like, but it is not a universal rule, as it might be possible in some exceptional case that Anne would like Bernard, but wouldn't treat him well, because, for example, she would be forced to do so for some greater good.

So there might be cogent arguments that are not valid as it is possible in some exceptional cases that the premises of such an argument would be true, while the claim would be false. We will encounter such arguments in section 3. For now let us address some terminological issues first.

A claim of a valid argument is said to be a *consequence* or a *conclusion* of its premises. In such case there exists a so called consequence relation between the premises and the claim. Cogent arguments are inconclusive, there might be some exceptional cases when their premises are true, while the claim isn't, so no universal connection between their premises and claim can be established.

However in cases both of valid and cogent arguments there exists a certain normative relation between their premises and the claim. Whenever premises of a rational argument are justified so is its claim. So it is normative that any rational reasoner who already accepted premises of a rational argument (and it doesn't matter whether those premises are true, or whether he was justified in accepting them in the first place), should also accept the claim of this argument.

Some authors (for example Feldman (1999)) use more traditional distinction and divide arguments into deductive and inductive. In this traditional definition the subject of the claim of a deductive argument is more specific then the subject of its premises. For example:

All men are mortal therefore **Socrates is mortal**.

In this argument, the concept of 'Socrates' is more specific than the concept of 'men'. In case of an inductive argument, the subject of its claim is more general than the subject of the premises. For example:

Socrates is mortal and *Alkibiades is mortal* and *Gorgias is mortal* therefore **all men are mortal**.

Now under standard definitions all deductive arguments would turn out to be valid, while most of inductive arguments would be at most cogent (argument of complete induction would be an example of valid inductive argument). Moreover, some inductive arguments would not be even cogent and there would be also cogent arguments that would not be inductive (I will present some examples in chapter 3).

Moreover, there would be valid arguments that would not be deductive. In what sense is a claim 'I am a human.' less specific than premises 'If it rains then I am a human.' and 'It rains.'?

The distinction of deduction and induction was originally yielded by Aristotle's syllogistic and motivated his terminological concerns. These original definitions have been later modified to include even other cases of valid arguments. Deductive arguments are sometimes defined as those whose claim is already contained in the premises. Also formal logic, which I will describe in detail in the next section, is often called deductive logic, because it is concerned with formalization of valid arguments only. However I believe such terminology is quite inexact and misleading.

It is unclear what it means that the claim is contained in the premises when we are looking at the arguments that do not have both premises and the claim in the subject-predicate form, as Aristotle's syllogisms do[9]. I will therefore use valid-cogent distinction rather then deductive-inductive one.

2. Formal logic as the science of formal arguments and problems of multitude of logical formalisms

Let us now look at valid arguments in detail. An argument is valid, according to the definition above, if it is impossible for its premises to be true while the claim would be false. So in each case when the premises are true, so is necessarily the claim. But what are those cases that make sentences true and what does it mean that a sentence X is true in case Y?

The logic should be able to determine whether the consequence relation between premises and a claim of a particular argument exists. In order to do so it must be able to describe all possible cases and conditions under which the premises and the claim of an argument are true.

But to give an exhaustive and well-arranged list of possible cases of the world and criteria determining truth value of each possible sentence of some language in those cases would be a miraculous accomplishment indeed, which would reduce all human knowledge to determining which state of the world is the case and then using logic. Logical evaluation of valid arguments must therefore use less ambitious methods. Instead of evaluating all arguments of some language one by one, arguments will be evaluated on basis of their form.

This approach to the study of valid arguments has proved extremely fruitful during the development of logic, which later gave rise to a subdiscipline of logic nowadays called formal logic. Formal logic does not evaluate natural language arguments, but studies formal arguments instead[10].

During the formalization of an argument its natural structure is preserved. The formalization consists of the formalization of its premises and its claim. For example, in

[9] These reservations towards the deductive-inductive distinction of arguments have been expressed also in Haack (2000).

[10] The development of formal logic is summarized in Bochenski's (2002) classical monograph and its origins are also mentioned in the introduction to Lorenzen's (1965) book. More recent approach to formal logic can be found in a textbook by Dirk van Daalen (2004).

sentential logic whole atomic sentences are replaced with variables ranging over the set of truth values, and sentential connectives are interpreted as logical constants. Therefore two sentences 'If it rains then the streets are wet.' and 'If you can whistle then the Moon is made of cheese.' share the same form in propositional logic:

$p \rightarrow q$.

Formal sentences can be combined to form a formal argument. So for example an argument:

If it rains then the streets are wet and it rains, therefore **the streets are wet**.

would have the form:

$(p \rightarrow q) \wedge p \therefore q$.

while an argument:

If you can whistle then the Moon is made of cheese and the Moon is not made of cheese, therefore **you cannot whistle**.

would have the form:

$(p \rightarrow q) \wedge \neg q \therefore \neg p$.

It is therefore the formalization of sentences that is carried out during formalization of arguments. A formal argument consists of a set of premises and a single claim (although arguments with multiple claims can also be studied), but those premises and claim are sentences of some formal language. These formal sentences are the result of a process of regimentation of all those components of natural language sentences which are considered to be irrelevant for determining their validity. Those components are replaced with variables.

Formal arguments are therefore certain argument schemes; upon substituting proper terms for the variables in them, we receive instance of these schemes, which are arguments.

For resulting formal arguments it is much easier to specify all possible relevant cases and conditions, under which the sentences they contain hold. Such cases for formal arguments are usually called *models*. A formal argument is said to be formally valid or analytical if and only if in each model where all its premises hold also its claim holds[11]. In such case we say the premises of the formal argument entail its claim. An argument is therefore formally valid if it is an instantiation of some formally valid argument scheme. A claim of an analytical argument with no premises is often called analytical truth.

Such approximation of the relation of consequence to the relation of entailment preserves the most important features of the consequence relation. Entailment rela-

[11] Often it is said that formal sentence is true in a model, or is satisfied in a model.

tion is necessary and normative, which means that if premises of a formal argument are true, then also its claim and that is necessarily true; and if someone accepts those premises it is also rational to accept the claim. Moreover, the relation of entailment is also formal. It holds for whole classes of natural arguments, sharing the same form, which has been regimented into a formal argument. Therefore, formal logic can study classes of arguments, rather than their particular tokens, in scientific manner, as well as a zoologist describes features of whole species of animals, rather than attributes of particular beasts[12].

Now do formally valid arguments really form a subclass of valid arguments? Is it possible that there would exist an argument that would have been formally valid, but not valid? Under our definition it would mean that there would be a logical formalization of such an argument, such that in all models where all the premises of the formal argument would hold, its claim would hold too, while it would be possible that all the premises of the original argument were true, while the claim would be false.

In such case we would probably say that this is not the proper formalization of the argument, as it violates the requirements for logical formalization, such as normativity and necessity. Therefore, formally valid arguments will be required to be valid by definition and from this requirement several restrictions on possible formalization and formal semantics may arise. For example we will require that in all formal models of a given language the logical constants will have the same interpretation and will obey the same laws. Also we will consider only those formal models where non-logical terms obtain theirs 'intuitive' interpretation.

For example the invalid argument:

(+) *Lassie is an animal*, therefore **Lassie is a dog**.

could be rendered formally valid if we inadequately formalized it as an instantiation of the argument form:

$p \wedge q \therefore p$

It could also come out as formally valid if we have defined models of its formalization in such a way that the interpretations of the concepts of 'animal' and 'dog' were the same in all the models.

However, such formalization or formal semantics would have to be considered seriously flawed. While it is now natural to ask how do we tell the correct formalizations and the formal semantics from the flawed ones, this question is beyond the scope of this article. For now, let us simply accept that we want formally valid arguments to be valid by definition.

But if the class of formally valid arguments forms a subclass of the class of valid arguments, does it mean it is a proper subclass?

[12] The ideas concerning formalization of arguments and its features are explained, in greater detail, in Chapter 2 of Beall and Restall (2006).

It seems there might be valid arguments which do not appear to be formally valid. For example such arguments as:

1) *The sun exists*, therefore **there must be some cause of its existence**.
2) *This rock is a body of mass*, therefore **this rock is spatial**.
3) *This rock is a body of mass*, therefore **this rock will fall downwards**.

Now some of those arguments would be considered valid at least for some time in the history of the western philosophy, in the sense of the definition that it is impossible for the premise to be true, while the claim would be false at the same time. But this kind of impossibility, we may be told, is based on other than semantical or logical grounds.

For example argument (1) could be said to be valid (and self-evident) on metaphysical grounds, argument (2) might have been interpreted as a synthetic apriori judgment, which means that it is valid because of the specific attributes of the faculties of our senses (valid on epistemological grounds), while the argument (3) could be interpreted as valid because of the necessary laws of physics.

As we have defined the concept of valid argument using concepts like 'truth', or 'impossibility' several rival accounts of valid arguments arise depending on how we interpret these terms.

It is a common solution in the analytical tradition to disregard metaphysical and epistemological interpretations of concepts like 'impossibility' and to limit its meaning to semantic impossibility only; i.e. to say that valid arguments are valid in virtue of the meaning of the premises and the claim. So if the premises of a valid argument are true, it is impossible that the claim would be false, because we assume that some constituents of the premises and the claim have kept fixed meanings. Therefore, formally valid and valid arguments are identified in this tradition.

It is not the purpose of this article to give an exhaustive explanation of those philosophical traditions, nor to argue for one or another. Readers interested in these problems can find an introduction in Coffa (1991). I am personally inclined to agree with the analytical tradition, while I am also acknowledging the possibility of different readings of 'impossibility'.

It appears to me that the determination of the border between valid and formally valid arguments (or the negation of the existence of these borders in case of analytical tradition) results from the understanding of semantics. If we have quite a narrow understanding of meanings of terms, for example as something that can be either defined explicitly or ostensively, we will probably end up with lots of arguments that will intuitively appear valid to us, but we will not be able to explain their validity on grounds of meanings of the terms. The analytical tradition in philosophy might be therefore understood as an open project motivated partly by the ambition to develop a semantics such that it would be possible to recover all valid arguments as arguments valid on purely semantical grounds.

But let us now look at certain aspects of formalization in more detail.

There are obviously many formal languages, many possible ways how to translate natural language arguments into formal sentences of formal languages and even many possible ways how to define the class of all possible cases and conditions under which some formal sentence is true in some case. However not all such formalizations would

be called logical formalizations and not all formal arguments are logical arguments. Let us take as an example the following argument:

(*) *Lassie is a dog*, therefore **Lassie is an animal**.

We can decide to abstract from concrete terms denoting individuals, replace the term 'Lassie' with a variable and obtain the following argument scheme:

(1) *X is a dog*, therefore ***X* is an animal**.

Under each substitution of an individual name for *X* into (1) the resulting argument will be valid due to the fixed meaning of the terms 'dog' and 'animal'. It is not possible that the premise would be true while the claim would be false under standard interpretation of those terms. No matter what individual we substitute for *X*, if *X* is a dog, then it is also necessarily an animal. So we can say that an argument scheme is valid if and only if all of its admissible instantiations (all variables of some kind are replaced with a proper term of the appropriate kind) are valid. Therefore (1) is a valid argument scheme.

However, we would still probably hesitate to call (1) valid on grounds of logic, because it still presupposes some specific knowledge, which we feel has nothing to do with logic, such as the knowledge of meanings of the terms 'dog' and 'animal'. If we were to abstract from the content of the original argument consequentially, we would have to replace predicates too. The resulting argument scheme would be:

(2) *X is A*, therefore ***X* is *B***.

This scheme is no longer valid, because we can substitute 'politician' for *A* and 'corrupt' for *B*. So which of the schemes (1) and (2) is the correct result of formalization of the argument (*)?

Beall and Restall (2006) present a list of similar argument schemes on p. 20 of their book, with decreasing degree of credibility of being valid on grounds of logic. They ask what it means that an argument is valid on grounds of formal logic and what is it that makes such logic formal. They argue that formal logic should not be simply any specification of conditions of validity for arbitrary set of argument schemes. Rather they suggest another three possible interpretations of formality.

Under the first interpretation, logical validity of formal arguments is always field-independent. To recognize logical validity of a formal argument, we need not have any specific knowledge, such as we have used in (1). This is a maxim stemming from Aristotle's ideal of logic as a universal tool of the sciences. In the traditional paradigm, logic is intended not only to be a special science among sciences, it should serve as a fundamental prerequisite to all other sciences.

However, in (1) we do not employ any specific scientific knowledge. We do not need to know zoological taxonomy to understand that all dogs are by definition animals. Therefore Beall and Restall (2006) suggest a stronger requirement. The goal of logic, which is the recognition of rational arguments, must be independent of any other knowledge, provided to us not only by special sciences, but also by the knowledge of the rules governing use of concepts in discourses.

The second interpretation of formality is the following: An argumentation scheme is a formal argument if all referring terms in it have been already replaced by variables. Under this definition the scheme (1) would not be a formal argument, because terms 'animal' and 'dog' still have not been replaced, but the scheme (2) would be a formal argument, although in this case it would not be logically valid. A similar approach can be found in Tarski's article (1994).

Under the third interpretation, formal arguments already abstract from all semantic content of premises and the claim of the original argument.

While these conditions are very similar, Beall and Restall (2006) argue that it would be a mistake to confuse or identify them. Still we can make a simple observation: Logically valid arguments are those arguments which are not valid on the basis on semantics. The determination of logical schemes is dependent on what we are willing to see as the 'rules governing the use of concepts', 'referring terms', or 'semantic content'. Logic therefore can be said to begin where semantics ends, or vice versa.

Terms which are not referring or have no semantic content will be called logical constants. Ideas behind introduction of the concept of logical constants are in greater detail explained in Chapter 2 of Susan Haack's book (2000).

We will therefore introduce another class of valid arguments, which is the class of logically valid arguments. Logically valid arguments will be arguments valid on the basis of their logical form, where only logical constants are treated as constants, as opposed to some less general form. Arguments that are formally valid, but not logically valid will be called material arguments.

But what are the logical constants? Where exactly lies the border between materially and logically valid arguments? No universal agreement exists here. Traditionally sentence connectives like 'and', 'or' etc. are treated as logical constants within sentential logic. Quantifiers like 'all' and 'some' are logical constants of predicate logics, while phrases such as 'it is necessary/obligatory/believed… that X' are often formalized by logical modalities. There are several options for the choice of the set of logical constants. Those sets are called 'logical languages'.

Now there is a multitude of 'logical languages', but sometimes even for the same language there is a multitude of specifications of models and of the notion of truth in a model. Consequently there arise multiple accounts of logically valid arguments. I will call such accounts logical formalisms. Examples of logical formalisms are 'classical propositional logic', 'intuitionistic propositional logic', 'classical predicate logic' etc.

Some of these formalisms differ in language, some differ just in their account of logically valid arguments. Such accounts need not always be specified in a model-theoretic way. For some logics, for example the substructural ones, it is more natural to use other methods to describe the set of logically valid arguments. So for example valid arguments of classical propositional logic can be defined using natural deduction calculi, Hilbert calculi, Gentzen calculi, truth tables, Beth's tableaux's, topological semantics, Boolean algebras etc.

We could be tempted to forget about the existence of the variance of methods as long as they lead to the same conclusion and simply identify logical formalisms with a pair constituted by the set of formal arguments of some language and the subset of this set designating logically valid arguments. However such a definition would be too broad and would include even curiosities unworthy of the name of logical formalism.

First of all, a language of a logical formalism should be some form of regimentation of natural language, where only logical constants remain. The list of logical constants is usually finite, but always recursive. Rules how logical constants can be used to form compound sentences from less complex ones are also described, therefore a whole logical language can be generated by an algorithm and is therefore always recursive. It should be always decidable in a finite number of steps whether a sequence of symbols is a sentence of a formal language of a logical formalism.

Therefore formal arguments of a given logical formalism can be identified in a finite number of steps.

The subclass of logically valid arguments can usually be generated by some algorithm as well and is not very high in the arithmetical hierarchy. So certainly not all subsets of a set of all formal arguments of a formal language deserve the name of logical formalism, as the class of all formal arguments is countably infinite and within this set of formal arguments we could find an arbitrarily complex subset of logically valid arguments.

Therefore for many of these subsets no reasonable algorithm exists, that could identify members of this subset. Such sets cannot be studied by methods of formal logic. Hence logical formalisms are usually not defined by listing of all logically valid arguments, but by the introduction of some algorithm which generates them.

Under this understanding of logical formalisms the big advantage of formal logic is that the formal languages can be made tidier and easier to manage with precise mathematical methods than the infinite set of all potential arguments in natural language. Therefore instead of describing natural language arguments themselves, one can transform premises and claim of these arguments into logical formulae, which are objects of a formal logical language and then decide the validity of corresponding formal argument using some algorithmic methods. Without the existence of these methods there is no need to formalize arguments in the first place.

For these reasons I believe that only such sets of formal arguments should be called 'logical formalisms' that are at most recursively enumerable (can be recognized by some recursive algorithm, but possibly in infinite time). In Chapter 4 we will see examples of sets of formal arguments that are reasonably motivated and are therefore traditionally called logics, but that do not satisfy this requirement, so I would hesitate to call them logical formalisms.

Existence of such methods presupposes certain attributes of the entailment relation between premises and the claim of a logically valid argument. Not only that each substitution of referring terms for variables in logically valid argument should result in arguments that are again valid, but also substitution of formal sentences of the language for variables should result in another formal argument, also logically valid. Such feature of the entailment relation is called the structurality of the relation.

For the reasons mentioned above, I will be hesitant to call formalisms (often referred to as nonmonotonic logics) of Chapter 4 logical formalisms, even though they are also dealing with formal arguments and therefore have something to do with logic. Despite these two properties, putting them in the same box with logical formalisms that do have the property of structurality would cause confusion and rise a lots of questions as to what actually formal logic is.

Still, even under those restrictions there remain a multitude of logical languages and even logical formalisms interpreting the same logical symbols as different logical constants. Therefore there arises a question whether there is a correct logical formalism (and if so, which one it is), or whether the multitude of logical formalisms presents just a multitude of tools to be used on different occasions for different purposes[13].

This question could be rephrased in the following way: 'Is there a logical formalism, which would help us to recognize all logically valid arguments? If so, which one is it?' But these questions are probably ill-phrased for how can we tell what arguments are logically valid independently of any logical formalism? So when I will use the term 'logically valid argument' throughout this article I will not use this as an absolute term, but relatively to some logical formalism, for example logically valid argument of classical propositional logic.

There is also a second reading of the question, which is as follows: 'Is there a logical formalism which would help us recognize all formally valid arguments?' Then the answer is 'NO'. The set of arithmetical truths, which certainly is a subset of analytical truths, is not recursively enumerable. If we require that logically valid arguments were recursively enumerable, we cannot include arithmetically valid arguments among them.

There is also another stance one can take when confronted with the problems of formalization of natural languages and multitude of logical formalisms, which is to disregard problems of formalization and relations of logical formalism to natural language argumentation and use logic to evaluate arguments couched already in a formal language.

This branch of formal logic is called 'mathematical logic' and is often contrasted with the so-called 'philosophical logic', which is primarily concerned with the application of logical formalisms to natural language.

Interest of philosophical logicians lies in the study of some cases of reasoning in natural language, often paradoxical, which cannot be faithfully formalized within current logical formalisms and in the creation of new formalisms suitable for the formalization of those problematic cases. Philosophical logic is therefore concerned with a design of logical formalisms and their application to understanding natural language arguments.

Therefore, the term 'philosophical logic' is sometimes used in a narrower sense as the description of properties of nonclassical logical formalisms[14], while mathematical logic is limited to the study of those logical formalisms which are commonly used in mathematics and their relation to other mathematical disciplines.

This is not a very fortunate distinction. While it is a historical fact that many nonclassical logical formalisms originally arose due to the effort of philosophical logicians to explain some natural language arguments including philosophically interesting concepts like 'necessity', 'time', 'knowledge', 'belief', 'obligation' etc. (explicated in modal logics), origins of some other nonclassical logical formalisms were motivated by mathematical interests (for example intuitionistic logic).

Moreover, those logical formalisms which are now considered as the object of research of mathematical logic were originally described in philosophical terminology motivated by concepts such as 'meaning', 'truth', 'consequence', or 'necessity'. It was only later that

[13] Such questions are addressed for example by Haack (2000) or by Beall and Restall (2006).
[14] That is in fact what one often finds when he looks into the *Handbook of Philosophical Logic*.

Theory of argumentation: interaction of arguments, strategies of argumentation, rhetoric… +

ARGUMENT RECONSTRUCTION ARGUMENTS *Linguistic theories, theories of literature…*

PSEUDOARGUMENTS WELL-FORMED ARGUMENTS

ARGUMENT EVALUATION

TRADITIONAL PARADIGM:

Dialectic = formal logic + informal logic = logic

RATIONAL ARGUMENTS IRRATIONAL ARGUMENTS

Fallacies are a subclass of irrational arguments studied within an eristic dialectic.

COGENT *Logic*

- *Lassie is a dog*, therefore **Lassie barks**.
- *Arnold is bigger than Bob and Bob is bigger than Charlie*, therefore **Arnold would beat up Charlie**.
- *x = 3 + 2*, therefore **if Anne gives Charlie three apples and bob gives Charlie two apples, Charlie will have enough apples for his kids**.

(FORMALLY) VALID

Semantics, Mathematics

MATERIAL

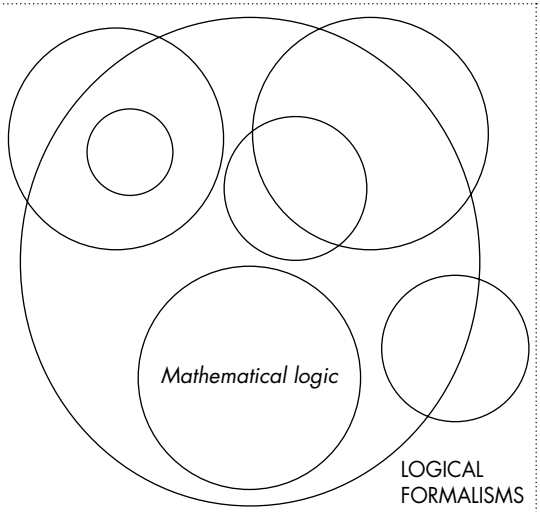
- *Lassie is a dog*, therefore **Lassie is an animal**.
- *Arnold is bigger than Bob and Bob is bigger than Charlie*, therefore **Arnold is bigger than Charlie**.
- *x = 3 + 2*, therefore **x = 5**.

Formal Logic

LOGICALLY VALID

- *Lassie is a dog and all dogs are animals*, therefore **Lassie is an animal**.
- *Arnold is bigger than Bob and Bob is bigger than Charlie*, therefore **Arnold is bigger than Bob if Bob is bigger than Charlie**.
- *x = 3 + 2*, therefore **2x = 2(3 + 2)**.

Philosophical logic



these concepts were replaced with mathematical concepts like 'reference', 'model', or 'entailment'. Also nonclassical logics have gone through the same development and most of these formalisms, which were motivated by philosophical interests, like modal logics, are now extensively and fruitfully studied by means of mathematical methods.

Therefore I am inclined to distinguish mathematical logic from philosophical logic only by identifying the first with studying the internal formal properties of logical formalisms, and the second with their interpretation and application to natural language arguments, which are often of some philosophical interest. This distinction can be best expressed in the terminology used by S. Haack (2000), who distinguishes between interpreted and uninterpreted logical formalisms.

It is clear that a formal language of mathematics can be understood as an extension of a logical language and a system of mathematical truths as an extension of a logical formalism, but that does not mean that mathematical logic should be understood narrowly as logic for studying the language of mathematics (however usual this may be), but rather as studying of logical formalisms, no matter which ones, by mathematical methods. Those logical formalisms, studied by mathematical logic, may also come from regimentation of natural language and must not be motivated by mathematical concepts. The adjective 'mathematical' therefore refers to the methods of the discipline, rather than to its subject.

So in this understanding mathematical logic would be a subdiscipline of mathematics, studying uninterpreted logical formalisms as mathematical objects, rather than the foundational discipline for all mathematical reasoning. If mathematics is studying certain material formal arguments (often including concepts like number, space, vector etc.), it should also include studying purely logical arguments.

3. Informal logic as a science of cogent arguments and limits of formal logic

In previous sections we have defined formal logic as the science of logically valid arguments and logic in a broader sense as the science of rational arguments. But what is a relation of formal logic to logic as the science of rational arguments? Is there a need for another discipline, which could be called informal logic, or is it the case that methods of formal logic are sufficient for determining rationality of all arguments?

Let us look at the case of analytical arguments. It has been demonstrated in the previous section, that reasonable logical formalisms can help us recognize only a limited subset of analytical arguments, namely logically valid arguments. Therefore logic could not be the science of all analytical arguments. However it still can serve as a universal tool of science, as with the aid of a set of additional premises we can turn any analytical argument into a logically valid one.

For example, let us consider following analytical argument from previous section:

(*) *Lassie is a dog*, therefore **Lassie is an animal**.

This argument can be easily turned into a logically valid argument (of classical predicate logic):

(x) *Lassie is a dog* and *whatever is a dog is also an animal*, therefore **Lassie is an animal**.

To see that this argument is valid we need to understand formulations like 'and', 'whatever is X is also Y', which are constructions that can be formalized within language of predicate logic.

Nothing further must be known, not even the meaning of the terms 'dog' and 'animal', because all relevant features of those meanings have been already made explicit in the argument. Replace 'dog' with 'vegetable' and 'animal' with 'plant' and you obtain equally logically valid argument.

The trick here was that we added a new premise, which made the relation between the meaning of the extralogical referring terms involved in the original premises and the meaning of the terms in the original claim explicit. Such premise is often a called 'linking premise' and it should be a trivial observation that in all logical formalisms (with certain 'reasonable' properties of the inference relation) any argument that is not logically valid can be turned into a logically valid form simply by adding a single linking premise to its premises.

However such trick can be used only for arguments that are already analytical. There still may be rational arguments that are not even analytical. Even if we are willing to accept the problematic philosophical assumption that all valid arguments are valid formally (and are therefore analytical), there still remains a large set of cogent arguments which are not valid. Can they be transformed into logically valid arguments as well? Let us look at the following arguments:

1) **Mr. Wilkinson is a Protestant**, because *he is a Swede*.
2) **All swans are white** because *swan 1 is white* and *swan 2 is white* and … and *swan 1 000 000 is white*.
3) **Mrs. Smith will get lung cancer**, because *she smokes*.
4) **It will rain during the night**, because *there are clouds this evening*.
5) **Earl has a good mood**, because *he smiles all the time*.
6) **This substance is toxic**, because *Mr. Smart said so*.
7) **This item is red** because *it looks red*.
8) **There are no unicorns**, because *if there were, we would have known it*.
9) **Mrs. Smith should not smoke**, because *smoking causes lung cancer*.
10) **Mr. Dread is the best writer**, because *his books are the scariest of all.*

If we recall our definition from the first chapter, cogent arguments were such arguments whose premises presented some substantial reason for accepting the claim. All arguments on the list seem to be cogent.

However, the arguments are not valid as there still can be some exceptional cases when the premises of any of the arguments would be true while its claim would turn out to be false. The premises of the arguments do not exclude such possibility. For example in argument (7) the item mentioned could be white and just look red because of a red light.

Therefore the arguments cannot be formally valid. The truth of the claim is not established due to meaning only, but requires additional justification. However, unlike in case of formally valid arguments, this justification cannot be simply added to the arguments as another linking premise, so they cannot be turned into logically valid arguments.

For example if we would have added a premise 'All Swedes are Protestants.' to the argument (1), we would turn this argument into a valid argument, which would be also logically valid. However unlike in the case of 'Lassie argument', this one would be an unsound argument, as this new premise would false. The new argument would then be of no use for rational argumentation.

But does that not mean that argument (1) is unsound as it is already? Perhaps the claim of (1) should rather be: 'Mr. Wilkinson is probably a Protestant.' In this case we could add several linking premises to this argument, including premises about precise observations of the ratio of Protestants among Swedes and some other premises connecting these observations with the claim (for example theorems of probability), and we could derive this modified claim in some logical formalism.

But do we really use the word 'probably' in such a complex manner? Certainly someone could claim we do, but this opinion is usually considered quite extreme. In the traditional paradigm it is commonly assumed that while all valid arguments are analytical, cogent arguments, on the other hand, cannot be analytical.

Now does that mean that cogent arguments cannot be formalized as formally valid arguments can?

In some literature (see, for example, Feldman, 1999), the cases of argumentation like those of our example are treated using certain argumentation patterns. In some cases, we are told, the cogent arguments have some distinct common pattern, although this pattern cannot be formalized so well as in the case of formally valid arguments, as it is still unclear what other information is needed to make the claim the inevitable conclusion of the premises. Those patterns are called 'argumentation schemes', but they are not proper argumentation schemes in the sense of the definition in Section 2, so I will refer to them as argumentation patterns.

Here are few examples of argumentation schemes:

Probabilistic arguments (1), inductive argument (2), causal arguments (3), predictive arguments (4), arguments of appearance (5, 7), arguments of authority (6), arguments of ignorance or lack of information (8), moral arguments (9), aesthetic arguments (10) etc.

In addition to the fact that some arguments can be classified as following such a scheme, some additional qualifications, depending on the argument scheme, have to be fulfilled, so that the argument could be considered as rational. For example we are told that the argument (6) is correct if 'The problem of determining whether a substance is toxic belongs to the field of chemistry and Mr. Smart is an expert in the field of chemistry.' and 'Mr. Smart has no reason to lie to us.' and 'Mr. Smart is not drunk.' and 'We understood Mr. Smart well.' etc.

Now it is obvious that if all those qualifications were added to the argument, together with an appropriate linking premise, we would again obtain a genuine logically valid argument. However there are few problems with such an approach.

In most cases exhaustive list of qualifications cannot be given and even if it could be, it would be probably so long and complex that checking whether all these premises are true would be almost impossible and there would be no reasonable procedure to determine, whether the argument is sound.

In case of some arguments it would be even difficult to state the truth conditions of those additional premises, as for example in case of ethical (9) or aesthetic

arguments (10). Is the argument (9) sound on the basis that the linking premises 'If something causes cancer it is a bad thing to do.' and 'If something is bad thing to do Mr. Smith should not do it.' are true? How do we evaluate truth of such statements? Is not an argument that has such statements as implicit premises already valid, as those explicit premises bring no new reason or support to the original arguments (9) or (10)?

It seems that this notion of 'substantial reason' cannot be explained without reference to psychology or some epistemological or metaphysical doctrine. Recognition of cogent arguments, as is widely acknowledged, cannot be carried out simply by application of some determinate algorithm[15] and requires some deeper understanding of human cognitive faculties.

Cogent arguments therefore must be treated within a discipline called informal logic. Informal logic can therefore be defined as logic minus formal logic as it is the science of rational arguments which are not analytical.

Some textbooks on informal logic cover also some elementary topics from the theory of argumentation (argument reconstruction, identification of fallacies etc. – see Feldman, 1999, once more), but in my understanding informal logic should be purely the science of rational arguments.

In other books (for example Fisher, 2004), formalization of cogent arguments is minimalized and they are studied in natural language. Informal logic therefore figures in theories of legal reasoning or critical thinking movement in education, described for example in Fisher (2001), rather than in mathematical discourse.

But wouldn't it be worthwhile to determine all qualifications of cogent arguments (despite the fact that it might not be always clear whether they are justified or not) and treat cogent arguments on the basis of their form? After all, the goal of logic is the determination of rationality arguments and not determination of their soundness and methods of formal logic are already well-developed.

While it might have been possible, there are certainly reasons why not to do so. Although such achievement would greatly help the logical evaluation of arguments, it wouldn't make their evaluation for the purposes of the theory of argumentation any easier, as the burden of evaluation of soundness of arguments would remain.

An influential book of Toulmin (2005) also opposes such an ambitious project. Toulmin observes that justifications of arguments are often presented in terms of impossibility. We argue that the claim of an argument must be true, on the basis of the premises, because the other case would be impossible.

But these impossibilities need not be always of the logical kind as in the case with logically valid arguments, not even to be the result of fixed meaning of some terms involved in the argument, as in the case of analytical arguments. Take for example following statements, uttered by a member of a logic department and addressed to some students:

A) You cannot levitate!
B) You cannot eat sandwich in our computer lab!

[15] This resistance of some rational arguments to proper formalization is explained as the gap between the art of practical reasoning (logica utens) and formal science about reasoning (logica docens), for example in Haack (2000).

C) You cannot call each teacher here 'professor'!
D) You cannot find a biggest prime!
E) You cannot affirm 'It is red and it is square' and 'It is not red' at the same time!

There are different reasons for those claims about impossibility, which could be made explicit in an argument as premises. In the first case it would be the body of general laws of physics that would justify this statement, the second statement is justified by certain obligations and prescriptions, the third points out certain linguistic and social conventions, the fourth one is justified because of the meaning of the terms 'prime' and 'biggest' and therefore a justifying argument for this command would have to be analytical, while the argument for the fifth statement would be valid because of the meaning of the terms 'and' and 'not' and would therefore be logically valid as it could be formalized within the sentential calculus.

Because each argument can be laid out in terms of some kind of impossibility or impropriety, Toulmin doesn't distinguish valid and cogent arguments and rather divides rational arguments directly into analytical and substantial. To motivate this distinction he introduces his own model of an argument.

The argument is not just a structure of premises (data) and claim (conclusion), but also of warrant and backing. The warrant is some kind of a rule that enables us to 'jump' from the data to the conclusion, while backing are some extra data, which in turn enable us to use this warrant. In case of logically valid arguments warrants can be understood as derivation rules and backing of those warrants can be provided by some semantics of logical connectives. The data together with warrant and backing constitute a qualification of the argument.

Further Toulmin introduces a concept of rebuttal, which is a piece of information that in turn invalidates the whole argument as it invalidates the whole qualification. It can undermine either backing of the warrant or the data. Suppose now we turn the statement C) into an argument: 'You cannot call each teacher there professor, because there adjective 'professor' refers to an academic degree, rather than the profession of teacher.' The backing of this explicitly stated warrant would be a convention that the university people are addressed by their academic degrees. The data here are represented by an implicit presumption that on this university some teachers are not professors.

The rebuttal of the data would be the fact that each teacher of the department would indeed have an academic degree of professor, the rebuttal of the backing would be that on this particular university different conventions hold.

Toulmin now suggests that logic should be a science of evaluating those reasons and warrants, some kind of jurisprudence of common arguments. Therefore not only logically valid arguments[16] should be subject of logic, but those arguments whose warrants are of some other kind should be studied by logic as well.

Toulmin further defines analytical arguments as those that become logically valid when we add the backing to the data. The remaining rational arguments are sub-

[16] He does not use the concept of logically valid arguments, but he recalls such classical reasoning patterns as syllogism or modus ponens, when he wants to speak about formal logic as it is traditionally used.

stantial[17]. To evaluate substantial arguments, one needs to look at the content of the premises and the claim and evaluate it due to the field-dependent criteria.

An example of a substantial argument is an argument presented in the previous chapter as the argument (8). I will rephrase this argument here for the greater convenience of the reader.

8) **There are no unicorns**, because *if there were, we would have known it*.

One can accept this argument as rational, because the argument about existence of some animal species belongs to the field of empirical science, where the principle of 'Occam's razor' holds, which tells us to avoid the stipulation of the existence of an entity unless there are some serious reasons for it. This principle is a backing for the warrant of our conclusion.

On the other hand, a physician should not accept arguments like 'The patient does not have cancer, because if he did, we would have known.' and rather perform more thorough diagnostics of the patient, because of the life risk. Also a judge should not just accept the argument 'The suspect committed the murder, because we have no evidence that he didn't.', because of the presumption of innocence, and he should rather free the suspect.

We have seen that although those arguments have a similar form, some of them are valid and some of them are not. Therefore the 'unicorn argument' cannot be sound on the basis of its form only and therefore cannot be analytical in principle.

The idea that all rational arguments can be reconstructed as analytical is what Toulmin calls an 'analytical ideal'. Due to the analytical ideal there are no genuinely rational substantial arguments. All rational arguments are analytical arguments or analytical arguments in disguise. According to the analytical ideal, to evaluate an argument one must only look at its form, possibly adding some self-evident premises, to determine whether it is valid, and then at its premises to determine whether they are true or justified. If both of these tests are successful, the argument as a whole is sound.

Analytical ideal implies that the rationality of arguments is something that is field-independent and is determined only by the universal laws of formal logic, which hold under any circumstances. Logic is therefore understood not as jurisprudence of arguments, but as a system of eternal truths. Cogent arguments are not rational as they lack some important proposition used as explicit justification in the premises. Some cogent arguments can be redeemed, if the missing justifications are added to the premises together with a liking premise, and so turned into analytical arguments. Each analytical argument that is not logically valid can then be turned into one. Therefore each rational argument can be turned into a logical one. Those arguments which cannot be turned into logically valid arguments must be disregarded.

But how can one defend this ideal in face of the difficulties mentioned? Toulmin identifies three attempts to fix this problem: Transcendentalism, phenomenalism and skepti-

[17] Given our identification of analytical arguments with valid arguments we can now identify substantial arguments with cogent arguments, but if we admit that there are valid non-analytical arguments there will be valid substantial arguments.

cism. While the first two approaches aim at bridging the 'logical gulf' between premises and the conclusion of substantial argument, the skeptical approach condemns substantial arguments as irrational, as arguments that simply cannot be justified by reason alone.

Transcendentalists turn substantial arguments into analytical ones by interpreting backing of warrants as a background data, which we obtain thanks to some non-empirical faculties of cognition. In the past, philosophers of this tradition took recourse to such sources of knowledge as a divine revelation or pure reason, which provides us with the so-called truths of reason, or some kind of intuition.

These truths of reason are used as a linking premise. Therefore we complete seemingly substantial arguments with a major linking premise, often in the form of some general statement, or implication and turn them into analytical ones.

Thus for example scientific judgments are justified, because we are endowed with knowledge of eternal truths of reason. Also other faculties of cognition can be considered: for example our conscience (categorical imperative, sense for common good etc.) when we are confronted with moral arguments, or feelings when we are confronted with aesthetic arguments.

The phenomenalist (or reductionist) approach, on the other hand, turns the claim of the argument into something which in fact is contained in our sensual data. Therefore claims about objects are in fact disguised claims about our recent collections of sensual data.

Toulmin gives an example of the application of the reductionist approach to arguments about other people's mental states in behaviorism. Behaviorists claim that statements about a person's feelings, emotions, thoughts etc. are in fact statements about the person's observable behavior.

Skepticism just dooms all substantial arguments as irrational. Such is an approach of David Hume, who's skepticism leads him to believing that we accept most arguments as a matter of mere habit rather than because of rational justifications.

Let me now summarize the main points of Toulmin's critique and draw some conclusions.

To recognize all rational arguments we need more than formal logic. But it is still questionable whether this recognition can be carried out in a scientific manner, using determinate algorithm or at least some heuristic method, or whether recognition of rational arguments isn't rather some kind of art than a matter of science.

Of course, informal logic can help us understand arguments and their use and it has for sure, certain scientific qualities, even when it does not rely on formal algorithmic method. After all, human sciences also deserved to be recognized as sciences and informal logic seems to fit into this pigeonhole of humanities, as it is concerned with human language, in particular its fragment involved in argumentation. Perhaps informal logic should be sufficient supplement of formal logic that could complete it in order to create the ultimate science of arguments.

But if we are satisfied with this solution, we have probably not taken Toulmin's critique seriously enough. Toulmin insists that logic should not be a science in any meaning of the word. He wishes logic to be a jurisprudence, rather than the science of argumentation. Logic as jurisprudence or art of arguing is quite appealing, but such endeavor could hardly be performed systematically and with the scientific rigor.

If logic is indeed supposed to be a science of arguments, it should not evaluate just singular tokens of arguments, but rather classify them first and then evaluate those classes of arguments instead. Botanists also do not describe each single plant by listing its attributes, but by observing particular plants they make up concepts and classes, then they categorize individual plants into the classes by their attributes and they further study those classes and their relation to each other in the whole taxonomy.

The need of classification of arguments for the purpose of their study in logic is even more important than the botanic classification. It is so, because there is only a finite number of plants on the planet, but the number of potential arguments one can use in an argumentation is unlimited.

We have seen that arguments can be classified according to their logical form. Formal logicians also create concepts of logical forms (designing new logical formalisms), categorize particular arguments into those forms (formalization of natural language arguments) and ascribe attributes to all arguments sharing the same form (description of mathematical properties of logical formalisms). Such strict classification has its limits as we have seen.

Informal logicians also aim at some kind of classification of arguments using argument patterns, despite the fact they cannot provide any uniform treatment of arguments within the same class and despite the fact that it is often unclear to which argumentation pattern a particular argument should conform. Such classification again must be carried out without referring to the context of use of those arguments and does not help us, in most cases, distinguish rational arguments from irrational ones, as we have seen in the case of the unicorn-cancer example.

Is therefore a science of arguments not possible? Is logic possible? I am sure there can be logic as a general science of arguments, but not as a science that tells us what rational arguments are, on the basis of the argument alone. While arguments can certainly be valid, and their validity can be surely recognized with highly scientific methods of formal logic, the notion of rational argument should be abandoned for good. An adjective 'rational' should rather be used to describe argumentation process as a whole.

We still might be able to discover some common feature to all cases of rational argumentation, which logic could help us evaluate. We just need to concentrate on the argumentation as a whole and give logic new purpose with this new paradigm of argumentation in mind.

4. Logic and argumentation revisited and what use could informal logic be

Now let us turn attention to the second paradigm of argumentation.

In the previous section we have seen that substantial arguments cannot be recognized on the basis of their form only, for one instance of the same argument scheme might be a rational substantial argument, while some other may be not. Whether some substantial argument is rational depends also on the fact whether all of its qualifications are met. The determination of satisfaction of the argument's qualification (much like the determination of the justification of premises), seems to be something that cannot be handled with formal methods without consulting empirical reality.

Substantial arguments are therefore arguments with incomplete information. Some information is missing to determine whether the argument is rational or not and that is the information expressed in the form of the argument's qualifications. However we often do not require information that this qualification is justified in order to consider the argument valid, though any evidence that the qualification is not met would lead to its instant invalidation. So the qualifications are something that is presumed to hold. Reasoning with substantial arguments is therefore reasoning with incomplete information and also presumptive reasoning[18].

An example of a presumptive argument is an argument using a presumption of innocence. All citizens are presumed to be innocent, unless they are found guilty. The 'unless' here indicates rebuttal of the argument. Substantial arguments are therefore defeasible. Their claims can be retracted from the list of conclusions of some argumentation process, if some information becomes available during this process, which invalidates the use of these presumptive arguments.

So it is important that we are presuming innocence here, and not for example, competence in piloting Boeing, because an argument 'This man should be allowed to pilot this flight, unless it is proved he is an incompetent pilot.' wouldn't be an example of a rational argument. So the fact that innocence can and must be presumed, while piloting competence mustn't, is an important factor in evaluating such arguments.

The most straightforward formalization of the ideas of qualification and rebuttal is that used in default logics, which are described for example by Antoniou (1997) or Poole (1994).

Default logics are an enhancement of a logical formalism with a set of the so-called default rules. Default rules, unlike logical rules, have three components: beside premises and a claim also a qualification. A rebuttal of this qualification is its negation (default logics therefore must have an explicit symbol for negation in their language). Given a set of formulae T, an extension of T is defined as a superset of T closed under application of logical rules and default rules, whose rebuttal is not in this actual extension. This definition seems to be circular, but can be handled using fixpoint definitions.

Extensions also have equivalent procedural description using an algorithm generating a tree which branches depending on the default rules being used. This algorithm is described for example by Makinson (2005), and in greater detail by Antoniou (1997).

Another example of presumptive reasoning is abductive reasoning. Abduction is a process where we move from particular facts as premises to a general laws that would explain them as a conclusion[19].

A useful insight into abductive reasoning is provided by Poole's Theorist formalism, described by Antoniou's (1997) and Brewka (1991). Poole suggests that we need not to design new logical formalisms, but to alter our using of existing ones, if we want to de-

[18] In Toulmin's model, this idea of qualification is elaborated into concepts of warrant and backing. In the models I will introduce the warrant will be represented by some default rules, while the backing of this rule will not be represented in the models, as it has the same role as data, which are presented as the premises of the argument. The justification of particular warrant as being applicable in such and such case is something that lies outside the scope of logic, as does the justification of the premises.

[19] The presumptive and defeasible nature of explanation, prediction and abductive reasoning in general is explained in great detail by Walton (2004).

scribe such kinds of rational activities as explanation or prediction. We should enhance logical theories consisting of a set of facts (axioms) by an additional set of hypotheses. A scenario is a set of facts united with some subset of hypotheses, consistent with the facts. An extension is a maximal scenario with respect to set inclusion.

There are several other formalisms that aim at a description of defeasible, plausible or commonsense reasoning like nonmonotonic modal logics (NML), or autoepistemic logics (AEL; see Brewka, 1991, or Konolige, 1994) using modal language to describe beliefs of an ideal introspective agent reasoning about these beliefs and the lack of them, to mention just few.

Now for all the formalisms mentioned above, the inference relation can be defined as either credulous (a formula A can be inferred from a set of formulae T if it is in some extension of T) or skeptical (A can be inferred from T if it is in all extensions of T). This inference relation is often nonmonotonic, in the sense that if we add more premises to some instance of it we do not get more conclusions. Therefore those formalisms are called nonmonotonic logics.

The properties of nonmonotonic inference relation are described in great detail in Makinson's summary (1994). Although Makinson shows that such inference relation of known nonmonotonic formalisms often share at least some reasonable properties with more common inference relation, it is quite debatable whether this is sufficient to consider nonmonotonic formalisms as logical formalisms.

For example: An inference relation that is nonmonotonic cannot be described using classical Hilbert calculi, because each proof of formula C from the set of formulae $A_1, \ldots, A_n$ is by definition also a proof of C from its superset $A_1, \ldots, A_n$, B.

Makinson (2005) further proves that each supraclassical (extending classical) inference relation over a propositional language is either total or it is not closed under the substitution of formulas for atoms.

For this reason it seems that nonmonotonic logics, resulting from the addition of some kind of defeasible rules to logical rules, cannot have standard model-theoretic or algebraic semantics.

Semantics of nonmonotonic logic does not fit within the old paradigm well. Although if we define an inference relation in them, we will have problems when we will try to describe it using standard methods of formal logic. In fact the nature of concepts of nonmonotonic logics is dialectical, as is mentioned in Prakken's and Vreeswijk's chapter in the *Handbook of Philosophical Logic* (2002).

Prakken and Vreeswijk argue here that the interpretation of defeasible rules cannot be founded on classical concepts like 'correspondence to the real world', 'truth' or 'entailment', but rather on notions like the 'burden of proof', 'attack', 'defeat', or 'rebuttal'.

The term nonmonotonic logic is therefore quite misleading, as nonmonotonic formalisms are not logical formalisms in sense of Chapter 2. Their purpose is not to recognize valid arguments, but to enlighten some procedures used in rational argumentation, such as making and retracting of defeasible arguments (default logics), using presumptions (Theorist), or reasoning with uncertain information (NML, AEL).

So if we are looking for justification of nonmonotonic logics, we should rather seek among formal approaches to describing argumentation that are not extensions of logi-

cal formalisms. One of them, capable of describing semantics of most nonmonotonic formalisms has been mentioned in the first chapter; it is the abstract approach to argumentation introduced by Dung (1995).

Dung's formalism offers natural semantics for nonmonotonic logics. The procedural nature of semantics of nonmonotonic formalisms is indeed related to procedures involved in argumentation. This fact is made quite obvious in an article by Bondarenko, Dung, Kowalski and Toni (1997), where semantics of great deal of nonmonotonic formalisms is described via Dung's abstract argumentation semantics, within so-called abduction frameworks. The extensions of nonmonotonic formalisms can be interpreted as the extensions of Dung's argumentation frameworks, where arguments are represented as assumptions (qualifications in default logics, hypotheses in Theorist or beliefs in NML or AEL) and one such argument defeats another if and only if it is possible to infer from it, together with some fixed set of facts, the contrary of the second argument.

So ABF's consist of some underlying logical theory (described by language, arbitrary relation of inference and arbitrary total mapping of contrariness among formulae of the language) and a set of presumptions.

It is obvious that the concept of contrariness is primary here, while the concept of inference is mandatory. Any reference to logic could be omitted as well as any reference to the set of undisputable facts and the relation of defeating could be simply identified with the relation of contrariness among two assumptions, which would lead to a minimalist theory of argumentation.

In default logic formalism this approach would result into a default theory in language consisting only of literals and with no role for logical rules, just the default ones – we would just drop the requirement that the extensions have to be closed under classical consequence.

Perhaps we should abandon such concepts as 'valid argument', 'truth', 'consequence', or even 'inference' and 'proof' and rather concentrate on the concept of counterargument, explaining it without a reference to previous concepts. We would get rid of the problematic concepts and give logic a more reasonable goal, which is defining of the relation of contrariness between arguments.

One possible approach to defining this relation is using traditional definition of an argument and defining the defeating relation in term of inconsistency of the claim of one argument with premises of another. This strategy has been followed in Besnard's and Hunter's (2008) recent monograph, where classical arguments of propositional logic are studied from this perspective.

However major problem of such an approach is that if we define arguments using an inference relation of classical logic and then define relation of defeating using its negation, the resulting Dung's argumentation frameworks become quite trivial. Besnard and Hunter (2008) therefore develop their own formalism for evaluating argumentation, based on the concepts of argument trees and structures, rather than on argumentation frameworks, and mention that nonmonotonic logics might be more suitable for defining arguments.

Instead of defining arguments as structures of formulae of some logical formalism, it is perhaps more promising to interpret them either as strict or defeasible rules. The

notion of defeasible rules is captured in defeasible logics (see Nute, 1994, or Prakken and Vreeswijk, 2002). Defeasible logic also introduces defeating rules that might defeat defeasible rules. However defeating rules must not be made explicit and the relation of defeating among rules might be defined on meta-logical level.

The improvement over an approach based purely on a strict undefeasible inference of logical formalisms is that defeasible rules can be defeated with some strict rule, while the same defeasible rule cannot defeat the strict rule so the resulting relation of defeating is less trivial. The review of such approach and its advantages and shortcomings is summarized in an article by Caminada and Amgoud (2007).

Using nonmonotonic or defeasible logics for defining arguments therefore seems to be the right path to take for the future research of argumentation, as is also suggested by Besnard and Hunter (2008).

The new paradigm suggests that we abandon such questions as 'Is this argument rational?' that is undecidable without a reference to a context, with questions like 'What set of arguments is admissible within a given constellation of arguments?'. This question can be answered using formal methods of abstract argumentation, while it is capable of modeling natural human reasoning in more faithful manner.

The problematic question whether some argument is rational is dismissed. Cogent arguments are evaluated not in relation to some implicit and unclear qualifications, but in the well-defined relation of defeating to other arguments that might refute them.

In the previous section we have seen that the defeasible rules are as important for understanding argumentation as the strict ones that can be defined within logical formalisms. However previous section begs even a more 'heretical' question, which needs to be addressed: 'Do we need logical formalisms to understand argumentation at all?'.

We have seen that the enhancement of some logical formalism with defeasible rules is sufficient for dealing with shortcomings of traditional paradigm, pointed out in Toulmin's critique, and that defeasible logic is capable of formalizing argumentation in a more faithful manner.

On the other hand we have also seen that nonmonotonic logics are not true logical formalisms as we have defined it, as the consequence relation cannot be described with usual methods employed in the study of logical formalisms.

But on the other hand it still seems that the definitions of basic concepts of defeasible logic are dependent on definitions of logical formalisms and that all nonmonotonic logics are in fact just supraclassical enhancements of those formalisms with defeasible assumptions (Theorist), defeasible valuations (Circumscription, preferential entailment, logic programming) or defeasible rules (defeasible logics, defeasible inheritance nets, default logics, nonmonotonic modal logics, autoepistemic logics).[20]

But again on the other hand this fact proves nothing more than that this reliance on classical concepts has been a design decision of their inventors.

We can omit strict rules from defeasible logic, as well as we can drop the requirement that extensions have to be closed under the inference relation of some logical formalism within default logic or Theorist, while keeping the notion of contrariness as our primitive

[20] See Makinson's book (2005) for motivations of this classification.

logical notion. We also do not need any logical connectives to serve as logical constants beside negation in Theorist and Default logics and we do not need one in defeasible logic, due to the existence of defeating rules.

It is not so obvious how we could get rid of a reference to the underlying strict inference from others non-monotonic logics (NML, AEL), however the observation that even their semantics can be defined within abduction frameworks suggest it would be possible. Within ABF's we can simply drop any reference to the strict inference relation and just define that one hypothese defeats another if it is its contrary. If we want to include the reference to the stable set of facts, we can do so provided that we will have a concept of joint contrariness not as the relation among single formula, but sets of formulae instead. One hypothesis would then defeat another, if the set of facts united with the first hypothesis would be contrary to the second hypothesis.

So we see it is possible to describe argumentation without an aid of logical formalisms that would help us recognize the strict inference, but is it reasonable to do so?

It is obvious that there are cases of argumentation which are undoubtedly wholly rational, but in which formal logic plays no role. There simply are no indefeasible arguments, represented by logically valid strict rules, involved in them, just undefeated ones, so the knowledge of rules of inference would not help us to evaluate this particular case of argumentation at all.

On the other hand formal logic is a respectable science and it would be hasty to ignore its results. The fact that formal logic alone cannot serve the purpose for which it was designed millennia ago by Aristotle does not mean it cannot serve any purpose at all. Formal logic has found its place in the system of sciences as the science of inference, rather than the science of argumentation in general. It does help us to understand argumentation better, if not always, then at least in most interesting cases. Formal logic has its justification in mathematics, but it is important for understanding argumentation as well. There are indefeasible inferences after all and there are even indefeasible inferences in natural languages, not just in formal languages of mathematics. So both mathematics and philosophy can benefit from formal logic, while the idea of science of universal methods of rational reasoning can be abandoned without any substantial loss.

Now I would like to summarize the main claims, which I hoped to establish in this article.

To understand argumentation it is often useful, though not necessary, to distinguish its logical and extra-logical components. In the traditional paradigm logic provides arguing agents with sufficient inferential mechanism, described more or less completely by rules of inference (sometimes understood as 'laws of logic'), while all the extra-logical information relevant to the process to argumentation is reduced to pure static data, often interpreted either as recognized facts or beliefs of the agents. We are told that this information can be added to premises of any of the arguments an agent is making together with a linking premise, because it is justified by the agent's 'background' knowledge or because it is a self-evident truth of reason. Justification of these premises is therefore considered as independent of logic.

However, as we have seen, this approach is quite inadequate to encompass substantial arguments without relying on quite problematic epistemological assumptions.

Now there are two ways out of this dilemma: Either to deny any relevance of universal schemes of argumentation for understanding particular cases of practical argumentation at all, and plead for another kind of logic than those presented by formal logicians, or to simply call any, no matter how cumbersome or *ad hoc* description of inference *logic* and dissolve the focus of formal logic into helpless categorizing of particular instances of inference.

What I believe is a solution to this Scylla-Charybdis problem is to recognize the importance of formal inference in many cases of argumentation, where the language of discourse allows some kind of schematization of some of the valid inferences, due to the shared meaning of logical constants (as is for example the case in mathematics), but admitting that many cases of reasoning are sanctioned by field-dependent information as well.

Formal logic as a science of logical formalisms is important for the development of argumentation formalisms, in which arguments can be evaluated reasonably and which should be the domain of logic in the broadest sense of the word.

References

Antoniou, G. (1997): *Nonmonotonic Reasoning*, MIT Press, Cambridge (Mass.).

Beall, J. C., Restall, G. (2006): *Logical Pluralism*, Clarendon Press, Oxford (NY).

Bench-Capon, T. J. M. & Dunne, P. E. (2005): *Argumentation and Dialogue in Artificial Intelligence*, http://www.csc.liv.ac.uk/~comma/IJCAI_05_Tutorial.pdf.

Bench-Capon, T. J. M. & Dunne, P. E. (2007): 'Argumentation in artificial intelligence', *Artificial Intelligence* 171.

Besnard, P. & Hunter, A. (2008): *Elements of Argumentation*, MIT Press, Cambridge (Mass.).

Bondarenko, A. & Dung, P. M. & Kowalski, R. A. & Toni, F. (1997): 'An abstract, argumentation-theoretic approach to default reasoning', *Artificial Intelligence* 93.

Bochenski, J. M. (2002): *Formale Logik*, Alber Studienausgabe, Freiburg (München).

Brewka, G. (1991): *Nonmonotonic Reasoning: Logical Foundations of Commonsense*, MIT Press, Cambridge (Mass.).

Caminada, M. (2008): *A Gentle Introduction to Argumentation Semantics*, http://icr.uni.lu/~martinc/publications/Semantics_Introduction.pdf.

Caminada, M., Amgoud L. (2007): *On the evaluation of argumentation formalisms*, http://icr.uni.lu/~martinc/publications/AIJ-postulates.pdf.

Coffa, A. J. (1991): *Semantic Tradition from Kant to Carnap*, MIT Press, Cambridge (Mass.).

Dung, P. M. (1995): 'On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games', *Artificial Intelligence* 77.

Feldman, R. (1999): *Reason and Argument*, Prentice Hall, Upper Saddle River (NJ).

Fisher, A. (2004): *The Logic of Real Arguments*, MIT Press, Cambridge (Mass.).

Fisher, A. (2001): *The Critical Thinking an Introduction*, MIT Press, Cambridge (Mass.).

Haack, S. (2000): *Philosophy of Logics*, MIT Press, Cambridge (Mass.).

Konolige, K. (1994): 'Autoepistemic Logic', *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3 (eds. Gabbay, D. M. & Hogger, C. J. & Robinsion, J. A.), Clarendon Press, Oxford, p. 217–438.

Lorenzen, P (1965): *Formal Logic*, Reidel, Dodrecht.

Makinson, D (2005): *Bridges from Classical to Nonmonotonic Logic*, King's College Publications, London.

Makinson, D. (1994): 'General Patterns in Non-monotonic Reasoning', *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3 (eds. Gabbay, D. M. & Hogger, C. J. & Robinsion, J. A.), Clarendon Press, Oxford, p. 35–110.

Nute, D. (1994): 'Defeasible logic', *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3 (eds. Gabbay, D. M. & Hogger, C. J. & Robinsion, J. A.), Clarendon Press, Oxford, p. 353–395.

Peregrin, J. (2004): *Logika a logiky*, Academia, Praha.

Poole, D. (1994): 'Default Logic', *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3 (eds. Gabbay, D. M. & Hogger, C. J. & Robinsion, J. A.), Clarendon Press, Oxford, p. 189–215.

Prakken, H. & G. Vreeswijk (2002): 'Logics for Defeasible Argumentation', *Handbook of Philosophical logic*, 2nd ed. vol. 4 (eds. D. M. Gabbay, F. Guenther), Kluwer, Dodrecht, p. 219–318.

Tarski, A. (1994): 'What are logical notions?', *Limits of Logic* (ed. Shapiro), Clarendon Press, Oxford, p. 119–130.

Toulmin, S. (2005): *The Uses of Argument*, MIT Press, Cambridge (Mass.).

Van Daalen, D. (2004): *Logic and Structure*, Springer Verlag, Berlin.

Walton, D. (2004): *Abductive Reasoning*, University of Alabama Press, Tuscaloosa (AL).

2007 ACTA UNIVERSITATIS CAROLINAE
PHILOSOPHICA ET HISTORICA 2
miscellanea logica (viii)
PAG. 71–112

Tarski's Method of Truth Definition: Its Nature and Significance

LADISLAV KOREŇ*

Department of Philosophy and Social Sciences, University of Hradec Králové

In the classic work, 'The Concept of Truth in Formalized Languages' (CTFL), Alfred Tarski set out to examine thoroughly under what conditions and by what methods it is possible to construct a satisfactory definition of the notion of truth as predicated of sentences.[1] In the end, what he achieved was not a definition of the general notion of truth, not even of sentential truth, but a general method of defining a truth-predicate restricted to sentences of some given language L, where L belongs to a comprehensive group of formalized (or formalizable) languages of a certain sort. Tarski's method of truth definition and his approach to semantics in general has various logical, philosophical and mathematical aspects, owing to the fact that truth is a notion that plays a very special role in mathematical logic as well as in philosophy, in which disciplines Tarski had both interest and education.[2] However, its reception in these disciplines has been very different. Logicians have concentrated mainly on 'formal' aspects of Tarski's method: the analysis and solution of semantic paradoxes, definability and indefinability theorems, formal machinery of semantic definitions and the relations between (recursive) meta-mathematical and (explicit) set-theoretical definitions, etc. In their view, Tarski showed how to define truth and related semantic notions by precise logico-mathematical methods, and they have been fairly widely agreed that his method of truth definition is a seminal contribution to their discipline. Philosophers, on the other hand, have focused more on 'material' aspects of the method: the adequacy criterion based on the so-called semantic conception of truth, the philosophical plausibility of the semantic conception of truth

* Work on this paper has been supported by the research grant No. 401/06/0387 of the Czech Science Foundation.

1 CTFL first appeared in Polish (Tarski, 1933), afterwards (revised and enlarged) in a German translation (Tarski, 1935), its English translation coming last (Tarski, 1956). Unless I indicate otherwise, I quote from the 2nd revised edition (Tarski, 1983). Other relevant papers are: 'Grudlegung der Wissenschaftlicher Semantik' – I shall quote from its English translation in Tarski (1983); 'Semantic Conception of Truth and Foundations of Semantics' – Tarski (1944); and 'Truth and Proof' – I shall quote from the reprint in Tarski (1993).

2 Tarski said of himself that he is 'a mathematician (as well as a logician, perhaps a philosopher of a sort)'; Tarski (1944: 369).

and its relation to other conceptions of truth, etc. Here the question of its importance has continued to be the subject of an ongoing controversy, whose participants argued sometimes truly antithetical claims. Whereas some people claimed that Tarski gave us a satisfactory explanation of the notion of truth and solved the problem of truth, other people argued that Tarski-style truth definitions so blatantly fail to explain the notion of truth that Tarski's method can have no philosophical value.

In this paper I want to argue for a different view of Tarski's method and its philosophical significance which, I believe, is more complex and plausible. To arrive at such a view, we need to understand: what philosophical, logical and mathematical aspects Tarski's method has, what role they were supposed to play in his project of scientific foundations of semantics, and which of these aspects hang in together (and how), and which should be kept separate. I start with a detailed exposition of the conceptual and technical underpinnings of Tarski's method of truth definition, situating it within his broader project of theoretical semantics. Having this in place, I pass on to discuss the question of its significance, reviewing a number of arguments that purport to show that the method fails as an explanation (or explication) of the ordinary notion of truth, in particular, that it is a confusion to think that it has semantic import. From this discussion the following view emerges. The critics are right that formal truth definitions in Tarski-style do not explain (explicate) our ordinary notion of truth. However, even though such definitions fail to explain (explicate) the ordinary notion of truth, it does not follow that we cannot think of Tarski's method as providing us with a valuable insight into truth of a different sort, namely as giving us a workable model of how truth conditions of sentences of a properly formalized language depend on semantic properties of their significant parts and their syntactic structure. This, of course, is how Tarski's method is viewed by those theorists who see in it the foundation of formal approach to semantics. What the critics overlook or do not appreciate is that there is more to Tarski's method than the particular formal definitions for particular languages. Its heart is the material adequacy criterion stated in Convention T which assures that a truth definition that meets it captures all and only the true sentences of a given language (the extension of truth with respect to that language), where a good intuitive grasp of the ordinary notion of truth is presupposed in the form of the semantic conception of truth. The formal technique in case of reasonably rich languages is recursion on a generalized semantic relation between expressions and objects, in terms of which truth for the language is defined in such a way that materially correct statements of truth-conditions can be delivered for each of the indefinite number of its sentences. In tandem, the two moments indicate to us where to look for the semantic import of Tarski's method, in which its philosophical value largely consists.

1. Tarskian semantics: between logic and philosophy

Tarski's aim in CTFL was to construct a satisfactory definition of sentential truth that satisfies the conditions of formal correctness and material adequacy. It is surely significant that he said that the problem belongs 'to the classical questions of philosophy'[3], or

[3] Tarski (1983: 152).

even, that the central problem (of establishing the scientific foundation of the theory of truth and semantics) 'belongs to the theory of knowledge and forms one of the chief problem of this branch of philosophy.'[4]

Semantics, as Tarski understood it, is concerned with the totality of considerations about the notions that express relations between linguistic expressions and extra-linguistic entities.[5] Let us adopt the following useful terminological suggestions due to Künne to make more precise this conception of semantics:[6]

> (A) *x* is a semantic notion (predicate) in the broad sense iff *x* expresses/signifies a property that only an expression can possess or a relation in which only an expression can stand to something such that *x* holds of an expression (of a language) in part in virtue of the expression's meaning (in the language);[7]
>
> (B) *x* is a semantic notion (predicate) in the narrow sense iff *x* is semantic in the broad sense and (a) *x* expresses a relation in which expressions stand (paradigmatically) to extra-linguistic entities, or (b) *x* is explained in terms of a broadly semantic notion that expresses such a relation.

We may call semantic notions falling under (Ba) 'directly relational' and semantic notions falling under (Bb) 'indirectly relational'.[8] Semantics, as Tarski conceived it, deals with narrowly semantic notions only, and is thus fairly restricted in its scope[9], because broadly semantic notions of meaning, synonymy or analyticity, which many would take to belong to the province of semantics, are neither directly relational on the face of them, nor is it clear that they are explainable in terms of directly relational notions.[10] Another question arises immediately. Clearly, notions such as denotation and reference (for singular expressions) or satisfaction and application (for predicates) belong to Tarskian semantics, since they are directly relational. But what should we say about the notion of truth as applicable to sentences? The word 'true' is a one-place predicate and hence it does not directly express a relation; so, if truth is a narrowly semantic notion belonging to Tarskian semantics, it is such in virtue of being explained in terms of a directly relational notion. Tarski indeed claimed that *the* intuitive notion of truth (or at least *one* intuitive notion of truth which he called *semantic*) is indirectly

---

4 Tarski (1983: 266–267).

5 Cf. Tarski (1983: 252, 401), (1944: 345).

6 See Künne (2003: 179). Tarski nowhere made the important qualification to the effect that semantic relations are those that hold between expressions and extralinguistic entities in part in virtue of the meanings of the expressions; and the qualification is often omitted in the literature. As Künne rightly points out, the predicate '*x* has more letters than *y* has legs' expresses a relation in which expressions stand to extralinguistic things, but it is obviously not a semantic predicate, since the relation expressed obtains independently of the meaning of a word that may be substituted for '*x*'.

7 'iff' is a short for a biconditional-forming connective 'if and only if'.

8 Cf. Tarski (1993: 112).

9 Cf. Tarski (1983: 401).

10 Although Tarski said that they may belong to the theoretical semantics after all, in spite of the fact that their 'intuitive content is more involved' and their 'semantic origin is less obvious'. See Tarski (1944: 354). In the footnote n. 20 he also referred to (1936b) for his classic definition of consequence and to Carnap (1942) for the definition of analyticity.

relational, in part because he sensed more than a grain of truth in the correspondence intuitions to the effect that *S is true just in case things are as S says they are*[11], and in part because he took truth to have certain intuitive connections to directly relational notions of denotation and satisfaction, which can (and in some cases, even must) be used in explaining it.[12]

However, semantics, conceived as a systematic theory of properties of narrowly semantic notions, was simply a non-existent discipline before Tarski embarked upon his own foundational investigations in the late 20s of the past century. Although truth and cognate semantic notions look fairly intuitive and easy to grasp both in their ordinary and technical usage (in logic, say), many of Tarski's contemporaries treated them with suspicion, and did not believe in the possibility of a systematic and respectable theory of their properties.[13] On the logical side of the problem, it was long known that the notion of truth gives rise to paradoxes of the Liar-variety, when a certain sort of self-reference is present in the discourse, and the same applies to the semantic notions of denotation, definition or satisfaction, for which similar paradoxes were ingeniously constructed (such as e.g. Berry's, Richard's and Grelling-Nelson paradox respectively). On the philosophical side of the problem, attempts at explaining truth in a clear and precise manner were not particularly successful; worse still, participants in traditional philosophical debates tended to entangle that notion in dubious speculations about metaphysical issues concerning the nature of reality, judgment, and the relation between them, and often couched their conceptions in terms that they seldom tried to give precise explications of. In reaction, some influential positivist theorists from Tarski's intellectual circle claimed that, both in its common and philosophical usage, truth is closely associated with the mysterious idea of a mirroring (correspondence) relation between language and language-independent reality that rejects explanation in scientifically (physicalistically) respectable terms (and much the same holds for intentional notions that apply not to language but to thought). The notion of truth, so understood, has no place in science or scientific philosophy, all the more so when semantic notions are logically ill-behaved, since they give rise to paradoxes. What empirical sciences really need is some workable epistemic notion of verification or confirmation, or, in the case of deductive sciences, some purely formal-syntactic notion of provability in a formal-deductive system. Both Neurath and Reichenbach held views of this sort.[14] While Neurath thought that in its common usage truth is absolute, which he held to be incompatible with the holistic and dynamic conception of scientific justification and knowledge, Reichenbach concluded much the same on

---

[11] For more on this see section 3.

[12] See Tarski (1944: 345). Being a pioneer in the theory of definability in mathematical logic, Tarski was well aware that the notion of satisfaction can naturally be used to explain the notion of definition, since a predicate defines an individual (or set of individuals of a given domain) iff the predicate is satisfied by that individual only (or by all and only the members of the set). The same applies to the notion of denotation, since *m* denotes *o* iff *o* satisfies the predicate '$x = m$'. Cf. Tarski (1944: 354), the footnote n. 20.

[13] Cf. Tarski (1983: 152, 401).

[14] See Neurath (1983b), Reichenbach (1938). Both proposed not to use in science the common notion of truth, which was in their view absolute, but rather some respectable epistemic notion (Neurath – coherence, Reichenbach – weight or degree of confirmation), which they tended to equate *with the only possibly useful notion of truth*.

the ground that truth is associated with absolute verification or certainty, which is unattainable in science.[15]

This tendency, to define truth using epistemic notions, is familiar in philosophy: to put what is a criterion or test of truth into the very definition of truth. Although Tarski shared the view that philosophical attempts to define truth were unsuccessful, and he had also some sympathies for the positivist's critique of the marriage of truth with metaphysics, he deemed epistemic conceptions of truth implausible. For such conceptions seem to violate the principle of excluded middle that intuitively holds of truth but not of epistemic notions. Thus,

X is true or X's negation is true

holds for any fully determinate sentence *X*, while

X is confirmable (provable) or X's negation is confirmable (provable)

does not hold in general, because there are many cases where we do not have enough evidence to decide a sentence one way or another.[16] Tarski's refusal to abandon truth altogether or equate it to some epistemic notion hangs in rather closely with his specialization in metalogic or metamathematics (what he called *the methodology of deductive sciences*). Metalogic studies properties of deductive disciplines axiomatized in formal-logical frameworks, as well as properties of the frameworks themselves. Now, consequence, validity and satisfiability are basic metalogical notions that have their intuitive semantic definitions in terms of truth (or in terms of the relative notion of truth in a structure), which notion is also needed to characterize consistency, soundness or completeness of a formalized deductive theory.[17] Tarski soon realized that the notion of truth, as it was used informally in metalogic, was intuitively different at the conceptual level from any notion that epistemic definitions could provide (even when the two happen to coincide, when a formalized deductive theory is complete). There were indeed proposals to equate truth, for purposes of formalized deductive theories, with a syntactic

[15] Also Carnap, under Neurath's influence and for a short period of time, seemed to subscribe to such a view. He soon became its critic and propounded Tarski's theory of truth. For Carnap's critique see his (1949); a good, critical discussion of Reichenbach's views is Soames (1999); a short exposition of Neurath and Carnap can be found in Candlish and Demnjanovic (2007).

[16] We shall have occasion to see that in making such claims Tarski had in mind only sentences that are fully interpreted in that their meaning/content is sufficiently determinate to render them either true or false. In this sense, he took truth to be absolute and guided by the bivalence principle. Once we restrict attention to such sentences, various potential objections based on contextualism or relativism are beside the point. Cf. Murawski and Wolenski (2008).

[17] In (1983: 199) Tarski distinguished the absolute notion of truth from the relative notion of true or correct sentence in an individual domain, which seems to be a direct predecessor of the now standard notion of truth in a structure, as defined in Tarski and Vaught (1956). From a certain point of view, the absolute notion can be viewed as a limiting case of the relative, where truth is defined for a language with intended interpretation in a given structure (but see Hodges 1985/86 for an interesting critique of the view that the notion of truth in a structure is already present in CTFL). CTFL was largely devoted to the notion of absolute truth, defined for fully interpreted languages, but Tarski made it clear that all the above mentioned notions (including the relative ones) belong squarely to his semantics, since, at bottom, they are definable in terms of satisfaction (or satisfaction in a structure). Cf. the footnotes n. 20 and 35 in his (1944).

notion of *provability in a formal deductive system*, the relevant notion of provability being purely formal-syntactic (or structural, as Tarski called it).[18] However, in the wake of Gödel's incompleteness theorems such proposals seemed doomed to failure. Tarski was one of the first to realize that truth (and related notions) cannot be reduced to proof (and related notions), that semantic notions need to be conceptually distinguished from syntactic notions in metalogic. The following passage, in which Tarski refers to Gödel's results, deserves to be quoted in full:

> […] Doubts continue to be expressed whether the notion of a true sentence – as distinct from that of a provable sentence – can have any significance for mathematical disciplines and play any part in a methodological discussion of mathematics. It seems to me, however, that just this notion of a true sentence constitutes a most valuable contribution to meta-mathematics by semantics. We already possess a sense of interesting meta-mathematical results gained with the help of the theory of truth. These results concern the mutual relations between the notion of truth and that of provability; establish new properties of the latter notion (which, as well known, is one of the basic notions of meta-mathematics); and throw some light on the fundamental problems of consistency and completeness […].[19]

This looks so familiar to us nowadays, when we are accustomed to distinguishing syntactic (proof-theoretic) and semantic (model-theoretic) approach to logic, that we may forget that there was no precise and systematic theory of semantic notions needed in metalogic when Tarski started his work on CTFL. Tarski's ambition was to show that metalogic, qua systematic theory of truth-theoretic and proof-theoretic properties of formalized deductive theories, can itself be construed as a deductive theory formalized within some respectable logico-mathematical framework, and based on precise definitions of semantic notions in respectable terms of the framework. It should thus enjoy the same respect as the framework in which it is expressed; in particular, it is consistent so long as the framework is consistent. What is of particular interest to us is that Tarski's mathematical and logical ambitions meet at this point his philosophical ambition. For, he not only hoped to persuade logicians and mathematicians that the semantic part of metalogic is itself a mathematically precise and interesting discipline that can be expressed within some respectable logico-mathematical framework and is thus consistent so long as the framework is consistent;[20] he also hoped to show that one can put to rest the positivist worries concerning the scientific (i.e. physicalistic) respectability of semantics. The idea is that semantic notions are as respectable from the scientific point of view as are the notions in terms of which they are defined.[21]

The following picture emerges from my exposition so far. A theory of truth forms the basis of semantics as conceived by Tarski, its heart being the definition that agrees reasonably well with the intuitive notion of truth used in metalogic and in science

[18] Carnap (1949) reports such a tendency in reaction to semantic paradoxes.
[19] Tarski (1944: 368). Tarski refers back to p. 354, where he mentions Gödel's results.
[20] Cf. Vaught (1974) and Feferman (2008). For more on this aspect of Tarski's work see section 10.
[21] Tarski (1983: 406).

generally. No notion defined in terms of confirmation or proof will intuitively serve the needs of logic and science: whether we consider truth in logic, mathematics or any other branch of science, truth does not seem to coincide with proof, confirmation, and the like ideas. That is not to say, of course, that there is no connection between truth and epistemic notions, but the relation between them is that epistemic ideas seem to presuppose the notion of truth, because we devise our proof procedures and epistemic procedures in general to track truth, the grasp of which notion guides us in our efforts.[22] At this point we should recall the platitude:

(P1) *S is true iff things are as S says they are.*

(P1) may not be the most precise statement, but it is hard to deny that the sound intuition behind it is that truth of a sentence depends both on what it says (its content or meaning) and on the way things are in reality. We shall see shortly that Tarski converted this intuition into the adequacy criterion of a definition of truth. However, he was well aware that this very adequacy criterion gives rise to infamous truth-theoretic paradoxes under certain conditions, and that there were certain worries about its allegedly metaphysical character. In view of all this, Tarski set out to provide a theory of truth based on its precise definition that meets the following desiderata:

a) it conforms to the adequacy criterion which in a way captures the above mentioned intuition;
b) it is consistent, that is, does not give rise to paradoxes;
c) it entails basic principles which intuitively hold of truth (e.g. the law of excluded middle and the law of non-contradiction);
d) it is meta-logically fruitful in that basic semantic notions of consequence, validity, etc., can be defined in terms of it, and basic metalogical results can be proved on its basis;
e) it is framed in terms that are scientifically (mathematically) respectable.

He also hoped to teach philosophers something important about truth-theoretic and semantic paradoxes in general: when and why they arise, what it takes to avoid them, and what consequences this has for the classical problem of truth definition – under what conditions it is possible to define truth (and semantic notions in general) both consistently and adequately, and under what conditions this is not possible.

2. Sentences and language-relativity of truth definitions

It was said that Tarski's method defines the notion of truth as predicated of sentences of a given language. In CTFL Tarski did not explain this choice, but in his later, more popular papers he made some remarks to the effect that 'true' is commonly predicated of sentences, and that this is its original use in natural languages.[23] One would like to

[22] The popular paper (1993), named 'Truth and Proof', makes this ambition of Tarski clear.
[23] Tarski (1993: 101).

know on what evidence he based this claim, as there are reasons to disagree with it.[24] But he did not tell us, although he was aware that the claim could be disputed by those thinkers who argue that truth should be defined for beliefs or judgements or propositions, in so far as such items are taken as primarily truth-evaluable. Tarski did not want to exclude that definitions of truth can be provided for such items but declared that he, at any rate, will be concerned with the *logical* notion of truth,[25] because for purposes of logic one needs be concerned only with truth as a property of declarative sentences. At the same time, though, he pointed out that it is convenient to define the notion of truth for sentences, since sentences are relatively unproblematic entities (compared to propositions, say).[26] We can appreciate this strategy nowadays, as there is still no agreed upon a conception of propositions.

What is rather important is that Tarski thought that once the truth definition is restricted to sentences, it has to be relativized to a particular language, since a sentence *S* that is true in language A may happen to be false in language B, depending on what it means in A and B respectively, and it may be neither true nor false in language C, in which it means nothing at all. So, the question whether *S* is true (or false) may have no definite answer, unless a particular interpretation of *S* is picked out. Such indefiniteness is to be removed by connecting sentences always with a particular language: a sentence has a definite meaning, hence definite truth conditions, only in the context of a particular language.[27]

3. Formal correctness and material adequacy

Tarski imposed two important conditions on a satisfactory definition of truth for a language: formal correctness and material adequacy. Now, *material adequacy* is nothing but the familiar desideratum (a), and formal correctness is the combination of the desiderata (b) and (e). As for *formal correctness*, he required that all the terms be listed that will be used in the definition, and that formal rules be specified in conformity to which the definition will be constructed (this, we shall see, requires the vocabulary and structure of the language to be exactly specified in which the definition will be given). The definition has to be a sentence that has the following standard form of (universally quantified) equivalence

(For every sentence *x* of L), *x* is true iff …*x*…,

where '*x*' is the only variable occurring in '…*x*…' (in the definiens), and what fills in the dots is an expression that contains neither 'true' nor any other term whose definition presupposes it. In a word: the definition must not be circular. Apart from that, Tarski required that the terms used in the *definiens* must not admit of any doubt or cause any methodological problems. In particular, no notion belonging to the province of semantics can occur in the *definiens*, unless it can be defined in non-semantic terms of

[24] The *locus classicus* is Carwtright (1962). See also Soames (1999).
[25] Tarski (1993: 101).
[26] See Tarski (1944: 342) and (1993: 101).
[27] Of course, the matter is more complicated on account of ambiguous, context-sensitive or vague sentences. See section 4.

the language (or theory) in which the definition is framed. Tarski thought this desirable, since, as we now know, many thinkers worried that semantic notions are paradoxical and/or that they cannot be explained in scientifically respectable terms (on account of the mysteriously metaphysical relation to language-independent reality). This definitional procedure ensures the consistency of the definition provided that the language (or theory) in which the definition is framed is consistent.

The motivation for the condition of material adequacy is not difficult to grasp. Clearly, formally correct definitions of the term 'true sentence' can be given that are intuitively inadequate as definitions of the notion of *true sentence* for a given language, e.g.:

(For every sentence *x* of English), *x* is a true sentence iff *x* has three words

This definition is no doubt formally correct, but it does not get things right: it is neither a sufficient nor a necessary condition for truth of a sentence of English that it has three words. The fact that Tarski wanted a criterion of adequacy makes it clear that he thought that there is something for a truth definition to be adequate or inadequate to, that it can get things right or wrong. He did not just want to introduce a new term via a purely stipulative definition, since there is nothing for a stipulative definition to get right or wrong – it simply states how a new term is to be used. Tarski was clear on this:

> […] The desired definition does not aim to specify the meaning of a familiar word used to denote a novel notion; on the contrary, it aims to catch hold of the actual meaning of an old notion […].[28]

Given that the definition aims to catch hold of the actual meaning of an old notion of truth, one would expect that Tarski took into account the way 'true' is actually used, because usage determines or at least constraints what a notion or word expresses. And, indeed, he repeatedly claimed that he intended to give truth definitions that agree reasonably well with what he variously called the 'common', 'intuitive' or 'ordinary' usage or meaning of 'true'. At the same time, though, he admitted that the actual usage – both common and philosophical – of the word 'true' is to some extent ambiguous and imprecise and some choice has to be made if one wants to give a precise definition of the notion of truth. We have seen that epistemic definitions of truth were found wanting by Tarski on the ground that they are hard to square with the intuition that the principle of exclude middle holds of truth. Other conceptions he explicitly mentioned were pragmatic (or utilitarian), which equate truth with kind of theoretic and/or practical utility of a belief, and usually also honour epistemic virtues. However, pragmatic theories of truth do not seem to fare any better than epistemic theories, since it is counterintuitive to equate truth with utility: there seem to be true beliefs that are not useful in any reasonable sense of that word, and false beliefs that are useful is in at least some reasonable sense.[29]

---

[28] Tarski (1944: 341).

[29] Furthermore, both epistemic and pragmatic properties appear to be context, time or subject relative in a way that truth does not appear to be, e.g.: the sentence 'The earth is not flat' was true even in those times when all the evidence available justified rather the attitude of holding true its opposite.

At any rate, Tarski complained that both epistemic (coherence) and pragmatist theories of truth 'have little connection with the actual usage of the term 'true'' and that 'none of them has been formulated so far with any degree of clarity and precision.'[30] He made it clear that his truth definitions are intended to make explicit and precise only the sound intuition about truth expressed in our familiar platitude

(P1) *S* is true iff things are as *S* says they are

or in the following statement (called the 'semantical definition' of truth)

(P2) […]A true sentence is one which says that the state of affairs is so and so, and the state of affairs is indeed so and so […].[31]

Tarski claimed that a conception or definition of truth that makes precise the intention behind them will agree 'to a very considerable extent with the common-sense usage.'[32] Such a conception of truth could rightly be called *classical*[33], since it would make clear the same intuition that Aristotle aimed to capture in his famous definition of true (false) statement in the *Metaphysics*:

(P3) […] to say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true […].[34]

The intuitions expressed in such statements are called by philosophers *correspondence platitudes*. Tarski himself said that according to the classical conception, truth of a sentence consists in its 'corresponding with reality'.[35] It has to be kept in mind, though, that one may embrace such platitudes involving the notion of truth, without endorsing a full-blooded theory that explains truth in terms of a dyadic relation (of a sort) between truth-bearers (of a sort) and truth-makers of a sort. It is, in particular, doubtful whether the platitudes (Pl), (P2) and (P3) support any full-blooded correspondential reading, and whether Tarski's conception and definition of truth that is based on such platitudes can be considered a correspondence theory.[36] I shall not address in detail this issue here. Let me just note that Tarski himself seemed to be uncertain in this respect:

> […] As far as my own opinion is concerned, I do not have any doubts that our formulation does conform to the intuitive content of that of Aristotle. I am less certain regarding the later formulations of the classical conception, for they are very vague indeed […].[37]

---

30 Tarski (1993: 103).
31 Tarski (1983: 155).
32 Tarski (1944: 360). See also Tarski (1993: 102).
33 Tarski (1983: 153), (1944: 343), (1993: 103).
34 *Metaphysics*, Γ 7, 1011 b. Aristotle (1923).
35 Tarski (1983: 404).
36 This applies both to modern theories based on the notion of correspondence to facts, as well as to more traditional theories based on the notion of correspondence to objects (derived from Aristotle). See Künne (2003) for a good discussion of these matters.
37 Tarski (1944: 360).

He rightly worried that, qua attempts at defining the notion of truth, such slogans are not satisfactory. The definition of *truth* as *correspondence with facts* (*agreement with reality*) is only as good as our understanding of the notions used in it, in particular, of *correspondence* and *fact*. If these are not clearer than the notion of truth itself (which they do not seem to be), and are not further explained in clearer terms, we have no explanation of truth but merely the illusion of one.[38] In general, Tarski preferred Aristotle's definition, because it is not couched in the imprecise idioms of *correspondence*, *fact* or *reality*. Still, he did not deem it sufficiently precise and general,[39] and claimed that he wanted to obtain

> […] a more precise explanation of the classical conception of truth, which could supersede the Aristotelian formulation preserving its basic intentions […].[40]

Now, it was arguably not the reference to states of affairs in (P2) that Tarski took to be the sound intuition behind the platitudes, but rather this: if a sentence *S* says that *p*, *S* is true if *p*, and untrue if not *p*; in short: *S* is true iff *p*. For instance, if we consider the sentence 'Snow is white' and ask under what conditions it is true, what according to Tarski comes immediately to our mind is that:

(1) 'Snow is white' is true if, and only if, snow is white

(1) seems to make an obviously true claim, because the sentence used on its right side expresses the content of the sentence mentioned on its left side, the same sentence being mentioned on the left side and used on the right side. That our intuitions about the validity of (1) are not misguided can be demonstrated as follows. (1) is a material biconditional and as such it is false only if its two sides differ in truth-value, that is, if either: (a) 'Snow is white' is false and "'Snow is white' is true" is true; or (b) 'Snow is white' is true and "'Snow is white' is true" is false. But if 'Snow is white' is false, then "'Snow is white' is true" is false and not true ((a) is excluded); and if 'Snow is white' is true, then "'Snow is white' is true" is true and not false ((b) is excluded). So, the two sentences under consideration are materially equivalent and (1) holds.[41]

Nothing in principle changes when we consider the biconditional for the German sentence 'Der Schnee ist weiss':

---

[38] Granted, some philosophers – Russell and Wittgenstein come immediately to mind here – attempted to explicate the correspondence intuition. But then their accounts faced formidable difficulties.

[39] Now, generality it indeed lacks, but not for the reason Tarski mentioned. He thought that it lacks generality because it covers only non-elliptical existential statements of the form '…is (not)' (Tarski 1993: 102). For a persuasive exegetical discussion of how to square Aristotle's dictum with his intentions, namely as covering categorical subject-predicate affirmations and negations, see Künne (2003). Improving on the intuition behind Aristotle's definition in light of (P1) (or something close to it) was a common practice in the Polish philosophico-logical tradition, whose members were quite agreed that the slogans in terms of correspondence or agreement with reality (facts) are problematic. For useful information about the relations of Tarski to this tradition see Murawski and Wolenski (2008).

[40] Tarski (1993: 103).

[41] If we allow sentences that are neither true nor false, the situation changes. For if *S* is a sentence that is neither true nor false, it may be argued that '*S* is true' is false, and the two sentences fail to be materially equivalent.

(2) 'Der Schnee ist weiss' is true iff snow is white,

in which the sentence used on the right side expresses the content of the sentence mentioned on the left side, being its translation. For, presumably, if *S′* translates *S*, *S′* and *S* mean the same, and hence they do not differ in their truth-value.[42] So, if 'Snow is white' is false, then 'Der Schnee ist weiss' is false, and " 'Der Schnee ist weiss' is true" is thus also false; and if 'Snow is white' is true, then 'Der Schnee ist weiss' is true, and " 'Der Schnee ist weiss' is true" is thus also true. Hence, the two sides of (2) do not differ in truth-value and (2) holds.

Tarski's simple but fundamental insight was that biconditionals like (1) or (2), which refer to a particular sentence of a particular language,

> […] explain in a precise way, in accordance with linguistic usage, the meaning of phrases of the form '*x* is a true sentence', which occur in them […].[43]

A particular biconditional of this kind explains wherein the truth of a particular sentence of a particular language consists, by way of specifying the condition under which the truth-predicate applies to it – its condition of truth. Such biconditionals can then be viewed as partial definitions of the notion of truth with respect to particular sentences of a given language L. This is generalized in the semantic conception of truth (henceforth, SCT), apparently so-called in allusion to the semantical definition (*viz.* P2), the general intention of which it aims to make more precise and definite in the following way:

(a) Every biconditional obtained from the schema *X is true iff p* by putting for 'X' the designator of a sentence of L and for 'p' either that sentence itself or its translation, holds and fixes the notion of truth for that particular sentence.[44]

(b) Taken together, such biconditionals for all sentences of L fix the notion of truth for L.

We shall follow the custom of using the label 'T-schema' for *X is true iff p* (or for its non-English versions) and 'T-biconditional' for any instance of T-schema obtained in accordance with the indicated rules of substitution for its dummy letters.

With SCT in place, Tarski was able to give a sharp formulation of the condition of material adequacy. If particular T-biconditionals for sentences of L fix the notion of sentential truth for particular sentences of L, then:

---

[42] Context-sensitive sentences form an important class of exceptions. See section 4.

[43] Tarski (1983: 187).

[44] More exactly, Tarski claimed that this holds only for those sentences that do not contain 'true' as their significant part. The reason is that the biconditional for such a sentence would contain on its right side 'true' and could not thus serve as a partial definition of 'true', because definitions are required to be non-circular. But, as he says, even though the biconditional is not a partial definition, it 'is a meaningful sentence, and it is actually a true sentence from the point of view of the classical conception of truth' (Tarski, 1993: 105). Furthermore, designators replacing 'X' should be perspicuous in that it is always possible to reconstruct from them sentences that they designate. As Tarski says: 'Given an individual name of a sentence, we can construct an explanation of type (2)' (namely: '*x* is true iff *p*'; my insertion), 'provided only that we are able to write down the sentence denoted by this name.' (Tarski, 1983: 156). Thus, quotational names, syntactic descriptions or Gödel numbers of sentences are all perspicuous designators in this sense.

> […] Not much more in principle is to be demanded of a general definition of phrases of the form '*x* is a true sentence' than that it should satisfy the usual conditions of methodological correctness and include all partial definitions of this type (as special cases; that it should be, so to speak, their logical product […].[45]

Tarski thus proposed that a definition of the truth-predicate for L will be called or considered materially adequate (to SCT), if it entails all the T-biconditionals for L.[46]

Some clarificatory comments are in order. [A] The material adequacy criterion is supposed to capture the intention of one conception of truth only, namely SCT. Although Tarski did not mean to exclude other conceptions of truth which may suggest different criteria of material adequacy for definitions faithful to them, he nevertheless thought that SCT is a neutral ground in that it does not commit one to take any stand on issues traditionally debated by philosophers with respect to truth (e.g. idealists vs. realists, or empiricists vs. metaphysicians), which he construed as being about what, if anything, warrants or entitles assertion of a sentence of this or that kind.[47] However, he deemed it strange to uphold a conception of truth that is incompatible with SCT, since it presumably implies the denial of some T-biconditional, which denial amounts to an assertion of a sentence of the form " 'p' is true iff not p". And this is surely not a particularly attractive position. [B] A truth definition for L that meets the material adequacy criterion is assured to be extensionally adequate in that all and only the true sentences of L fall under it. But this does not mean that it also supplies a criterion of truth for L, that is, a method or test effectively deciding whether a given sentence of L is true or not. Clearly if we are unable to decide the truth-value of a sentence, say, 'An extra-terrestrial life exists', it does not help us very much to be said that the sentence 'An extra-terrestrial life exists' is true iff an extra-terrestrial life exists. As Tarski noted, we cannot in general expect that a definition of a notion will provide an effective method of deciding what falls under it.[48] [C] Given Tarski's claim that his truth definition 'aims to catch hold of the actual meaning of an old notion' and repeated claims to the effect that T-biconditionals are partial definitions of 'true' for sentences of L, that together *explain the meaning* of 'true' with respect to the whole of L, it is likely that he thought that a materially adequate truth definition for L will be more than extensionally adequate. Unfortunately, Tarski used the notion of meaning informally and rather loosely and it is hard to figure out whether he really meant his truth definitions to capture the meaning as opposed to the extension of 'true' (as applied to L).[49]

---

45 Tarski (1983: 157).

46 Tarski (1993: 106).

47 Tarski (1944: 361).

48 See Tarski (1944: 364) and (1993: 116). The proponents of epistemic theories of truth often complained that correspondence conceptions of truth are useless as a method of discovering or recognizing or checking what is true. This seems right, but as Tarski pointed out, to object against a conception or definition of truth on the ground that it does not enable us to recognize what falls under it – to tell truth from falsehood – is a special case of asking from a definition that it supply the criterion or test enabling one to effectively decide what falls under the notion defined. Tarski rightly retorted that this is too demanding a standard of definition, which is frequently violated in scientific practice.

49 See Patterson (2008) for a detailed reconstruction of Tarski's views on meaning. Field (1972) claims and Hodges (2008) argues that Tarski intended his material adequacy criterion to amount to no more than the extensional adequacy of truth definition, while Künne (2003) and Patterson (2008) claim that Tarski wanted, in some sense, to capture the meaning of 'true'.

At any rate, in the last section of this paper we shall have an occasion to see that if it was Tarski's goal to explain the meaning of 'true' (as applied to a particular language), then there are rather good reasons to think that he was not successful in achieving it.

It is sometimes said in the literature that SCT – especially the claim that T-biconditionals are partial definitions of the notion of truth – is but a sentential version of the redundancy theory of truth. The redundancy theory was originally propounded for sentences such as 'It is true that snow is white', in which 'true' is a part of the sentence-forming operator 'it is true that' that behaves semantically as the double-negation operator (turning truths into truths and falsehoods into falsehoods). The starting point is nicely captured by Frege's observation that instances of the equivalence schema

It is true that p iff p

are platitudes that display the transparency property of the notion of truth: we can see through 'it is true that', as applied to a given sentence S, to the content of S itself.[50] Frege then made the following point, which Ramsey, taken by many as the father of the redundancy theory, approved:

> […] If I assert 'It is true that sea-water is salt'. I assert the same thing as if I assert that 'sea-water is salt' … the word 'true' has a sense which contributes nothing to the sense of the whole sentence within which it occurs as a predicate […].[51]

> […] 'it is true that Caesar was murdered' means no more than that Caesar was murdered, and 'it is false that Caesar was murdered' means that Caesar was not murdered […].[52]

Since the operator 'it is true that' does not seem to add anything to the content of what it is predicated of, some people – though not Frege – concluded that 'true' is a redundant word that can always be erased without loss, and at best performs specific pragmatic functions in language (e.g. putting an emphasis on a claim, commending or endorsing it). This claim was sometimes wedded to another: being a content-redundant device, 'true' expresses no property at all, hence no property calling for a deep philosophical analysis.[53]

Tarski himself criticised one form of the redundancy theory under the name 'the nihilistic approach to the theory of truth',[54] according to which 'true' makes sense only when it occurs (syncategorematically) in the contexts 'It is (not) true that p', and all

---

[50] See Frege (1918/1919). But Frege was not a redundancy theorist. He held a peculiar view, on which truth is an important, indefinable notion, characterized by the transparency property, which shows that it is not best construed as a property (rather, it is to be understood as that at which we aim in making judgments and assertions or forming beliefs).

[51] Frege (1978: 251).

[52] Ramsey (1999: 106).

[53] Ramsey (1999, 2001) is usually considered the father of this approach, although he did not make the claim that 'true' does not express a property. This claim was made famous only later by Ayer (1951: 87–90).

[54] Tarski (1993: 111). In (1944: 358–359) he used the label 'redundancy of semantic terms – their possible elimination'.

other (categorematic) uses of it are to be treated as illegitimate (in part because they give rise to paradoxes). The nihilists are right to say that the word 'true', as it occurs syncategorematically, seems to be a sort of redundant device. If 'true' were used only in such contexts, one could well wonder if it performs other than purely stylistic or ornamental functions in the discourse. Tarski's interesting and correct response to the nihilists was that they propose to eliminate from the discourse precisely those uses of 'true' that do not seem to be purely stylistic or ornamental. As he pointed out, the notion of truth serves a very useful expressive function. Thus, for instance, it is often attached to a description of a sentence that one is not in a position to reproduce *verbatim* but has reasons to accept/assert or reject/deny (the so-called 'blind ascriptions of (un)truth'):

Though I cannot quite recall it, the first sentence in the *Tractatus* is true

Moreover, it enables us to express generalizations such as the following:

All consequences of true sentences are true

Everything that the pope said 24. 12. 2008 in Rome was true

or

Every alternation of a sentence and its negation is true

The nihilistic approach to truth, and the redundancy theory of truth in general, is hard pressed to explain such uses of truth. While SCT gives pride of place to the intuition that a sentence of the form "'p' is true" is sort of equivalent to 'p' under the usual notion of truth, Tarski saw clearly that 'true' cannot be explained away (eliminated) from such contexts in any automatic way, which strongly suggest that truth is a property of a sort after all, and a useful one (in part due to its expressive uses mentioned above).[55] Ramsey was also aware of the fact that in the ordinary language 'true' is not easily eliminable from such contexts.[56] What he proposed by way of a solution was the following definition (expressed in a semi-formal English):

(For all x): x is true iff $\exists p$ (x = the proposition that p and p).

The success of this strategy depends on whether there is a working account of quantification that will make the definition intelligible, because under the usual objectual reading of quantifiers it does not make good sense. Before presenting his own method of truth definition, Tarski considered a closely related definition of sentential truth:

[55] Here Tarski was closer to modern deflationist theorists who maintain that the point of 'true' in the discourse is that is serves a certain logical or expressive function, enabling us to express such generalizations (infinite conjunctions or disjunctions), and that it is capable of performing this function because its use is governed by something like the schematic principle expressed by T-schema. See Quine (1970), Leeds (1978), Horwich (1982, 1990) or Field (1994).

[56] Ramsey (1999, 2001). But Ramsey did not make the point that modern deflationists are so proud of: that the point of the notion of truth in language consists precisely in such uses.

(For all x): x is true iff $\exists p\,(x =$ 'p' and $p)$

At first blush, this definition seems to be a straightforward generalization of SCT that subsumes T-biconditionals as special cases, but Tarski rejected it on the ground that it either does not make sense (under the view that "'p'" is simply an unstructured name of one particular letter, which does not make sense to bind by a quantifier) or else quotation marks are to be treated as a name-forming functor that assigns an expression, qua its argument, its quotational name. Such a functor, however, is not extensional, and Tarski complained that the nature of intensional functors is an unclear matter.[57]

Tarski also discussed an objection against SCT to the effect that what T-biconditionals show is that 'true' can always be eliminated when applied to a quotational name of a sentence. And what is eliminable is in a way redundant or sterile (as he put it). As we shall see, Tarskian truth definitions have the potential of eliminating truth-predicates from all contexts of the language (or theory) in which they are defined. However, Tarski did not regard this as a sign of the fact that SCT is a redundancy theory of truth, since, by parity of reasoning, one would have to conclude that all defined terms (in science) are useless or sterile – which he deemed absurd.

4. Exactly specified languages

Once the condition of material adequacy was sharply formulated, the problem of defining truth took the form of the question: For what languages and by what methods is it possible to construct formally correct definitions of truth that entail the T-biconditionals for all their sentences? In the sections to come we shall see that although it is easy to construct partial truth definitions for particular sentences in the form of their T-biconditionals, it is not a trivial task at all to construct a materially adequate truth definition for a reasonably rich language that contains an indefinite number of sentences. But let us first tackle the question as to for what languages it is possible to give such truth definitions.

Tarski famously argued that it is not possible to provide satisfactory truth definitions for natural languages. One reason he had for this negative claim was that natural languages are too loose and ill-behaved phenomena. He maintained that

> […] The problem of the definition of truth obtains a precise meaning and can be solved in a rigorous way only for those languages whose structure has been exactly specified […].[58]

But in the same breath he claimed that

> […] Our everyday language is certainly not one with an exactly specified structure. We do not now precisely, which expressions are sentences, and we know even to a smaller degree which sentences are to be taken as assertible […].[59]

[57] It is fair to say, though, that many thinkers, especially from the camp of deflationists, nowadays believe that there is a coherent substitutional reading of that definition (see Soames, 1999). The question then is whether such a reading does not presuppose the very notion of truth (Cf. Horwich, 1990, or Field, 1994).

[58] Tarski (1944: 349).

[59] Ibid.

Tarski's point is that a truth definition for a language L is materially adequate only if it entails T-biconditionals for all sentences of L. However, if it is not fixed what belongs to the set of sentences of L, it is not fixed either what belongs to the set of T-biconditionals for sentences of L, hence the problem of constructing a materially adequate truth definition for L becomes moot. And Tarski thought that it is not settled what words belong to a natural language such as English, and that it is therefore not settled what truth-evaluable sentences belong to English, since sentences are formed from words. Moreover, he seemed to believe that the set of sentences, *a fortiori* of truth-evaluable (or assertible) sentences of a natural language, cannot be specified in syntactical terms.

One may well doubt if the foregoing considerations really disqualify natural languages from being given satisfactory truth definitions: for purposes of theorizing, we can conceive of a natural language as having fixed vocabulary (at a given time, say), and linguists nowadays approach natural languages by formal methods, trying to determine the category of meaningful or grammatical sentences. But Tarski had further reasons to despair of natural languages in this respect. We have already seen that a sentence is to be referred to a particular language if the question as to when (or under what conditions) it is true or false is to have a definite sense. But, of course, the truth about sentential truth is more complicated. If we take sentences to be capable of being true or false, we see at once that many of them can be evaluated as true or false only if we take into account some particular context of their utterance, and that their truth-value might change across such contexts of utterance, depending on various features of contexts such as who speaks, where, when, to what addressee and with what intentions. Tarski was well aware of this, because he said that many sentences do not satisfy the condition that 'the meaning of an expression should depend exclusively on its form' and, in particular, that 'it should never happen that a sentence can be asserted in one context while a sentence of the same form can be denied in another'[60], obviously meaning that their truth-value can change across contexts of their utterance.

Now, owing to the fact that the content of a context-sensitive sentence – e.g. 'I am hungry' – varies across contexts of its utterance (in this case, depending on who utters the sentence and when), it cannot be plausibly assigned the truth conditions in the form of the biconditional

(3) 'I am hungry' is true iff I am hungry,

because by using (3) one at best captures the truth-conditions of 'I am hungry' as uttered by himself at that time, but there is no telling what the truth conditions of (3) are as uttered by different speakers at different times, or as uttered by himself at other times. To capture this, we need a generalization that makes explicit the dependence of the truth conditions of (3) on who utters it and when

(4) For all *s* and *t*, 'I am hungry' is true as uttered by *s* at *t* iff *s* is hungry at *t*

[60] Tarski (1993: 113).

This has some intuitive appeal, but we no longer have on the right side the sentence mentioned on the left side (or its translation), and hence we can no longer use the criterion of material adequacy as Tarski stated it. Tarski's strategy thus works only for eternal sentences – sentence-types all of whose tokens have the same truth conditions. In order to get a new criterion of adequacy that would apply to non-eternal sentences, we need to generalize T-schema.[61]

Other troublemakers in natural languages are ambiguous sentences that often occur within a single language, whose truth conditions might vary under their different readings or interpretations. So far as I know, Tarski did not explicitly speak of vague predicates or non-denoting terms in this respect, but it is likely that he would have considered sentences containing such expressions as equally problematic, provided that we treat them as not having a determinate truth-value and as being neither true nor false respectively. For, in general, Tarski seemed to think that a precise truth definition in his style can be given only for sentences that have a sufficiently determinate sense (content) to be either true or false.[62]

5. Truth and paradox in natural language

Tarski had a second, more serious reason for claiming that formally correct and materially adequate truth definitions cannot be given for a natural language L, even if we could somehow approach L as having an exactly specified structure (and ignored context sensitive and other troubling sentences). According to him, a natural language like English is universal in that it can in principle express everything that can be expressed. In particular, then, of any English sentence we can say in English that it is true (or untrue), by appending 'is true' (or 'is untrue') to an English designator of it. Once we realize this, Tarski said, we should realize the possibility of forming a self-referential sentence of English that can (be used to) say something about itself, in particular, we can construct a sentence of English that says of itself that it is untrue (or is equivalent to a sentence that says that). But such a sentence starts a logically plausible chain of reasoning that ends with a plain contradiction. To illustrate this, we shall use the letter 's' to denote the following sentence, which looks unexceptionably English:

The only bold printed italicized sentence in this paper is not true.

---

[61] The following is a tentative attempt in this direction: A definition of truth for a language L is materially adequate if it entails all instances of the schema *X is true in C iff p*, where 'X' is replaced by a perspicuous designator of a sentence *x* of L, and 'p' by a sentence that says the same (or expresses the same content) as *x* would have said (expressed) if uttered in C.
Indeed, context sensitivity has been intensely studied in essentially this spirit in modern truth conditional approaches to semantics heavily influenced by Tarski's own ideas – Davidson (1984), Field (1972), Montague (1974), Lewis (1983), Kaplan (1989). One may say, though, that Tarski would not have endorsed this proposal on the ground that it uses more involved ideas of 'saying the same' or 'expressing the same content'. To this it may be retorted that, in general case, he himself needed the idea of correct translation that presupposes the idea of sameness of meaning, which does not seem to be less problematic than the other two ideas.

[62] It may well be Tarski's strictures here are connected to his absolutist view of truth. See Murawski ad Wolenski (2008) for a short discussion of this idea in relation to the Polish school and Tarski.

According to the material adequacy criterion stated in Convention T, we are licensed to assert the following T-biconditional:

s is true iff the only bold printed italicized sentence in this paper is not true.

But we easily confirm that the following holds:

The only bold printed italicized sentence in this paper is identical with *s*.

Given this identity, we can replace the name 'the only italicized sentence on this page' by the name '*s*' in the T-biconditional, without affecting its truth-value (the principle of substitutivity of identicals). What we get is this:

s is true iff s is not true

And from this sentence the contradiction "'*s*' is true and *s* is not true" follows within classic logic.

What the paradox teaches us is that the T-biconditionals for a language entail a contradiction (within classical logic) if a certain kind of self-reference is present in the language. In particular, English has means of designating any expression or sentence that belongs to it, as well as the truth predicate 'true' ('untrue') that can be meaningfully appended to any such designator, the result being itself an English sentence. Tarski called such a language *semantically closed*. We then arrived at the contradiction assuming only that: (A) classical bivalent logic holds; (B) T-biconditionals are valid (part of the motivation behind the criterion of material adequacy).[63] Now, unwilling to abandon (A) or (B), Tarski famously put blame on the factor of semantic closure and argued that no consistent language for which classical bivalent logic holds (call such a language classical) expresses its own truth-definition (or truth theory) that entails all T-the biconditionals for its sentences. By contraposition, then: no classical language that expresses its own truth definition (truth theory) that entails all the T-biconditionals for its sentences is consistent. Consequently, a formally correct and materially adequate definition of truth can be constructed only for a semantically open language L that does not contain its own truth-predicate, and the definition must therefore be framed in another, expressively richer language L (called 'metalanguage'), that has resources needed to define truth for L. In general, if truth for a classical L is definable (or usable) in ML in a consistent and materially adequate manner, a principled distinction must exist between the object-language L and the meta-language ML: the two must not be identical or inter-translatable (though L may be a proper part of ML). If we wish to define

[63] Actually, an additional factor was that in English an empirically established premise 'The only bold printed italicized sentence in this paper is identical with *s*' can be formulated and accepted as true. But Tarski (1983: 168) remarked that indirect self-reference by means of empirical descriptions is not necessary, because we can reconstruct a version of Grelling's paradox of *heterological predicate* in it, which does not rely on any empirical premise: let '*F*' abbreviate the predicate 'not true of itself'. Then we can apply '*F*' to itself (self-apply '*F*'), getting the sentence: "'*F*' is not true of itself". It can be shown the the sentence is false if true and true if false.

truth for the stronger ML, we have to ascend to a yet more powerful language, since ML cannot define its own truth predicate (on pain of inconsistency). According to Tarski's hierarchical strategy, we can then imagine a whole (possibly transfinite) sequence of increasingly richer languages L_0, L_1, L_2, …, where L_0 does not contain any truth-predicate, and for any n ($0 < n$), L_n contains the truth predicate 'true$_n$' that applies only to lower level sentences of L_m ($m < n$), but never to sentences of the same or higher level.[64] The paradox is thus avoided, because the appropriate version of T-schema that contains an indexed truth predicate

X is true$_n$ iff p

generates meaningful (or well-formed) sentences only for sentences of a lower level than n (as substitutes for 'X'), and hence no troubling biconditionals can be asserted.

The solution of the truth-paradox along these lines is fine for properly restricted and formalized languages that Tarski decided to focus on (*see* the next section), but it does not apply to natural languages, which are universal (expressively unrestricted), and because of that no principled distinction between object-language and meta-language can be made for them, since any candidate metalanguage is translatable into them. Tarski admitted that it is possible to consider fragments of a natural language L that are semantically open and provided with exactly specified structure (of a certain sort) and complete vocabularies, and then construct satisfactory truth-definitions for them on the model of formalized languages, which will be framed in richer fragments of L (possibly of another natural language). Following this suggestion we can again imagine fragments E_0 and E_1 of English, where E_0 contains only English sentences that do not contain 'true' as their significant part, and E_1 contains E_0 as well as every sentence formed by appending 'is (not) true' to the name of any sentence of E_0. If we pursue this strategy, we might arrive at the whole hierarchy E_0, E_1, E_2,… of exactly specified and semantically open fragments of English such that for every n ($0 < n$), E_n contains the predicate true of all and only the sentences of E_{n-1}, but never of sentences of the same or higher level. Since the restricted truth-predicates in the hierarchy differ in their extensions, we had better to distinguish them as before by different indexes: E_1 contains 'true$_1$', E_2 contains 'true$_2$', and so on. This is a perfectly rational procedure, but Tarski thought that what one is imagining are not so much fragments of a natural language but well-behaved products of an artificial reform of a natural language:

> […] It may be doubted, however, whether language of everyday life, after being 'rationalized' in this way, would still preserve its naturalness and whether it would not rather take on the characteristic features of the formalized languages […].[65]

[64] It should be noted, though, that Tarski's views concerning the paradox and impossibility of defining truth for a language in the language itself, dominant as they once were, are no longer universally accepted. What Tarski showed is not that truth for a semantically closed L cannot be consistently defined, period; he showed that truth for such L cannot be defined if we assume (A) and (B). So one may ask whether truth can be consistently defined for a natural or other semantically closed language L once (A) or (B) or both are abandoned.

[65] Tarski (1983: 253).

6. Formalized languages

In view of all this, Tarski proposed to focus on a comprehensive group of semantically open and well-behaved languages used by logicians to formalize deductive-mathematical theories (originally expressed in informal or semiformal fragments of natural languages). They are formalized in that in the specification of their structure only properties and relations of their sings are appealed to.[66]

For truth definitions given relative to formalized languages, Tarski was able to give the sharpest formulation of the material adequacy criterion in the form of Convention T, which reads as follows:

> A formally correct definition of the term 'true' in the metalanguage (ML) will be called a materially adequate definition of truth for a given language (L) if it has among its consequences all sentences obtained from T-schema
>
> *X is true iff p*
>
> by replacing the schematic letter 'X' by a structural-descriptive designator of any sentence of L and 'p' by its translation into ML.[67]

Some comments are in order. [A] Although Convention T may appear to be general, as it stands it applies only when ML is English (or better, a well-behaved fragment of it), since instances of *X is true iff p* are English sentences. For a truth definition framed in another meta-language, we have to reformulate the convention and use an appropriate version of T-schema (e.g., if ML is a fragment of German, we have to use *X ist wahr wnw p*).[68] Convention T itself is formulated in a meta-meta-language, which may or may not contain ML. [B] If a definition meets Convention T, it is extensionally adequate – all and only the true sentences of L fall under the defined notion. For, any two definitions that satisfy Convention T are bound to be equivalent. [C] Though the 'if' in Convention T suggests that it specifies only a sufficient condition of material adequacy, it was arguably intended to state both a necessary and sufficient condition of material adequacy – *viz.* the qualifications 'it is to be demanded' and 'it should be' in the passage:

> […] Not much more in principle is to be demanded of a general definition of phrases of the form '*x* is a true sentence' than that it …include all partial definitions of this type (as special cases; that it should be, so to speak, their logical product […].[69]

---

[66] But they are fully interpreted – their expression are equipped with determinate meanings.

[67] For the original formulation differing in certain aspects, see Tarski (1983: 187–188). A structural-descriptive name is a perspicuous designator of the following sort: 'the sentence that consists of the word 'Snow' followed by the word 'is' followed by the word 'white' '; alternatively: 'The sentence that consists of three words, the 1st of which is the string of the letters Es, En, O and DoubleU, the 2nd is the string of the letters I and Es, and the third is the string of the letters Double-U, Eitch, I, Te, and E'.

[68] As David (2008) shows, the situation is even more complicated with Tarski's original formulation of Convention T, which refers to the particular object-language that Tarski considers (the language of the calculus of classes) and to the particular meta-language, though it makes the appearance of generality.

[69] Tarski (1983: 157).

Now, if the definition of truth for the object language L is to be formally correct, it is imperative to specify exactly the structure of L as well as of the meta-language ML in which the definition will be formulated. Now, as Tarski's analysis of semantic paradox showed that a consistent truth definition for L can be given only if L does not express its own semantics (in particular, does not contain its own truth predicate), and hence cannot be given in L itself, L and ML may not be identical or inter-translatable (though L may be a proper part of ML). Consequently, if truth for L is to be definable in ML, ML has to contain means that are necessary for constructing a materially adequate definition of truth for L: (a) means enabling it to talk about signs and expressions, sequences of signs and expressions, as well as of sets of expressions, operations on expressions, etc. (in short: means sufficient to express syntax of L – conceived of as an axiomatic theory); (b) all non-logical and logical expressions of L or their translations (this is essential, if ML is to express T-biconditionals for sentences of L); (c) variables whose order exceeds the order of any variable of L, or quantifiers ranging over arbitrary subsets of the universe of discourse associated with L (ML has to be, in Tarski's words, essentially stronger than L).[70]

7. Truth-definition for finite languages

Tarski's strategy was to demonstrate how the method of truth definition works for a particular formalized language belonging to a certain comprehensive group (that is, fully interpreted extensional languages with quantifiers, whose syntactic structure is exactly because purely formally specified, and that are free of context-sensitive or ambiguous expressions), and then to argue that the method can be extended to other language from that group. For our purposes, however, it is not necessary to explain Tarski's method of truth-definition using his example (the language of the calculus of classes). On the other hand, it will be useful to demonstrate how the method works for languages L_0, L_1 and L_2 (or better, fragments of a full-blooded language) of increasing complexity, ending up with L_2 whose structure is analogous to the structure of those languages that belong to Tarski's group.

Let's start with L_0, a formalized fragment of German containing three sentences 'Der Schnee ist weiss'; 'Das Grass ist grün' and 'Der Himmel ist blau', with their usual English interpretations (we assume that a well-behaved fragment of English is our meta-language). The following, list-like definition for L_0

(D1) *s* is a true sentence of L_0 iff
(*s* = 'Der Schnee ist weiss' and snow is white) or (*s* = 'Das Grass ist grün' and the grass is green) or (*s* = 'Der Himmel ist blau' and the sky is blue)

[70] Tarski's interest was in the formalization of deductive theories in logical languages (within the system of simple theory of types). For Tarski, then, specification of the structure of a formalized language was closely connected with specification of logical axioms and derivation rules, as well as axioms specific to the particular deductive theory formalized in the language. All this was to be done in a purely formal manner, so that it could always be decided whether a string of expressions of L counts as an axiom (logical or theory-specific) or proof. Strictly speaking, this feature of L is inessential to the definition of the notion of truth-in-L, but it allows Tarski to give important meta-mathematical results about deductive theories.

is perfectly good by Tarski's lights, because it is explicit and uses no undefined semantic terms, and all the partial definitions of 'true' for L (T-biconditionals for sentences of L) can be deduced from it, given certain obvious syntactical facts: 'Der Schnee ist weiss' ≠ 'Das Grass ist grün'; 'Der Schnee ist weiss' ≠ 'Der Himmel ist blau'; 'Das Grass ist grün' ≠ 'Der Himmel ist blau'. To check this, let us substitute "Der Schnee ist weiss" for '*x*' in the definition, thereby obtaining the following:

> 'Der Schnee ist weiss' is true iff ('Der Schnee ist weiss' = 'Der Schnee ist weiss' and snow is white) or ('Der Schnee ist weiss' = 'Das Grass ist grün' and the grass is green) or ('Der Schnee ist weiss' = 'Der Himmel ist blau' and the sky is blue)

Since we can eliminate the obviously true identity in the first disjunct as well as the obviously false second and third disjunct, what we are finally left with is the T-biconditional:

> "Der Schnee ist weiss" is true iff snow is white

As wanted. This method, however, does not work for languages containing more than finitely many sentences, unless we are prepared to allow infinite disjunctions as acceptable definitions (as Tarski was not). But languages of theoretical interest have, as a rule, more complex structure, which forced Tarski to look for a more general method of truth-definition.

8. Truth-definitions for propositional languages

Let us move to Tarskian truth-definitions for more complex languages. We obtain one, namely L_1, by expanding L_0, adding the constructions "Es ist nicht der fall, dass", "und" and "oder" (in their usual English interpretations), by repeated application of which an infinitude of compounded sentences can be formed from the three atomic sentences of L_0. We can then define 'true' for L_1 in a well-known recursive (inductive) manner (where '*A*' and '*B*' are taken to range over arbitrary sentences of L_1 and italicized complex expressions are to be read is if they were in between Quine's corner quotes):

> (D2) *s* is a true sentence of L_1 iff
> (*s* = 'Der Schnee ist weiss' and snow is white); or (*s* = 'Das Grass ist grün' and the grass is green); or ('Der Himmel ist blau' and the sky is blue); or (*s* = *Es ist nicht der fall dass A* and *A* is not true); or (*s* = *A und B* and both *A* and *B* are true); or (*s* = *A or B* and *A* is true or *B* is true)

This definition is implicit, since the defined predicate appears in clauses fixing its application-conditions for compound sentences. With help of some elementary set-theory, however, it can be turned into an explicit definition, in which the original clauses are transformed to specify the conditions on membership in a certain set, call it 'TR':

> (D3) *s* is a true sentence of L_1 iff there is a set TR such that s belongs to TR, and for every sentence *y* of L_1, *y* belongs to TR iff

(a) y = 'Der Schnee ist weiss' and snow is white; or (b) y = 'Das Grass ist grün' and the grass is green; or (c) y = 'Der Himmel ist blau' and the sky is blue; or (d) y = *Es ist nicht der fall dass* A and A does not belong to TR; or (e) $y = A$ *und* B and both A and B belong to TR; or (f) $y = A$ *or* B and A belongs to TR or B belongs to TR

It can easily be shown that, for any given sentence of L_1, its T-biconditional is derivable from this definition, the material adequacy of the definition being thereby ensured.

9. Truth-definitions for quantificational languages

We have shown how to define the truth-predicate for L_1 so that T-criterion is satisfied, by fixing the truth-conditions of any complex sentence of L in terms of the truth-conditions of its less complex component sentences, ultimately in terms of the truth-conditions of simple sentences whose truth-conditions are fixed directly in their corresponding T-biconditionals. If, now, we expand L_1 by adding to it iterative constructions of a different kind, namely quantifiers, we obtain a potentially infinite number of complex sentences that cannot be dealt with in this way, because their immediate constituents are no longer truth-evaluable sentences, but sentential functions (a generalized version of Frege-type predicates), e.g.: 'x is blue' or 'x loves y', 'x is a man and x loves y', 'For all x, if x loves y, x is a man', etc. Sentential functions are neither true nor false as they stand: (a) they are true or false only relative to specific assignments of appropriate values to their free variables; alternatively: (b) they are true or false only relative to specific substitutions of denoting terms for their free variables. If every object in the universe of discourse associated with a language L (e.g., a first-order language of arithmetic) is denoted by a term of L, we can well follow (b) and define the truth-conditions for quantified sentences of the form $\forall x_k A(x_k)$ in terms of the truth-conditions of sentences, in this way:

> A sentence of the form $\forall x_k A(x_k)$ is true iff $A(t)$ is true, whatever (denoting) term t of L we put for 'x_k' in '$A(x_k)$'.

But, of course, L may not happen to contain a name of every object in its universe of discourse (L may even have no names, as was the case with the language of the calculus of classes, for which Tarski provided his truth-definition). Consequently, truth for such L cannot be defined recursively on the complexity of sentences of L. Because sentential functions can have more argument-places that are marked by different variables which can be replaced by more terms or closed by prefixing to them more quantifiers, it is better to talk, generally, about satisfaction of n-argument sentential functions by ordered n-tuples of objects. Tarski actually defined something more general: the relation *x satisfies y*, where x is an infinite sequence of objects (from the universe of discourse associated with L) and y a sentential function of L. This satisfaction relation is the converse of the relation *y is true of x*, generalized so as to cover sentential functions with arbitrary (but finite) number of argument-places.

To explain how this works, let us consider a regimented fragment of German, L_2, that has no simple or complex terms (to keep the definition simple), and its only non-logical constants are: 'ist ein Mann', 'ist eine Frau', 'liebes' (all having their usual English interpretations). Atomic sentential functions of L_2 are formed by attaching one or two variables (taken from the infinite sequence: 'x_1', 'x_2',…,'x_n') to its non-logical constants, while complex sentential functions and sentences of L_2 are formed from its atomic sentential functions via the operations of negation '¬', conjunction '∧', disjunction '∨', and universal quantification '∀' (all with their usual interpretations). Sentential functions of L_2 are defined inductively:

(D4) f is a sentential function of L_2 iff either
(a) $f = x_iP$, for some predicate P of L_2 and positive integer i; or
(b) $f = \neg A$, for some sentential function A of L_2; or
(c) $f = A \wedge B$, for some sentential functions A and B of L_2; or
(d) $f = A \vee B$, for some sentential functions A and B of L_2; or
(e) $f = \forall x_i\,(A)$, for some sentential function A of L_2 and positive integer i.

The definition of the satisfaction relation for L_2 then mimics that recursive definition of sentential function, proceeding by recursion on the complexity of a sentential function of L_2: it is first specified under what conditions a given infinite sequence p of objects satisfies simple sentential functions of the form x_iA, and then the conditions are specified under which p satisfies complex sentential functions of the forms $A \wedge B$, $A \vee B$, $\forall x_i\,(A)$, in terms of the conditions under which p satisfies (or does not) less complex sentential functions that are their immediate components:[71]

(D5) p satisfies f iff p is a sentential function and f is an infinite sequence of objects such that either
(a) f = 'x_k ist ein Mann' and p_k is a man; or
(b) f = 'x_k ist eine Frau' and p_k is a woman; or
(c) f = 'x_k liebes x_l' and p_k loves p_l; or
(d) $f = \neg A$ and p does not satisfy A; or
(e) $f = A \wedge B$ and p satisfies A and p satisfies B; or
(f) $f = A \vee B$ and p satisfies A or p satisfies B; or
(g) $f = \forall x_k\,A$ and every infinite sequence of objects p^* satisfies A that differs from p at most at the k-th place

Now, whether or not p satisfies f depends only on those members of p that are paired with the free occurrences of variables of f, provided it has any. But, mind you, sentences are 0-argument sentential functions with no variables free. Accordingly, whether or not

[71] Henceforth, I omit the qualifications: 'for some sentential function A (A and B) and positive integer k (k and l)'. 'p_k' denotes the k-th member of the infinite sequence p. Since variables and objects in a sequence are ordered, every variable occurring in a sentential function gets paired with exactly one object in the sequence, via its numerical index. Such pairing can be considered as completely non-semantic so that Tarski could avoid talking about assignments of objects to variables.

a sentence is satisfied by *p* does not depend on what members of *p* are paired with its free variables. So, there are only two possibilities to consider: either a sentence is satisfied by all sequences, in which case it is true, or it is satisfied by no sequence, in which case it is not true. Truth for L_2 is thus defined directly:

s is a true sentence of L_2 iff every infinite sequence of objects satisfies *s*

The truth definition for L_2, based on (D5), enables us to prove T-biconditionals for all sentences of L_2. But Tarski also required that no semantic notion be used in the definition of truth unless it can be shown that it is definable in purely non-semantic terms of the language in which the truth definition is framed. Now, owing to its recursive character, (D5) fixes the application conditions of the satisfaction predicate for L_2 (fixes what sequences satisfy its sentential functions), but it does not allow us to eliminate the predicate from all contexts (we do not have a formula free of that predicate that could replace it in all contexts of the metatheory). Fortunately, (D5) can be turned to an explicit definition, provided that our meta-language (meta-theory) has a sufficiently strong set-theory, or higher order variables than L_2:

(D6) *p* satisfies *f* iff there is a set S such that $<p, f>$ belongs to S and for every sentential function *g* and infinite sequence of objects *q*, $<q, g>$ belongs to S iff
(a) $g =$ 'x_k ist ein Mann' and q_k is a man; or
(b) $g =$ 'x_k ist eine Frau' and q_k is a woman; or
(c) $g =$ 'x_k liebes x_l' and q_k loves q_l; or
(d) $g = \neg A$ and *A* does not belong to S; or
(e) $g = A \wedge B$ and both $<q, A>$ and $<q, B>$ belong to S; or
(f) $g = A \vee B$ and $<q, A>$ belongs to S or $<q, B>$ belongs to S; or
(g) $g = \forall x_k A$ and $<q^*, A>$ belongs to S, for any infinite sequence of objects q^* that differs from *q* at most at the *k*-th place

Once we have this explicit (eliminative) definition of satisfaction, truth can itself be defined in purely non-semantic terms of the metalanguage as follows:

s is a true sentence of L_2 iff there is a set S satisfying the conditions (*a*), …(*g*) in (D6) above, and for every infinite sequence of objects *p*, $<p, s>$ belongs to S

Since the conditions for the membership in S (in D6) have been characterized in non-semantic terms of the meta-language, Tarski succeeded in showing how to define truth for L_2 in non-semantic terms (the definition being formally correct by his lights). With such a definition at hand, we can eliminate the semantic notion of satisfaction in favour of non-semantic terms of the metalanguage – (a) the translations of expressions of L_2, (b) logical and/or set-theoretic expressions, and (c) syntactic expressions – and semantic notions are only as controversial as the apparatus in terms of which they are introduced. In particular, a consistent metalanguage (or metatheory) remains consistent if enriched by semantic notions defined explicitly. Moreover, owing to its recursive character, a Tarski-style truth definition for a quantificational L not only

entails T-biconditionals for all sentences of L, but it entails also important generalizations couched in terms of truth, e.g.: that a conjunction is true just in case both its conjuncts are true; or that of a sentence and its negation exactly one is true and exactly one is false, and more of this sort. On this basis, then, basic metalogical theorems can be precisely stated and proved concerning consistency, soundness or completeness of deductive theories expressed in the object-languages. Given that Tarski focused primarily on formalized languages of mathematical theories, he succeeded in showing that their metalogic, including the truth theory and semantics, can be expressed (interpreted) in their logically (or set-theoretically) stronger extensions, whose subject matter is mathematical too.

10. The significance of Tarski's conception: logic

In the 20s of the past century, the pioneers of model theory Löwenheim and Skolem commonly worked with the informal, metatheoretic notion of truth or satisfaction (in a structure) and the related semantic notions. And other leading mathematical logicians – the members of Göttingen school, Gödel and Tarski himself – soon followed the suit. That usage was considered to be not only intuitive but also safe with respect to known semantic paradoxes.[72] As Robert Vaught remarks in his useful survey of model theory before 1945, since the notion of a sentence σ being true in a given structure

> […] is highly intuitive and (perfectly clear for any definite σ), it had been possible to go even as far as the completeness theorem by treating truth (consciously or unconsciously) essentially as an undefined notion – one with many obvious properties […].[73]

In view of this, it does not seem, *pace* many commentators of Tarski, that there was an urgent need for an analysis of truth (and related notions) to ensure its clarity or consistency in the eyes of Tarski's fellow logicians.[74] As was already suggested in section (2), what bothered Tarski as a mathematical logician was rather the fact that truth and related ideas were used intuitively and informally and there were no precise definitions or theories making their properties precise, and that there were no definitions of such notions in terms of some respectable deductive framework in which all of mathematics could be expressed (if not completely axiomatized – as Gödel's first incompleteness theorem states), which meant that basic metalogical results formulated in terms of semantic notions could not be stated, still less proved within such frameworks.[75] As Vaught continues:

---

[72] Ways were known how to avoid truth-paradoxes; e.g., in his ramified type theory, Russell (1908) anticipated to some extent Tarski's own hierarchical solution.

[73] Vaught (1974: 161).

[74] As Feferman (2008: 79) says: 'there were no uncertainties or anomalies in informal model-theoretic work that would have created any urgent need for conceptual analysis of semantical notions to set matters right.'

[75] See Tarski (1944: 369). Cf. Vaught (1974), Feferman (2008).

> […] But no one had made an analysis of truth[…] At a time when it was quite well understood that 'all of mathematics' could be done, say, in ZF, with only the primitive notion of ∈, this meant that the model theory (and hence much of metalogic) was indeed not part of mathematics. It seems clear that this whole state of affairs was bound to cause a lack of sure-footedness in metalogic […].[76]

The framework that Tarski and his contemporaries had in mind was either a standard axiomatization of set theory (Zermelo-Fraenkel system), or some version of simple type theory, which was assumed to be consistent (though, of course, its consistency cannot be proved within it, as Gödel's second incompleteness theorem states). Tarski's primary logical aim was to show that metalogic, qua theory of truth-theoretic (semantic) and proof-theoretic (syntactic) properties of formalized deductive theories and their relations, could itself be dealt with in purely mathematical spirit. He achieved this aim by showing how to construct mathematically precise and extensionally adequate definitions of truth for properly regimented and fully interpreted object-languages in logically (set-theoretically) stronger meta-languages (meta-theories). As he said:

> […] meta-mathematics is itself a deductive discipline and hence, from a certain point of view, a part of mathematics; and it is well known that – due to the formal character of deductive method – the results obtained in one deductive discipline can be automatically extended to any other discipline in which the given one finds an interpretation […].[77]

Metalogic should thus enjoy the same respect as the framework in which it is expressed; in particular, it is consistent so long as the framework is consistent, because notions introduced via explicit definitions cannot intrude any inconsistencies into the theory, provided that the theory was consistent before their introduction.

Furthermore, Tarski suggested a generalization of the method of truth definition to uninterpreted quantificational languages, namely the relative (as opposed to absolute) definitions of *truth (satisfaction)* in a structure or interpretation, in terms of which the notions of *validity*, *satisfiability*, and *logical consequence* are standardly defined. If L is an interpreted language, the method of absolute Tarskian truth-definition applies to it, defining what it means for a sentence of L to be true, period, based on the previous definition of what it means for a predicate to be satisfied by a sequence (assignment) s of elements from the universe of discourse associated with L. If, on the other hand, L is a formal language whose extra-logical expressions (constants) are given no fixed (intended) interpretation, the method of relative truth-definition applies to it, defining what it means for a sentence of L to be true in a structure *M* (*M* being a model of it), based on the previous definition of what it means for a formula to be satisfied by a sequence (assignment) s of elements (of the domain of *M*) in *M*.[78] In particular, we owe

[76] Vaught (1974: 161).
[77] Tarski (1944: 369).
[78] See my footnote n. 50.

to Tarski the classic *semantic* definition of the notion of logical consequence: sentence *A* is a logical consequence of set of sentences *B* iff each structure that is a model of all the sentences belonging to *B* is also a model of *A*.[79] Although there is a lively dispute about whether the now standard model-theoretic account of logical consequence (which we also owe to Tarski) is essentially the same as Tarski's classic definition, and about whether the standard account is conceptually adequate, nobody disputes that Tarski contributed greatly to logic by his method of language-relative truth definition.[80]

11. The significance of Tarski's Conception: philosophical questions

Truth is not only a notion dear to logicians, but also one of those notions that always provoked attempts at philosophical explanation or analysis. Now, philosophers are usually not interested in purely stipulative definitions, but in definitions that capture or elucidate the meaning of some interesting notion that is already in use. As we have already noted, it does not make sense to ask of a purely stipulative notion whether or not it is adequate or good, whether or not it gets things right. If, on the other hand, one gives a definition of a notion that is already in use, we can ask not only whether the definition gets the extension of the notion right, but also whether it does a good job in capturing its meaning.[81] Judging from Tarski's own notorious claim that 'the desired definition… aims to catch hold of the actual meaning of an old notion',[82] he seemed to think that his truth definition for L is more than extensionally adequate surrogate for the ordinary sentential truth-predicate restricted to L. However, Tarski surely did not want to give a direct analysis of the ordinary notion of truth, for that notion does not seem to apply to sentences only but also to beliefs, statements, or propositions (if there are such things). He did not even want to give a direct analysis of the ordinary notion of sentential truth, which can be applied to any sentence whatever of any language, because he argued that this notion cannot be consistently defined in its generality. So, if Tarski really intended his truth-definitions to be more than extensionally adequate (agreeing in extension with 'true' over particular languages), then his truth definitions are to be considered rather as partial explications (in something like Carnap's sense) of the ordinary notion of truth, in the sense that they are consistent approximations of the semantic concept of truth predicated of sentences. Convention T then states in more precise terms the necessary and sufficient conditions that a truth definition must meet to be faithful, qua such an explication, to the semantic notion of truth. And if the definition is materially adequate in that it satisfies convention T, it is thereby assured to be extensionally adequate.

If we see the situation in this light, the question arises whether Tarskian truth definitions are really good or plausible explications of the notion of truth (as restricted to particular languages). The opinions of philosophers on this question have differed

[79] Tarski (1936b).

[80] The standard model-theoretic account goes back to Tarski and Vaught (1956). Etchemendy (1988) argues that Tarski's classic definition of logical consequence differs in important respects from the standard model-theoretic one; in his later book (1990) he levels some interesting objections against the standard account.

[81] For a good account of definitions see Belnap (1999).

[82] Tarski (1944: 341).

considerably. Thus, whereas some folks claim that Tarski succeeded splendidly in explaining (or explicating) the notion of truth in clear and precise terms, and gave us a satisfying account of that notion,[83] other people retort that Tarskian definitions fail so blatantly as explanations of the ordinary notion of truth that they cannot simply be of any philosophical interest.[84]

We should first note that Tarski's method is considerably limited in its scope. First, it is designed to apply only to a certain family of properly restricted extensional languages, without any suggestion whether, or how, Tarski's criteria and techniques might be supplemented or modified to cover syntactically richer languages, including natural languages or languages approximating their complexity (one immediately thinks of so-called intensional or hyper-intensional constructions). Second, the method is designed to apply only to languages in which there is no context-sensitivity, no ambiguity or vacuous expressions. But natural languages, on which formal semanticists focus their attention, abound in such phenomena. Third, in view of semantic paradoxes, Tarski argued that no consistent definition or even theory of truth that meets his adequacy condition can be given for a language that contains or can express its own notion of truth (or satisfaction) unrestrictedly applying to all its sentences (predicates), provided that we assume that classical bivalent logic holds. But natural languages certainly do seem to contain such notions of truth (satisfaction), and so Tarski concluded that neither natural languages nor properly regimented languages that approximate them in expressive power can be given consistent truth-definitions in his style.

In spite of the fact that Tarski's method is so limited in its scope, Davidson, Montague and others have persuaded many theorists that Tarskian truth definitions (or their model-theoretic extensions) provide at least a *Muster* that may guide constructions of more sophisticated semantic theories for richer languages, including substantial fragments of natural languages. It is natural to call a theory semantic if it systematically articulates the truth-conditions of sentences on the basis of the semantic properties of their significant syntactic parts and the syntactic mode in which the parts are combined. Sentences thus need to be thought of as constructible from a finite stock of simple expressions according to a couple of *syntactic* rules of admissible combination. The semantic theory first fixes the interpretations of each simple expression, by assigning to it a semantic property of the type appropriate to its syntactic category, and then gives a couple of *semantic* rules that determine the semantic properties of complex expressions, given the semantic properties of their simpler constituent parts and their syntactic mode of combination. All this is arranged in such a way that the semantic properties delivered by the rules for sentences turn out to be the conditions under which the sentences are true. Plausible accounts along these lines have to cover elementary as well as more complex sentences,

---

[83] Carnap, Popper and Quine are the most prominent figures belonging to this camp. Popper saw in Tarski's method a rehabilitation of the correspondence theory of truth. The position of Carnap and Quine, on the other, is closer to modern deflationism (though they too claimed that Tarski's work captures in a precise form a certain vital intuition behind the correspondence theory). Indeed, inspired by Quine (1970), modern deflationists like Horwich (1982) or Leeds (1978) claimed that Tarski laid solid foundations of a satisfactory theory of truth that shows truth to be a pleasingly light-weight, logical notion, with no philosophical implications.

[84] See Putnam (1985, 1994 a, b) for arguments in this spirit.

the details depending on the complexity of the language under consideration. Now, the two Tarskian methods of truth-definition – absolute and relative – influenced respectively the two most influential truth-conditional approaches to formal semantics: the truth-theoretic of Davidson and his followers and the model-theoretic of Montague and others. In fact, it has even been claimed – although Tarski himself would not have gone so far – that plausible semantic theories for substantial fragments of natural languages could or should have *something like* Tarski-style truth-definitions as their basis, though modified or supplemented to accommodate features not present in languages with simple logical syntax. Donald Davidson went even so far as to suggest that we can use Tarski's method to give a compositional theory of meaning for a given natural language L:

> […] There is no need to suppress, of course, the obvious connection between a definition of truth of the kind Tarski has shown how to construct, and the concept of meaning. It is this: the definition works by giving necessary and sufficient conditions for the truth of every sentence, and to give truth conditions is a way of giving the meaning of a sentence. To know the semantic concept of truth for a language is to know what it is for a sentence – any sentence – to be true, and this amounts, in one good sense we can give to the phrase, to understanding the language. […] Indeed … a Tarski-type truth definition supplies all we have asked so far of a theory of meaning […].[85]

In view of this, the significance of Tarski's method of truth definition would seem to be beyond reasonable question. But several philosophers argued that strictly Tarskian truth definitions are of no use as theories of meaning or even as semantic theories. In general, these arguments aim to show that there is more to the notion of truth than we can read off from particular Tarskian truth definitions given for particular languages.

The point can be pressed from various more-or-less related directions. As regards Davidson's influential program, at least in its early form, one obviously pertinent observation is that we cannot expect from a Tarskian truth-definition for L (even in the 'revealing', recursive form) that it tells us, via its satisfaction-conditions, truth-conditions or denotation-conditions, something revealing about the meanings of expressions and sentences of L, because we have to know in advance the meanings of non-logical and logical expressions in order to define truth for L in Tarski's style.[86] Thus, for instance, we cannot expect from the following sentence:

'Der Schnee ist weiss' is true iff snow is white

to tell us both what the mentioned sentence means and what 'true' means. For it cannot not serve as a partial explanation of the meaning of 'true' in Tarski style, unless we already understand the mentioned sentence and know that it means what the used

[85] Davidson (1984: 24).

[86] This observation is commonly attributed to Dummett (see 1973, 1978b), but Tarski was fully aware of it. See also Quine (1970) and Künne (2003).

sentence means. And it cannot serve as an explanation of the meaning of the mentioned sentence, unless we already know what 'true' means in it. Similarly, then, the recursive clause of the type:

… *A und B* is true iff *A* is true and *B* is true

cannot explain both what 'true' means and what 'und' means. The project of defining truth for L in Tarski's style is thus incompatible with the project of a theory of meaning for L, as conceived of by Davidsonians.[87]

Furthermore, it has been argued that the language-specific character of Tarskian definitions, based on list-like definitions for basic cases (be it simple sentences, predicates or terms of a language), makes them non-extendible: a particular truth definition for a given language does not state anything to guide one in extending it to new cases (when a new sentence, predicate or term is added to it, or a new language is considered).[88] As Dummett once put it: Tarskian truth definitions do not tell us what the point of introducing truth predicates is.[89] Note that the worry is not that there is no hint as to how to extend a truth definition to expressions of a more complicated logical form (or to new languages containing such expressions). The point is that the definition in itself gives us no hint as to how to extend it to new cases that are logically familiar. We may illustrate this on the trivial truth definition for L_0. Thus, let us imagine that we add to the language L_0 a new sentence 'Die Sonne ist gelb', thereby obtaining a new language L_4. It may seem that we can easily extend the definition (D1) to a new truth definition by adding an additional clause for the new sentence.

s is a true sentence of L_3 iff
(a) *s* = 'Der Schnee ist weiss' and snow is white or (b) *s* = 'Das Grass ist grün' and the grass is green, or (c) *s* = 'Der Himmel ist blau' and the sky is blue, or
(d) *s* = 'Die Sonne ist gelb' and the sun is yellow

This extension of (D1) seems natural, provided that we know what the German sentences mean, or, at least, that they mean the same as the English sentences that are used to express the conditions they are paired with. However, nothing in the definiens of (D1) – which is supposed to explain all the meaning of the defined notion – dictates that we should extend it by pairing the sentence 'Die Sonne ist gelb' with the condition that it expresses (by using an English sentence that translates it). For all we can read off from (D1), it may be that (d) should read like this: 'Die Sonne ist gelb' and Jupiter is big (in which case, of course, truth for L_3 is not adequately defined). To be able to extend (D1) to an adequate truth definition for L_3 one needs to have some information not stated in (D1) (as is clear from the fact that possession of it is not necessary to understand (D1)): namely, (a) that

[87] Meanwhile, Davidson came to concede this point, and in the mid-70s started to talk about reversing Tarski's strategy: instead of taking the meanings of L-expressions for granted and defining truth-in-L, he proposes to take truth for granted (as a primitive concept) and give the meanings via a recursive truth-theory (as opposed to definition) for L.

[88] The point was made by Field (1972) and Dummett (1978b).

[89] Dummett (1978a: xx–xxi).

the English sentences used to express the conditions paired with the German sentences are their translations, and (b) that (D1) is intended to satisfy Convention T.

A closely related objection is that Tarski showed us at most how to define various predicates 'true in L_1', 'true in L_2',…, or 'true in L_n', each of which agrees in extension with 'true' with respect to a particular language L_1 or L_2 … or L_n, but he did not explain or elucidate our pre-theoretical notion of truth that can be applied to any given language. To put it slightly differently, Tarski did not explain or elucidate the notion of *truth in L*, where L is a variable (not even when L ranges only over properly restricted and regimented languages), because there is no hint in particular Tarskian truth definitions for particular languages as to what they have in common – why they deserve the common label of definitions of *truth*.

Another well-known type of objection is that the modal status of T-biconditionals differs from the modal status of the biconditionals in which 'true' is replaced by its explicit Tarskian definiens so that the Tarskian definiens cannot be an accurate explanation (explication) of 'true' (even when the later is restricted to a given language).[90] To explain what is at stake, let us again recall the truth-definition (D1). We know that with help of the obvious logical and syntactic principles we can derive from (D1) the T-biconditional for 'Der Schnee ist weiss':

'Der Schnee ist weiss' is a true sentence of L_0 iff snow is white

Now, as both Putnam and Etchemendy claimed, being a consequence of the metatheory that contains obviously valid syntactic and logical axioms and the definition (D1), the T-biconditional above is a logical or logico-syntactical, certainly non-empirical, necessary truth that does not depend on the meanings of expressions and sentences of L_0, or on the facts about their use by speakers. In particular, the biconditionals holds in all possible circumstances in which snow is white, irrespectively of what the 'Der Schnee ist weiss' means in such circumstances. But this is intuitively absurd, since the T-biconditional seems to make an empirical and contingent claim. Had Germans used 'white' to apply to black things (in a possible circumstance *C* in which snow is white), the biconditional would have turned false. If so, Tarskian definitions capture the extension of truth with respect to particular languages but do not capture its intension or meaning. As Putnam put it:

> […] A property that the sentence 'Snow is white' would have (as long as snow is white) no matter how we might use or understand that sentence isn't even doubtfully or dubiously 'close' to the property of truth. It just isn't truth at all […].[91]

However, this line of argument can be blocked. Several people argued that Tarski takes L to be individuated in part by the fact that particular expressions belong to it that are equipped with particular meanings.[92] So, if there is a single expression e whose meaning is different in L and L* respectively, then L and L* are different languages,

[90] See Putnam (1985, 1994 a, b) or Etchemendy (1988).
[91] Putnam (1985: 333).
[92] See Davies (1981), Garcia Carpintero (1996), Künne (2003).

even though they may be syntactically indistinguishable (comprised by exactly the same type-expressions). It follows that by adding a single expression to L or by changing the meaning of a single expression of L, we no longer have L but a different language L*. If we are imagining that a sentence or expression actually belonging to L could change its meaning contingently on its use by the community actually using L, we are imagining, strictly speaking, that a different, though closely related language L* would be used by the community. And once we individuate languages in this strict way, even informal T-biconditionals turn out necessary.[93]

A closely related but, I think, much more serious line of argument is concerned with one specific aspect of Tarski's method of truth definition: whether simple or complex, implicit or explicit, Tarskian definitions rest, in one way or another, on list-like or enumerative definitions of semantic notions (or truth, satisfaction, or denotation). If we look at the explicit definiens in (D1)

> (*s* = 'Der Schnee ist weiss' and snow is white) or (*s* = 'Das Grass ist grün' and the grass is green) or (*s* = 'Der Himmel ist blau' and the sky is blue)

we see that there is nothing obviously *semantic* about it, even though it meets Convention T and is thus faithful to the semantic conception of truth. It does not seem to throw any light upon how the truth-conditions of sentences are determined by the semantic properties, including semantic relations objects, of their significant constituents and their manner of composition (it is no use to say that it relates sentences to extra-linguistic entities, facts or states of affairs, since it obviously does not do that). In light of this, the truth definition for L_2 certainly appears to be more illuminating of the semantic structure of L_2 than the two previous truth-definitions are of the semantic structures of L_0 and L_1 respectively.

So, what this shows is that it is possible for a definition of truth for a language (at least for a simple language) to satisfy the demands of Tarski's semantic conception of truth, while not being *semantic* in any natural sense of that notion. That the truth predicate defined in this way does not have the right connections to semantic facts becomes transparent once we realize that if we did not understand German but had reliable information that the following informal T-biconditional is true

> 'Der Schnee ist weiss' is true (in German) iff snow is white

we would have at least some information about the meaning that 'Schnee ist weiss' has as a sentence of German, hence as a sentence of L_0 (assuming we understand the meta-lan-

[93] It should be noted that Putnam's argument could be blocked in a different way. We may, for instance, allow that L retains its identity even though its expressions change their meanings and semantic properties (contingently on their use by speakers of L), so that sentences that actually serve as true T-biconditionals for L can turn out false, other sentences serving as T-biconditionals for L. We may refuse to accept that Tarskian biconditionals are necessary truths derived from definitions framed in a mathematico-syntactical metatheory via logic. We may argue, for instance, that their modal status depends on the modal status of the definitions, and there is no hint that Tarski considered them necessary as opposed to materially true (rather, it is arguable that he preferred to avoid intensional notions). For lack of space I cannot deal with this interesting response here. But see Patterson (2008b).

guage in general, and 'true' in particular). We could infer, for instance, that it does not mean that snow is not white, that snow is black and other things incompatible with the fact that snow is white. Now, it is standardly assumed that an explicitly defined predicate can always be replaced by the definiens without any loss (throughout extensional contexts). But when we replace 'true' in the informal T-biconditional for 'Der Schnee ist weiss', by its explicit Tarskian definiens what we get is this:

> ('Der Schnee ist weiss' = 'Der Schnee ist weiss' and snow is white) or ('Der Schnee ist weiss' = 'Das Grass ist grün' and the grass is green) or ('Der Schnee ist weiss' = 'Der Himmel ist blau' and the sky is blue) iff snow is white

After performing admissible simplifications, this is equivalent to the following triviality

> Snow is white iff snow is white

which, obviously, does not state any information at all about the meaning or semantic properties of the sentence 'Der Schnee ist weiss'. Clearly, one can understand what this list-like truth-definition states without knowing anything at all about the semantics of L_0.

We have said that the truth definition for L_2 appears more illuminating of the semantic structure of L_2 than the truth-definitions for L_1 is of the semantic structures of L_1, and this in turn appears more illuminative than the truth definition for L_0 (because using recursion). But the appearance may be delusive. In fact, the critics argued that essentially the same considerations apply, *mutatis mutandis*, to the truth definition for L_1 and L_2. Especially if they are put in their explicit forms (and only these meet all Tarski's strictures), it becomes clear that one could understand what they state without knowing any semantic fact at all about L_1 or L_2. As Etchemendy and Soames have pointed out[94], when we replace in the T-biconditional for 'Der Schnee ist weiss' (as belonging to L_1), the predicate 'true' by its explicit Tarskian definiens, we get the following long claim:

> There is a set TR of sentences of L_1 to which 'Der Schnee ist weiss' belongs such that…

It could then be shown that after admissible simplifications of this claim we get something that, again, does not state any information at all about the meaning or semantic properties of the sentence 'Der Schnee ist weiss' in L_1.[95] Even at the intuitive level: we readily understand this claim, but unless we know that TR is the set of true sentences of L_1 (which is not stated in the definition), we cannot, on the basis of this claim, infer anything concerning the meaning that 'Der Schnee ist weiss' has as a sentence of L_1.

Moving finally to the truth-definition for L_2, we immediately observe that the part of it that deals with the satisfaction conditions of simple sentential functions is trivially list-like or enumerative:

---

[94] Etchemendy (1988: 56–57); Soames (1999: 102–105).
[95] Cf Soames (1999: 104).

p satisfies *f* iff (*f* = 'x_k ist ein Mann' and p_k is a man) or (*f* = 'x_k ist eine Frau' and p_k is a woman) or (*f* = 'x_k liebes x_l' and p_k loves p_l)

But then essentially the same line of argument can be used to show that the whole definition has no semantic import.

12. So what, if anything, is the philosophical importance of Tarski's method?

It would seem that we must grant the critics that Tarski's method of truth definition has no semantic import. One cannot introduce semantic notions for L via purely logico-syntactico-mathematical definitions, and at the same time expect that such definitions will state something relevant about the semantic properties and structure of L. Once satisfaction or denotation and, in terms of them, truth is defined explicitly, there remain no semantic terms in their definiens, because they all disappear completely in favour of the terms of a semantics-free metalanguage (meta-theory). This is all to the good, if one is after extensionally adequate definitions of semantic notions in terms of set-theory, logic, syntax and the vocabulary of the object-language to which the definitions are relativized (which was arguably Tarski's main logical aims, for reasons discussed in the earlier parts of this paper). However, set-theoretical definitions, like those definitions that are trivially list-like, can capture only the extensions of restricted semantic notions but cannot possibly explain, explicate or reduce them to terms that are conceptually, ontologically or epistemically more fundamental. While mathematical logicians tend to emphasise that Tarski showed us how to do formal semantics (model theory) by precise mathematical methods (indeed, within set theory), philosophers are naturally not so impressed by this aspect, as is clear from the fact that they do not widely share Tarski's belief to have captured the meaning of an old notion of truth, or even to have established the scientific foundations of truth theory and semantics by showing how to define semantic notions in terms of respectable syntactic, logical and set-theoretical notions. This reaction is, I think, justified.[96] The fact that Tarski's succeeded in showing how to interpret the truth theory and semantics of L in a (logically stronger) formalized mathematical theory does not mean that he showed something philosophically important about truth and semantics.

I think that the critics are right in so far as they claim – contra some theorists with affinity to the so-called truth-deflationism[97] – that Tarski's method of truth definition does not tell us everything there is to the notion of truth by way of a philosophical account. In view of this, also Tarski's notorious claim that his truth definitions catch hold of the *actual meaning* of our intuitive notion of truth is unfortunate and misleading, to say the least. However, all this is of a limited importance as a critique of Tarski's method of truth definition. Tarski could well be confused or simply careless in his

[96] The *locus classicus* of the claim that Tarski did not really succeed in reducing semantic notions to scientifically (physicalistically) respectable notions is Field (1972).

[97] See e.g. Leeds (1978), Horwich (1982). In his (1990) Horwich no longer thinks that Tarski told us everything there is to know about truth, though he retained his deflationism (minimalism about truth). As we shall see, the relation of Tarski's conception to modern deflationist theories of truth is a rather delicate matter.

claims on this matter, but even though his truth definitions do not in fact catch hold of the actual meaning of an old notion of truth, his method might still provide a different sort of insight into the notion of truth than its analysis – showing us how to reconstruct the truth conditions of sentences in a language (of a certain type) as systematically depending on the semantic properties of their significant parts, based on their syntactic structure. What should be clear anyway but is often ignored or overlooked by the critics is that there is more to Tarski's method or conception of truth definition than particular formal definitions. Clearly, its heart is Convention T, stating the material adequacy condition for a satisfactory definition of truth for a given language, and the technique is, in general case, a recursive characterization of satisfaction (or denotation) based on syntax. In tandem, these two aspects indicate to us where the semantic import of Tarski's method lies.

It is an inherent part of Tarski's method that the truth definition for L is meant to be a materially adequate definition of truth for L, faithful to one powerful intuition about our ordinary notion of truth. This intention takes the form of the requirement of its implying all T-biconditionals for the language under study. And it is the fact that it has such consequences that confers upon it the status of the truth definition for L, for then the claim that all and only true sentences of L satisfy the definition holds. Consequently, a Tarskian truth definition for L cannot be taken as a purely stipulative definition, although it cannot be construed as a meaning or concept giving definition either (as the foregoing arguments have shown). The fact that it is intended to be materially adequate also shows that it is not divorced from meaning at all, but depends in a way on it (*viz*. Convention T and its appeal to the notion of translation), since every change in meaning of sentences calls for a brand new truth definition entailing a new set of T-biconditionals.[98] The semantic import of a Tarskian truth definition for L, as based on the recursive definition of satisfaction for predicates (or denotation for terms) of L, rests simply on the claim that the definition is a materially adequate definition of truth for L. True, in making this crucial claim we must use our ordinary notion of truth. But there is no vicious circularity involved, because the claim itself is not part of the definition framed in the metalanguage, but a higher level claim about our success in achieving what has been our goal all along. This, incidentally, is the reason why Tarskian truth definitions, without further ado, look semantically insignificant. However, once we know that the definition is materially adequate, we can read it as containing relevant information about the semantic properties of L, based on its compositional structure (it is here where the recursive technique plays its role). Indeed, Tarski himself did not hesitate to use the notion of truth in his original version of Convention T:

> […] A formally correct definition of the symbol 'Tr', formulated in the metalanguage will be called an adequate definition of truth if…[…].[99]

[98] This holds whether we consider this as a change of language (the standard response to the objection of Putnam and Etchemendy), or not (the non-standard response to the objection sketched in the footnote n. 119).

[99] Tarski (1983: 187–188).

What this shows, as Donald Davidson pointed out,[100] is that we are not wrong to interpret Tarski's method of truth definitions as a method of fixing the extension of 'true' for particular languages (properly formalizable), that takes full advantage of our pre-theoretical grasp of the notion of truth in the form of the semantic conception of truth. As Davidson and Richard Heck persuasively argued,[101] the claim that a Tarskian truth definition for a full-blooded quantificational L is materially adequate (satisfies Convention T) makes the definitions in a sense equivalent to an axiomatic theory of truth for L, whose axioms mimic the clauses of the recursive truth definition (*viz.* e.g. (D5)), with semantic notions construed as its primitives.

(A) Base axioms for simple predicates:
- p satisfies 'x_k is a man' iff p_k is a man
- p satisfies 'x_k is a woman' iff p_k is a woman
- p satisfies 'x_k loves x_l' iff p_k loves p_l

(B) Recursive axioms for predicates formed from simple predicates by means of constructions of conjunction, disjunction, implication and universal quantification:
- p satisfies $\neg A$ iff p does not satisfy A
- p satisfies $A \wedge B$ iff p satisfies A and p satisfies B
- p satisfies $A \vee B$ iff p satisfies A or p satisfies B
- p satisfies $A \rightarrow B$ iff p does not satisfy A or p satisfies B
- p satisfies $\forall x_k A$ iff every sequence p', which differs from p at most in its k-th member, satisfies A

For various reasons that I shall not rehearse here, the question of whether axiomatic theories along these lines can serve as empirical theories of understanding or interpretation, as Davidson famously claimed, is rather controversial.[102] But, at the very least, such theories do seem to throw some light upon the semantic structure of quantificational languages: how the truth conditions of sentences systematically depend on the truth-relevant semantic properties of their significant parts based on their logico-syntactic structure (thus revealing the inferential structure of quantificational languages). And that is no mean achievement.

Moreover, Tarski seemed to be fully aware that his method of truth definition had this dimension, because he said that in order to capture the truth conditions (T-biconditionals) for each of the indefinite number of sentences of a reasonably rich language L, the most *simple* and *natural* way is to proceed through a recursive characterization of the satisfaction conditions for complex predicates in terms of the satisfaction conditions of simpler predicates that are their immediate significant constituents (if L has complex terms, a recursive characterization of the denotation conditions is added).

---

[100] Davidson (1990).
[101] Etchemendy (1988), Davidson (1990), Heck (1998).
[102] Davidson (1990).

> […] it turns out that the simplest and the most natural way of obtaining an exact definition of truth is one which involves the use of other semantic notions, e.g., the notion of satisfaction […].[103]

He considered it natural, I dare say, because it is a rather intuitive thought that semantic properties of complex expressions, and hence the truth conditions of sentences, systematically depend on the semantic properties of their significant constituents. Logicians have always honoured intuitions such as the following: a sentence of the form *name + predicate* is true just in case the predicate is true of what the name denotes (alternatively: …just in case what the name denotes has the property that the predicate denotes).[104] Such intuitions were elaborated and generalized in various ways, but the essential idea remained: truth or falsity of sentences (of at least certain forms) depends on the semantic properties of their significant syntactic parts, so that it is possible to specify the condition under which a sentence (of an appropriate form) is true in terms of the semantic properties of its parts. Tarski knew that in order to define adequately the notion of truth for a quantificational language that has an infinitude of sentences each of which has a certain exactly specifiable logico-syntactic form, it is *natural*, indeed inevitable, to take full advantage of the fact that truth or falsity of sentences of increasing logico-syntactic complexity depend on the semantic properties of their immediate constituent parts, and ultimately on the semantic properties of their simple parts.

When we are concerned with languages worth of that name (and not with oversimplified fragments like L_0 and L_1) a satisfactory characterization of truth conditions for sentences of a language, like the one we gave for L_2, is naturally going to be framed in terms of the relations of satisfaction or denotation (or something analogous) between expressions and objects; and, typically, such relations will have to be defined recursively on the logico-syntactic structure of expressions. Such a definition displays the contribution of that structure to the truth-conditions of sentences of L_2, and provides also the basis for a precise account of logical consequence. This is the real import of Tarski's method of truth definition, in which its philosophical importance largely consists.

13. Conclusion

Tarski's method of truth definition has various aspects: logical, philosophical and mathematical. I have tried to show that once we properly understand and distinguish the various aspects of the method, its contribution to semantics, *a fortiori* to philosophy, dwells in the fact that a recursive truth-definition for a reasonably rich L is equivalent to a compositional axiomatic truth-theory for L – with semantic terms construed as its primitives – which illuminates the semantic structure of L. However, Tarski's method does not give us satisfying answers to so-called metasemantic (foundational) questions in semantics and theory of meaning: *Why (in*

[103] Tarski (1944: 345).
[104] Such intuitions are arguably present already in the works of Plato (Cf. the *Sophist*).

virtue of what facts) does L have the semantics it has?; How (on the basis of what evidence) can we tell that a truth theory (or semantics) for L is correct? In particular, we should not buy Tarski's claim to have captured the meaning of the notion of truth, or even to have established the scientific foundations of truth theory and semantics by showing how to reduce semantic notions to respectable syntactic, logical and set-theoretical notions.

References

Almog, J., Perry, J. and Wettstein, H., eds. (1989): *Themes from Kaplan*. Oxford University Press, Oxford.

Ayer, A. (1952): *Language, Truth and Logic*. Dover Publications, New York. Reprint of the second edition 1946.

Blackburn, S. and Simmons, K., eds. (1999): *Truth*. Oxford University Press, Oxford.

Butler, R. J., ed. (1962): *Analytical Philosophy*. Blackwell, Oxford.

Belnap, N. (1999): 'On Rigorous Definitions'. *Philosophical Studies* 72: 115–146.

Carnap, R. (1942) *Introduction to Semantics*. Harvard University Press, Cambridge MA.

Carnap, R. (1949): 'Truth and Confirmation'. In: H. Feigl and W. Sellars (eds.) *Readings in Philosophical Analysis*, Ridgeview Publishing Company, Atascadero CA, 1949, pp. 119–127.

Candlish, S. and Demnjanovic, N. (2007): 'A Brief History of Truth'. In: D. Jacquette (ed.) *Handbook of the Philosophy of Science. Volume 5: Philosophy of Logic*, Elsevier BV, Netherlands, pp. 273–369. (Handbook editors: Dov Gabbay, Paul Thagard and John Woods).

Cartwright, R. (1962): 'Propositions'. In: Butler (1962): 81–103.

David, M. (1996): 'Analyticity, Carnap, Quine and Truth'. *Philosophical Perspectives* (10): 281–296.

David, M. (2008): 'Tarski's Convention T and the Concept of Truth'. In: Patterson (2008a): 133–156.

Davies, M. (1981): *Meaning, Quantification and Necessity*. Routlege, London.

Davidson, D. (1984a). *Inquiries into the Truth and Interpretation*. Oxford University Press, Oxford.

Davidson, D. (1984b): 'Truth and Meaning'. In: Davidson (1984a): 17–36.

Davidson, D. (1990). 'The Structure and Content of Truth'. *Journal of Philosophy* (87): 279–328.

Dummett, M. (1973). *Frege: The Philosophy of Language*. Duckworth, London.

Dummett, M. (1978a): *Truth and Other Enigmas*. Duckworth, London.

Dummett, M. (1978b): 'Truth'. In: Dummett (1978): 1–24.

Etchemendy, J. (1988): 'Tarski on Truth and Logical Consequence'. *Journal of Symbolic Logic* (53): 51–79.

Feferman, S. 2004: 'Tarski's Conceptual Analysis of Semantical Notions', in A. Benmakhlouf (ed.), *Sémantique et Épistemologie*, Casablanca–Paris: Le Fennec-Vrin, pp. 79–108. Reprinted in Patterson (2008a): 72–93.

Field, H. (1972): 'Tarski's Theory of Truth'. *The Journal of Philosophy* (69): 347–375.

Field, H. (1994): H. Field: 'Deflationist Views of Meaning and Content'. *Mind* (103): 249–285.

Frege, G. (1918/1919): 'Der Gedanke. Eine logische Untersuchung'. *Beiträge zur Philosophie des deutschen Idealismus* (1): 58–77. English translation in Blackburn and Simmons (1999): 85–105.

Frege, G. (1979): *Posthumous Writtings*. Blackwell.

Garcia Carpintero, M. (1996): 'What is a Tarskian Definition of Truth?' *Philosophical Studies* (82): 113–144.

Heck, R. (1997): 'Tarski, Truth and Semantics'. *The Philosophical Review* (106): 533–554.

Henkin, L. et al., eds. (1974): *Proceedings of the Tarski Symposium*. In: Proceedings of Symposia in Pure Mathematics, vol. XXV, RI: American Mathematical Society, Providence.
Hodges, W. (1985/6): 'Truth in a Structure'. *Proceedings of the Aristotelian Society* (86): 135–151.
Hodges, W. (2008): 'Tarski's Theory of Definition'. In: Patterson (2008a): 94–132.
Horwich, P. (1982): 'Three Forms of Realism'. *Synthese* (51): 181–201.
Horwich, P. (1990): *Truth*. Blackwell, Oxford.
Hughes, (1993): *A Philosophical Companion to First-order Logic*, Hackett Publishing Company, Indianapolis.
Kaplan, D. (1989): 'Demonstratives'. In: Almog et al. (1989): 565–614.
Künne, W. (2003): *Conceptions of Truth*. Oxford University Press, Oxford.
Leeds, S. (1978): 'Theories of Reference and Truth'. *Erkentniss* (13): 111–130.
Lewis (1983): *Philosophical Papers I*. Oxford University Press, Oxford.
Montague, R. (1974): *Formal Philosophy. Selected Papers of Richard Montague*. Yale University Press, New Haven.
Murawski, and Wolenski, J. (2008): 'Tarski and his Polish Predecessors on Truth'. In: Patterson (2008a): 21–43.
Lynch, M. P., ed. (2001): *The Nature of Truth*. MIT Press, Cambridge MA.
Neurath, O. (1983a): *Philosophical Papers (1913–1946)*. D. Reidel, Dordrecht.
Neurath, O. (1983b): 'Physicalism'. In: Neurath (1983a): 52–57.
Patterson, D., ed. (2008a): *New Essays on Tarski*. Oxford University Press, Oxford.
Patterson, D. (2008b). 'Tarski's Conception of Meaning'. In: Patterson (2008a): 157–191.
Putnam, H. (1985): *Representation and Reality*.
Putnam, H. (1994a): *Words and Life*. Ed. J. Conant, Harvard University Press, Harvard.
Putnam, H. (1994b): 'On Truth'. In: Putnam (1994a): 315–329.
Putnam, H. (1994c): 'Comparison of Something with Something Else'. In: Putnam (1994a): 330–350.
Quine, W. O. V. (1970): *Philosophy of Logic*. Englewood Cliffs, Prentice Hall.
Ramsey, F. P. (1999): 'Facts and Propositions'. Excerpt in: Blackburn and Simmons (1999): 106–107. Originally published in *Foundations of Mathematics*, ed. R. B. Braithwaite, Routledge and Kegan Paul, 1931.
Ramsey, F. P. (2001): 'The Nature of Truth'. In: Lynch (2001): 433–445.
Reichenbach, H. (1938): *Experience and Prediction. An Analysis of the Foundations and the Structure of Knowledge*. University of Chicago Press, Chicago.
Russell, B. (1908): 'Mathematical Logic as Based on the Theory of Types'. *American Journal of Mathematics* (30): 222–262.
Soames, S. (1984): 'What is a Theory of Truth?' *Journal of Philosophy* (81): 411–429.
Soames, S. (1999): *Understanding Truth*. Oxford University Press, Oxford.
Tarski, A. (1933): 'Pojęcie prawdy w językach nauk dedukcyjnych'. Prace Towarzystwa Naukowego Warszawskiego, Wydział III, no. 34. German translation 'Der Wahreitsbegriff in den formalisierten Sprachen' in *Studia Philosophica* (1): 261–405. English translation 'The Concept of Truth in the Languages of the Deductive Science', in Tarski (1956, 1983a): 152–278.
Tarski, A. (1936a): 'Grundlegung der wissenschaftlichen Semantik'. In: *Actes du Congrès International de Philosophie Scientifique*, vol. III, Paris, pp. 1–8. English translation in Tarski (1956, 1983): 401–408.
Tarski, A. (1936b): 'Über den Begriff der logischen Folgerung'. In: *Actes du Congrès International de Philosophie Scientifique*, fasc. 7 (Actualités Scientifiques et Industrielles, vol. 394), Hermann et Cie, Paris, pp. 1–11. English translation 'On the Concept of Logical Consequence' in Tarski (1956, 1983a): 409–420.
Tarski, A. (1956): *Logic, Semantics, Metamathematics*. Ed. J. Woodger, Oxford University Press, Oxford. 2nd edition Tarski (1983).

Tarski, A. (1983): *Logic, Semantics, Metamathematics, papers from 1923 to 1938*. Ed. John Corcoran, Hackett Publishing Company, Indianapolis.
Tarski, A. (1993): 'Truth and Proof'. In Hughes (1993): 101–125. Originally published in *Scientific American*, 1969, (220): 63–77.
Tarski, A. (1999): 'The Semantic Conception of Truth and Foundations of Semantics'. In: Blackburn and Simmons (1999): 115–143. Originally published in *Philosophy and Phenomenological Research*, 1944, (4): 341–376.
Tarski, A. and Vaught, R. L. (1956): 'Arithmetical extensions of relational systems'. *Compositio Mathematica* 13: 81–102.
Vaught, R. L. (1974): 'Model Theory before 1945'. In: Henkin (1974): 153–172.

ACTA UNIVERSITATIS CAROLINAE
PHILOSOPHICA ET HISTORICA 2/2007
miscellanea logica (viii)

Foundations of Logic

Layout and typeset: Kateřina Řezáčová
Published by Charles University in Prague
Karolinum Press, Ovocný trh 3, 116 36 Prague 1, Czech Republic
Prague 2010
Printed by Karolinum Press
First edition Number of copies: 200

ISBN 978-80-246-1815-9
ISSN 0567-8293
MK ČR E 18598

Distributed by the Faculty of Arts, Charles University in Prague,
2, Jan Palach Sq., 116 38 Prague l, Czech Republic
edice@ff.cuni.cz