Logic and Philosophy. A Reconstruction

(Preprint. Printed version forthcoming in: *Oxford Handbook of Philosophy of Logic*, ed. by E. Brendel, M. Carrara, F. Ferrari, O. Hjortland, G. Sagi, G. Sher, F. Steinberger, Oxford: Oxford University Press.)

Hannes Leitgeb

The article recapitulates what logic is about traditionally and works out two roles it has been playing in philosophy: the role of an instrument and of a philosophical discipline in its own right. Using Tarski's philosophical-logical work as case study, it develops a logical reconstructionist methodology of philosophical logic that extends and refines Rudolf Carnap's account of explication and rational reconstruction. The methodology overlaps with, but also partially diverges from, contemporary anti-exceptionalism about logic.

Keywords: logic, philosophy, methodology, explication, rational reconstruction, Tarski, Carnap, logical pluralism, normativity of logic

> I am convinced that we now find ourselves at an altogether decisive turning point in philosophy, and that we are objectively justified in considering that an end has come to the fruitless conflict of systems. We are already at the present time, in my opinion, in possession of methods which make every such conflict in principle unnecessary. What is now required is their resolute application.
>
> […] That the situation is unique and that the turning embarked upon is really decisive can be understood only by becoming acquainted with the new paths and by looking back, from the standpoint to which they lead, upon all those efforts that have ever passed as "philosophical".
>
> The paths have their origin in logic. (Schlick 1930: 5-6)

1. Introduction: What is Logic?

Logic is traditionally concerned with logical truth, consequence, and validity, with the logical form of linguistic expressions, and with logical concepts, such as negation or universal quantification.

Let me explain this by the following simple argument:

P_1 All Austrians are EU citizens.
P_2 Biden is not an EU citizen.
Therefore:
C Biden is not an Austrian.

Each of P_1, P_2, C happens to be empirically true, but that is logically irrelevant. What is relevant is that, intuitively, *if* P_1 and P_2 are *true*, *then* it *must* be the case that C is *true*, and hence one would be perfectly justified in arguing for C on the supposition of P_1 and P_2. In more logical terms, tradition has it that the premises P_1 and P_2 logically imply the conclusion C, or equivalently, that the argument is logically valid. Accordingly, the if-then sentence that can be constructed in correspondence with the argument,

S: If all Austrians are EU citizens and Biden is not an EU citizen, then Biden is not an Austrian.

is logically true — it *could not* be *false*. And one would come to the very same verdicts if "Austrian(s)", "EU citizen(s)", and "Biden" were replaced by other terms of the respective syntactic types, since the resulting argument and conditional would be of the same logical form as the respective previous ones, and all linguistic expressions of the same logical form must share all their logical features. Indeed, one might say that linguistic expressions have their logical features in virtue of their logical forms having them.

In first approximation, the logical forms themselves can be determined from the natural language expressions by replacing each descriptive term by a variable of the same syntactic type, while leaving all logical (and auxiliary) expressions invariant. In the present case, this would yield

P_1^* All *P* are *Q*.
P_2^* *a* is not *Q*.
Therefore:
C^* *a* is not *P*.

and

S^*: If all *P* are *Q* and *a* is not *Q*, then *a* is not *P*.

For logical purposes, it does not matter what the place-holders *P*, *Q*, *a* stand for: as long as the logical expressions "all", "not", "if-then", and "therefore" retain their meaning and are applied to the place-holders in the syntactically correct manner (which, in English, may involve the help of auxiliary copula expressions, such as "are" and "is"), it holds that P_1^*, P_2^* logically entail C^*, the *-argument is logically valid, and its corresponding conditional is logically true. That is: logical consequence is not just *necessarily truth-preserving* ("*if*… are *true*,… *must* be… *true*") and logical truths are not just *necessarily true* ("*could not* be *false*"), but logical consequence/validity/truth are also *formal* — they only depend on logical form.[1] Necessity and formality taken together entail that the conclusion C^* makes truth-conditional information explicit that is contained implicitly in the truth of the premises P_1^* and P_2^*, presupposing only the truth-conditional meaning of the logical constants and how these logical constants occur syntactically in the premises and the conclusion.

At this point, it is no longer obvious to which language the *-argument and its corresponding S^* belong, as they involve a mixture of natural language and formal symbols. This should be rectified, and with the arrival of modern logic in the second half of the 19th and the first half of the 20th century it did get rectified by the construction of artificial formal languages in which arguments from natural, mathematical, scientific or philosophical language could be logically represented so that only expressions from the vocabulary of the relevant formal language were used in the resulting logical forms, and where the logical forms satisfied the grammatical formation rules of the formal language in question.

In the example from above, if we determine its standard logical representation in the *formal language of first-order logic*, we get

P_1^{**} $\forall x(P(x) \rightarrow Q(x))$
$P2^{**}$ $\neg Q(a)$
$\therefore$
C^{**} $\neg P(a)$

and

S^{**}: $(\forall x\big(P(x) \rightarrow Q(x)\big) \wedge \neg Q(a)) \ \rightarrow \ \neg P(a)$

in which all occurrences of "is" have been replaced by predications, "All… are" by a universally quantified material if-then-formula, and the original logical expressions by the respective

[1] Aristotle was the first to systematically work out the formality of simple types of logically valid arguments, the syllogisms. For discussions of the necessity and formality of logic, see Sher (1996), Shapiro (1998), and Dutilh Novaes (2012: Section 1).

logical constants ∀, →, ¬, ∴ of first-order logic. In this way, the logical forms of the original sentences have been pinned down completely as well-formed formulas. (There would be alternatives to this logical representation, but let us put them to one side here.)

Finally, formal languages, such as first-order languages, can be supplied with a formal semantics or model theory by which some of the informal semantic locutions used before, including "*true*", "*stand for*", and "*meaning*", can be made precise, and by which the necessity and formality of logical consequence are captured by metalinguistic universal quantification over all models or interpretations of the language: *whatever semantic interpretation* is assigned to P, Q, a, if P_1^{**} and P_2^{**} are true under the interpretation, the same holds for C^{**}. In this way, logical consequence/validity/truth can be defined precisely on semantic grounds. And by laying down a deductive system of formal axioms and rules, one can define a logical derivability relation in equally exact terms, such that, at least for first-order logic, one can give a metalinguistic mathematical proof that the derivability relation coincides extensionally with semantic consequence. On that basis, the initial intuitive verdicts ("*Intuitively*, if…") can be made precise and backed up by proving the logical validity of the **-argument-form and the logical truth of its corresponding conditional-form: either semantically, by a metalinguistic mathematical proof based on the semantic definitions, or by using the axioms and rules of first-order logic to logically derive C^{**} from P_1^{**} and P_2^{**}, and to prove S^{**} without premises.

That is thus, *in nuce*, the task of logic: the systematic study of logical consequence, validity, and logical truth with their properties of necessity[2] and formality[3], the construction of formal languages, the logical representation of linguistic expressions as logical forms in these languages, the systematic investigation of these logical forms and of the underlying logical concepts, and the precise definition, study, and application of formal semantics and proof systems by which pre-theoretic logical verdicts can be sharpened and assessed and through which logic may benefit from the expressive and inferential power of mathematics.

That is logic as *discipline*. In contrast, *a* logic as *a system* is what we get when we put together a particular formal language or a family of such languages, L, a set of logical constants in L expressing logical concepts, a semantics with a logical consequence relation for L, and a deductive system with a logical derivability relation for L that is a subset of (if not equal to) the consequence relation, combined with a method of logically representing a fragment of natural, mathematical,… language in L. In this sense, logic as a discipline may be said to construct, study, and apply logical system*s* – in the plural, as there are many such systems. If one prefers, one may also speak of *theories* (of logical validity, form, concepts,...) instead of *systems*, though one should be aware that the analogy with scientific theories, which are supposed to describe, predict, and explain natural phenomena *as they are*, may be misleading in so far as logical theories are usually meant to *correct* and *improve* inferential practices.

This construction, investigation, and application of logical systems is not just carried out in philosophy but also in mathematics and computer science (and in linguistic semantics), which is why logic belongs to more than one academic subject. Which aspects of logical systems are emphasized, how the systems are dealt with, and what the aims and methodology are, differs

---

[2]In other semantics, "necessity" may not refer to the necessary preservation of *truth* but of something else that is relevant to rational reasoning and argumentation: *assertability*, *probability*, *existence of a winning strategy*,… In yet others, truth preservation may be weakened to *approximate* truth preservation, e.g. allowing for logical consequences to be only "tolerantly true" when the premises are "strictly true" (see Cobreros et al. 2012).

[3]There are logics the sets of logical truths of which are not closed under uniform substitution: e.g., substituting a formula for a propositional variable may turn a logically true formula of Public Announcement Logic into one that is not logically true. In such logics, propositional variables do not represent arbitrary sentences *tout court* but rather arbitrary sentences *with a certain syntactic form or subject matter*. See Holliday et al. (2011: Section 1.2).

from one subject to the next: the focus of mathematical logic, with its classical (maybe by now outdated) division into model theory, proof theory, recursion/computability theory, and set theory, is to prove theorems about logical systems of mathematical interest by mathematical methods. Logic in computer science elaborates the computational part of logic by (theorems about) formal languages and algorithms for model checking, program verification, automated deduction, logic programming, and logic in AI, and by studying their computational complexity. And logic in philosophy designs, investigates, and uses logical systems for philosophical purposes that differ from those of the purely formal disciplines. But boundaries are blurry, and logic in the *foundations* of mathematics/computer science may overlap with logic in philosophy: e.g., much of contemporary modal logic is located at the borderline between computer science and philosophy, as was Turing's classical work on the foundations of computation and computability. In what follows, I will concentrate on the philosopher's focus, way, aim, and methodology of doing logic: on logic in philosophy, that is, on logic as an *instrument in philosophy* (the application of logical methods in philosophy, Section 2) and on logic as a *philosophical discipline in its own right* (philosophical logic, Sections 3 and 4).

2. The First Role of Logic in Philosophy: Logic as an Instrument

Since its inception, logic has been playing two distinct roles in philosophy: first, as a propaedeutic "instrument" or "tool" (an *organon*) for philosophy, that is, a foundation and source of methods for rational reasoning and argumentation that may be put to use in one's philosophical work. Indeed, the Aristotelian tradition in late ancient philosophy did not view logic as an actual part of philosophy but rather as a *mere* such instrument. In contrast, the Stoics, who dealt with the logical analysis of language and rhetoric, logical concepts such as *if-then*, and logical paradoxes such as the Liar paradox and the Sorites paradox, regarded logic as one of the three main disciplines of philosophy itself (next to physics and ethics). Their understanding and work exemplify the second role of logic: as a proper philosophical discipline in its own right.[4] See Figure 1 for a summary.

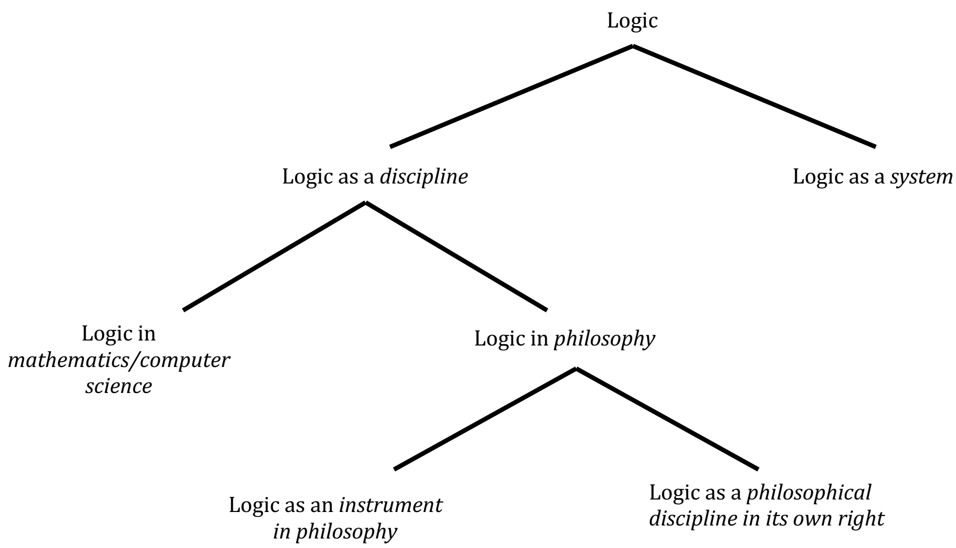

Figure 1: Logic – some terminological distinctions

[4] For more on the two roles of logic in ancient Greek philosophy, see Adamson (2014: Chapter 30) and Adamson (2015: Chapter 8).

Fast-forwarding 2000 years, we can still observe the same two roles that logic has been playing in philosophy. It is the service role that got emphasized in the Schlick quote from the beginning of this chapter, in which he expresses his excitement about the progress he believes the application of modern mathematical-logical methods will achieve in philosophy. While his optimism was exaggerated and the logical empiricists' idea of philosophy as logic of science has long since gone, the application of logical systems in philosophy since the days of Frege, Peirce, Russell, Tarski, and Carnap certainly amounts to a success story. Let me turn to the instrumental role of logic in philosophy in more detail now with the help of some examples.

The utility of logical methods in areas close to logic is perhaps not so surprising: this applies to the philosophy of language, with its logical analysis of definite descriptions, its formal models for truth, intension, conditionals, vagueness, and conversational common grounds, and its use of intuitionistic logic and/or proof-theoretic semantics in the study of anti-realist (constructive or inferentialist) conceptions of truth, meaning, and assertability. The same holds for the philosophy of mathematics, as witnessed by analyses of Gödel's Incompleteness Theorems (which themselves belong to proof theory) in provability/epistemic logic and applications thereof in the epistemology of mathematics, and by the essential role that higher-order logic plays in logicism, neo-logicism, and mathematical structuralism.[5]

For other parts of philosophy, the success of logical methods is less expected: as far as general epistemology is concerned, epistemic logic is exploited in the analysis of Fitch's Paradox of knowability, and belief revision theory, nonmonotonic logic, and dynamic epistemic logic are used to assess rationality postulates for belief change. Logic in general philosophy of science, which was of prime interest to the logical empiricists, includes the application of inductive logic and formal learning theory in the study of induction, the logical reconstruction of theories by Ramsey sentences or as sets of (perhaps partial) logical models, the study of criteria for the (e.g. definitional) equivalence of theories, and the theory of definitions as part of the methodology of science; while e.g. quantum logic can be of use in philosophy of physics. In metaethics, methods from deontic/nonmonotonic logic and formal semantics are employed to evaluate moral arguments, to determine the logical form of conditional defeasible norms, and to explore the Humean is-ought problem and moral expressivism. In practical philosophy, epistemic game theory utilizes epistemic logic and possible worlds semantics, and the logic of action is relevant to the philosophy of action. And so forth (modal logic in metaphilosophy,…).

Ironically, given the logical empiricists' aversion to the subject, metaphysics might be the philosophical discipline in which logical methods have been applied with the greatest success (for a survey, see Zalta 2011). Here is an incomplete list of metaphysical topics and the *logical methods* used in their study:[6]

---

[5] In recent years, the philosophy of mathematical practice has gained importance in philosophy of mathematics. While philosophical work on mathematical practice is not necessarily invoking logical methods, it is not necessarily excluding them either (see e.g. Mancosu 2017).

[6] Why this special success of logic in metaphysics? Here is a tentative explanation: perhaps the same logical systems that logicians construct, study, and apply by describing them metalinguistically *from outside* are constructed, studied, and applied by metaphysicians by describing the world object-linguistically *from within them*: where a logician would say $\mathcal{M} \vDash \forall P \forall x (P(x) \vee \neg P(x))$ and $\mathcal{M} \vDash \forall x \forall y (x = y \rightarrow N(x = y))$ about a second-order Kripke model $\mathcal{M}$, a metaphysician would say from inside of such a model that every property applies or does not apply to an object and that identity holds necessarily. From outside, the logical truths of a system reflect the formal structure of its class of models and how the logical constants have been defined to track that structure; from inside, the same logical truths appear as necessary metaphysical laws about the world.

- higher-order metaphysics (properties, relations, concepts, universals, propositions): *higher-order logic, lambda calculus, possible worlds semantics*
- existence, ontology, ontological commitment, quantification over everything: *first-order logic, free logic, formal theories of truth, higher-order logic*
- identity, criteria of identity, principle of identity of indiscernibles: *logic of identity, higher-order logic, theory of equivalence and congruence relations*
- modal metaphysics (necessity, possibility, impossibility, essence, actualism, possible worlds, situations, truthmakers, dispositions, laws): *modal logic, logic of counterfactuals, possible worlds semantics, situation semantics, truthmaker semantics*
- metaphysics of time (and space): *tense logic (and logic of space[7]), many-valued logic*
- mereology, parthood, composition: *theory of partial orders and Boolean algebras, algebraic logic*
- abstract objects, abstraction, abstraction principles: *higher-order logic, object theory[8]*
- nominalism: *higher-order logic, plural logic, modal logic*
- metaphysics of fiction: *modal logic, free logic, paraconsistent logic, impossible worlds semantics, object theory*
- metaphysics of logical objects, numbers, sets: *higher-order logic, object theory*
- functionalism about mental states: *computability theory*
- metaphysics of infinity, actual vs potential infinity: *set theory, non-standard analysis, modal logic*
- ontic vagueness and vague identity: *modal logic and possible worlds semantics*, *supervaluation semantics*, *many-valued logic*, higher-order logic
- supervenience, ontic dependence, grounding: *modal logic, truthmaker semantics*
- causality: *logic of counterfactuals*
- deflationism about ontology: *quantification, formal theories of truth*
- realism vs antirealism: *model theory, formal theories of truth*
- ontological arguments for the existence of God: *free logic, modal logic, higher-order logic, automated theorem-proving*[9]
- Buddhist metaphysics: *paraconsistent logic, relevance logic, many-valued logic*[10]

Last but not least, logic serves a pedagogical role in philosophical education, as most philosophy programs require their students to take logic courses[11] that should help them learn to speak and think more clearly and precisely, to question language and reasoning, to make hidden premises and logical forms of arguments explicit, to determine the logical (in-)validity of arguments, to derive conclusions from premises in a logical system, to structure philosophical theories deductively, to prove the (in-)consistency of such theories, to build and study formal models for such theories, and to understand limitative results about axiomatic reasoning. All these issues comprise the role of logic as an instrument in philosophy.[12]

Sections 3 and 4 will turn to logic as a philosophical discipline in its own right. It is constitutive of logic as a philosophical discipline to reflect on logical systems philosophically (otherwise it is just mathematical or computational logic); as such, it must talk *about* language and logical systems.[13] While Section 3 is devoted to two case studies — Tarski on consequence and truth

---

[7] See e.g. Aiello et al. (2007).

[8] In the sense of Zalta (1983).

[9] As in Benzmüller (2020).

[10] As in Priest (2018).

[11] The data collected by Attridge et al. (2016) suggest that logical education succeeds in teaching reasoning skills.

[12] Next to logical methods, other formal-mathematical methods have experienced great success in philosophy in recent decades, too, such as probability theory, game theory, and computer simulation. Accordingly, these methods have started to seep into philosophical education. So long as they are not meant to replace logical methods but to complement them, this will not threaten the pedagogical role of logic in philosophy.

[13] Since this is sometimes disputed (e.g. Williamson 2017), here are some typical instances of the kinds of *meta*statements one finds everywhere in philosophical logic: "$\mathcal{M} \vDash A$", "$A, A \to B \vdash B$", "$\frac{\Gamma, \neg A \Rightarrow B \wedge \neg B}{\Gamma \Rightarrow A}$ is not a rule of this calculus", "There are no logical truths in SKL", "M(odus) P(onens) is logically invalid in LP", "it is

— Section 4 will demonstrate that Tarski's work conforms to a methodology suggested by Rudolf Carnap and others, it will propose an extension and refinement of that methodology as a useful methodology for philosophical logic, and it will argue that the resulting *logical reconstructionism* about logic overlaps with, but also partially diverges from, contemporary anti-exceptionalist methodologies for logic. Along the way, I will draw conclusions on some topics that have been hotly debated in the philosophy of logic in the last couple of years, such as logical pluralism and the normativity of logic.[14]

3. Logic as a Philosophical Discipline in Its Own Right: Tarski as Case Study

Alfred Tarski, who was educated in the Polish Lvov-Warsaw school, was not just one of the leading mathematical logicians of the 20th century but also one of its preeminent philosophical logicians. It is the philosophical part of his work from which I have chosen two contributions: Tarski (1936, 1944). My main reasons for discussing them is that they are paradigmatic both of philosophical logic and philosophy of logic, and they are sufficiently short and non-technical to be addressed here. But one should keep in mind that they are supplemented by other works of Tarski, some of them collected in Tarski (1983a), which deal with same topics in much more formal/philosophical detail; e.g., contrary to Tarski (1944), Tarski (1983b) works out his definition of truth and its formal and philosophical consequences in great detail.[15]

In what follows, I will be as much interested in *what* Tarski does as in his methodology: *how* he proceeds and how he *justifies* that. I will summarize his central moves while re-structuring his discussion and adding some explanations to it. Page numbers refer to the English versions of his articles.

3.1 Case Study: "On the Concept of Logical Consequence" (Tarski 1936)

This paper derives from a lecture Tarski gave at the International Congress of Scientific Philosophy that members of the Vienna Circle organized in Paris in 1935. (The definition of logical consequence sketched in our Section 1 descends from this paper of Tarski's.)

Tarski starts by considering the pre-theoretic concept of consequence, which he says is only expressed vaguely and insufficiently clearly in everyday language, and usage of which "fluctuates" is subject to mutually contradictory "tendencies" (p. 409). His goal is to state a formal definition of logical consequence: any such definition will exhibit some more or less arbitrary features, reflecting some of the choices that need to be made when correcting the deficits of pre-theoretic logical consequence. For the same reason, the resulting precise notion will only approximate the pre-theoretic notion without capturing it exactly.

After sketching proof-theoretic accounts of logical consequence understood as derivability and finding them wanting, he prepares his own definition of logical consequence by discussing a precursor to it (Carnap's); later he will return to the existing literature when he compares

counterintuitive that MP is invalid", "MP should be valid", "the deductive system is incomplete relative to this semantics", "logical truth in… is undecidable", "natural language conditionals obey C(onditional) E(xcluded) M(iddle)", "conditionals in this system obey CEM", "identity is second-order definable", "the semantic value of this term is a set of sets of individuals", "truth is compositional", "Sentences of kind… should be represented as formulas of form…", "the belief operator in doxastic logic does not express a logical concept" etc.

[14] See Van Benthem (2007) for a different kind of survey of logic in philosophy.

[15] For more on Tarski's philosophical-logical work, see e.g. Patterson (2012).

his own definition with those of his forerunners. He wants to do better than them by describing a general method for constructing a definition of logical consequence that is not just applicable to formal languages of a very restricted type, as he thinks Carnap's method is. While he regards the basic ideas behind the definition as familiar, he states that it is only the methods developed recently for the purpose of a scientific semantics that make the formulation of an exact definition possible.

At next, Tarski enumerates some "intuitive" conditions that form the starting point of his definition (p. 414): first, logical consequence cannot lead from true premises to a false conclusion; secondly, logical consequence only depends on form and hence must be independent of empirical matters and invariant under the replacement of the designations of non-logical terms by other designations. One conceivable way of making this latter kind of invariance more precise would be to utilize syntactic substitutions, that is, to require that consequence be invariant under the replacement of any non-logical *term* in a formula by any other non-logical *term* of the same type: however, the strength of the resulting criterion of formality would very much depend on the expressiveness of the object language for which consequence is to be defined. For that reason, Tarski regards substitution invariance as a merely necessary condition on an adequately defined concept of logical consequence.

So far as the *definition* of consequence itself is concerned, he rather suggests to take the semantic character of replacing *designations* seriously by defining logical consequence in semantic or model-theoretic terms, based on the concepts of truth and satisfaction of sentences/formulas relative to sequences/models which Tarski defined elsewhere (and which were sketched in our Section 1):

> "The sentence X follows logically from the sentences of the class K if and only if every model of the class K is also a model of the sentence X" (p. 417).

As intended, this definition of logical consequence resembles the pre-theoretic understanding, and it can be shown to have the consequence that logical consequence is truth-preserving, substitution-invariant, and "independent of the richness of the language being investigated" (p. 417).

Tarski concludes by pointing out that his definition is open to extension and improvement and leads to further open questions.

3.2 Case Study: "The Semantic Conception of Truth and the Foundations of Semantics" (Tarski 1944)

This paper may be regarded as the less technical and more broadly accessible companion of the much more substantial and significant Tarski (1983b). The article is divided into the main systematic part and an additional "polemical" one in which he addresses potential worries about his systematic contribution. I will focus mainly on the former part while highlighting only a couple of his "polemical" remarks that are of methodological interest.

Tarski begins by considering the notion of truth as used in everyday language and as considered by historical sources, such as Aristotle. He observes that traditional attempts at defining truth have not been "sufficiently precise and clear" and that therefore "a more precise expression of our intuitions" (p. 343) is called for. Later in the paper he will claim that his definition does justice to these intuitions. Since the meaning of the common term "true" is only vaguely specified, the meaning of "true" according to an exact definition must deviate from the pre-theoretic understanding of "true". He mentions that he has no ambition to capture the "essence" of truth (p. 361), whatever that might be. Instead, he aims for his definition to resemble the common usage of the truth predicate, and he points to empirical-statistical work by which this could be judged and cites Arne Naess' empirical work[16] on that topic.

Other than lacking clear and precise definitions, semantic concepts such as truth are problematic also due to the semantic paradoxes, such as the Liar paradox which is analysed in some detail in the paper. He dismisses the option of avoiding the paradoxes by "changing our logic", due to its "consequences" (p. 349). Instead, he notes that the paradoxes only emerge for "semantically closed" languages in which semantic terms can be applied to names of sentences that include those very semantic terms again, and suggests to avoid the paradoxes by stating the definition of truth for languages that are not "semantically closed" (p. 349). This will still allow him to apply such definitions to a great variety of formal object languages that approximate fragments of natural or scientific language. The definition of truth for any such object language will be carried out in a metalanguage that extends the object language and which is "essentially richer" by containing "variables of a higher logical type than those of the object language" (p. 351). He adds that if one wanted to develop a theory of truth in a metalanguage that is not "essentially richer" than the object language, the truth predicate should rather be regarded a primitive term, and instead of giving a definition for it, one should constrain its meaning with the help of axioms involving it; in general, such an axiomatic approach may "prove useful for various purposes" (p. 352).

But for now his goal is: to give "a *satisfactory definition*" of truth, that is, "a definition which is *materially adequate* and *formally correct*" (p. 341), where formal correctness and material adequacy are meant to be conditions by which the satisfactoriness of a definition of truth can be determined.

What Tarski has in mind with formal correctness is that the formal structure and logic of the metalanguage in which the definition is formulated ought to be exhibited precisely, and that the definition should conform to the requirements that the general theory of definitions imposes on the logical forms of explicit definitions. Definitions like that do not only determine the meaning of their defined terms exactly and uniquely but can also be proven to have attractive logical features; in particular, the so-defined "semantic concepts will not involve us in any contradictions" (p. 351). By that he alludes to the fact that if any such definition is added to a consistent theory, the resulting package is provably consistent again. More generally, he thinks that introducing new concepts by definition can constitute great scientific progress.

Furthermore, "*we shall call a definition of truth 'adequate' if all these equivalences follow from it*" (p. 344), that is, if all equivalences of the same form as (assuming "snow is white" is a formalized sentence of the object language)

"The sentence 'snow is white' is true if, and only if, snow is white" (p. 343)

follow from the definition and the metatheory. As Tarski notes, any such definition will then be certain to get the extension of "true" right.

Finally, Tarski briefly sketches how his definition of truth is formulated: first one states a recursive definition of satisfaction for formulas by (sequences of) objects, which he has done elsewhere. Then one can define truth for the sentences of the object language by:

"*a sentence is true if it is satisfied by all* [sequences of] *objects, and false otherwise*" (p. 353).

[16] See Barnard and Ulatowski (2016) for more on this.

The resulting definition is formally correct, can be proven materially adequate, and it has various important consequences: in particular, well-known general laws of truth are derivable from it (and the metatheory), such as that for any two contradictory sentences of the object language exactly one of them is true. If applied appropriately, the definition can be used to prove the consistency of mathematical theories, and, for languages of sufficiently comprehensive formalized mathematical disciplines, one can demonstrate that truth outruns provability-in-a-recursively-axiomatized-system for any such discipline, as follows by combining Tarski's definition of truth with the Incompleteness Theorems. More generally, definitions of truth can serve as a foundation for semantics and have applications to all areas in which semantic concepts are of use, including science and the methodology of science.

He adds that there are other pre-theoretic concepts of truth, too, which should be clarified through definition or axiomatization, distinguished from each other syntactically, and to which his semantic concept of truth should be compared and related systematically.

4. A Methodology for Philosophical Logic

It should have become clear from Section 3 that each of Tarski's articles proceeds as follows: (i) it takes a pre-theoretic/informal concept *X* of logical interest (logical consequence, truth) that is not clear or precise enough and use of which may be prone to "contradictory tendencies" or plain inconsistency, (ii) it defines a clear and exact concept *X*' which, it is argued, resembles *X* extensionally, serves a similar function, and usage of which will be fruitful in its consequences, and (iii) proposes to replace *X* by *X*' for mathematical, semantic, methodological or philosophical purposes in the corresponding contexts.

This conforms almost exactly to what Carnap famously called "explication"[17]:

> "By an explication we understand the transformation of an inexact, prescientific concept, the explicandum, into an exact concept, the explicatum" (Carnap 1950: 1).
>
> "A concept must fulfil the following requirements in order to be an adequate explicatum for a given explicandum: (1) similarity to the explicandum, (2) exactness, (3) fruitfulness, (4) simplicity" (Carnap 1950: 5).

(Simplicity is not mentioned by Tarski but also only of secondary importance to Carnap and only relevant for the comparison of alternatives.) In fact, at least in the case of Tarski on *truth*, it is hardly an original insight that Tarski gave an explication in Carnap's sense.[18] More generally,

> "The task of making more exact a vague or not quite exact concept used in everyday life or in an earlier stage of scientific or logical development, or rather of replacing it by a newly constructed, more exact concept, belongs amongst the most important tasks of logical analysis and logical construction" (Carnap 1956: 7–8).

Much the same can be said about the many modern continuations of Tarski on logical consequence and on truth, which are explications or at least close to explications, too. E.g.,

[17] For further discussion and references, see Leitgeb and Carus (2022, Supp. D).

[18] Carnap (1950: 5, 1956: 8) himself presents Tarski on truth as an instance of explication; see also Leitgeb (2013). See Scharp (2013) for more on the general idea of conceptually engineering truth.

see Fig. 2 for a sketch of how formal theories of truth and semantic paradox unfolded after Tarski:[19]

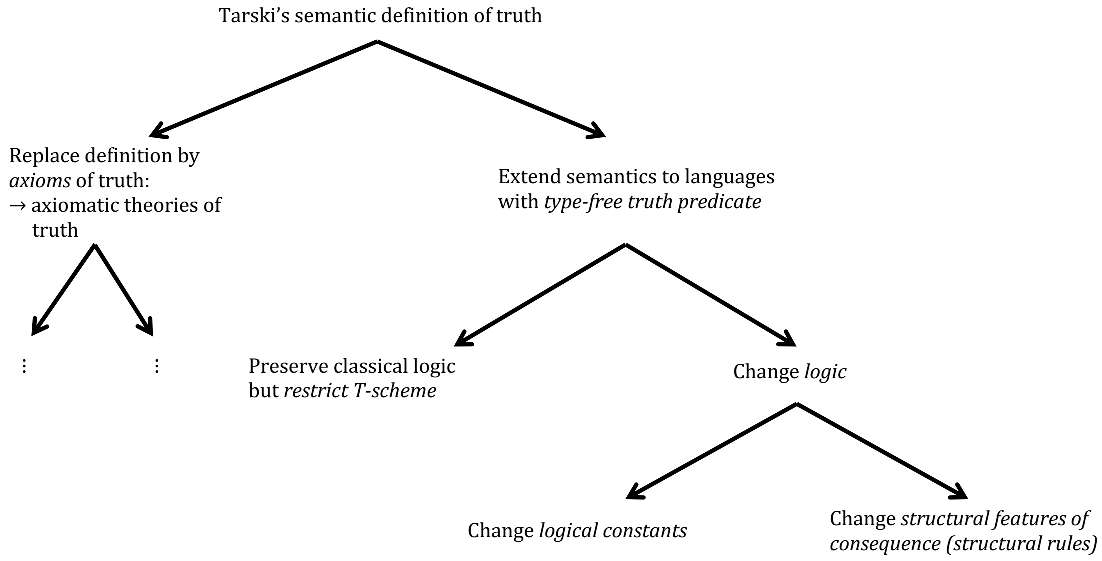

Figure 2: Formal theories of truth and paradox after Tarski

Without being able to support this here, I want to claim that the logical works to which this diagram alludes share the properties (ii) and (iii) from above. The only difference concerns (i): what these newer sources start from is not necessarily the pre-theoretic *X* that Tarski had considered, instead, it is often the very *X*′ defined by Tarski or his successors (or a combination of *X* and *X*′) that is meant to be improved by a new *X*′′. Since the "input" *X*′ as defined by Tarski is already sufficiently exact, desiderata *other than exactness* drive the transformation towards the improved "output" *X*′′. That is one reason why the methodology of explication needs to be generalized a bit in order to be turned into a general methodology for philosophical logic.

A second reason is: in contrast with explication, logical work is not always concerned just with *concepts*. E.g., the aim of Frege's *Begriffsschrift* was, quoting from its preface, to construct a "formula language for pure thought", which is a "device invented for certain scientific purposes", namely a notation and method by which *general patterns of reasoning* found at work in mathematics and science, independent of subject matter, could be clarified, sharpened, systematized, and generated. Thus, it would be off the mark to say that the point of the *Begriffsschrift* was to explicate some concept(s) (even when Frege's notation and rules illuminate logical concepts and determine a logical concept of derivability).

Thirdly, I want to leave open whether philosophical-logical undertakings are revisionary — as Tarski's from Section 3 — or whether they "merely" logically analyse an *X* of logical interest in a logical variant of conceptual analysis. If the latter, the intended improvement does not consist in the proposal to replace *X* by some (in some respects) better *X*′ but in the improved *understanding* of the logical structure of the original *X*. E.g., Russell's famous logical representation of definite descriptions in "On Denoting" has traditionally been regarded as such a logical analysis. (Though in fact it is not clear at all whether Russell is really just

[19]Note that explications are not necessarily carried out by stating a definition but may proceed axiomatically or by introducing formal rules or in some other manner.

"zooming into" the *already existing* logical structure of definite descriptions or whether he is actually proposing to supply them with a *new* such structure.)

So we need a term for something *like* explication that is not necessarily concerned with precisification, concepts, or revision. Conveniently, stretching its original meaning just a bit, such a term can be found in Carnap's early work (and others', e.g., Reichenbach's): *rational reconstruction*,[20] which does not necessarily involve precisification, applies more broadly (to concepts, reasoning processes, bodies of belief, methods,…), and may consist in just analysis. The "re-" refers to an *X* from which it starts, the "rational re-" to the intended output making what is rational about *X* more transparent and/or improving features of *X* that are relevant to rationality. This may involve imposing more formal precision on *X*, but only if *X* is imprecise in a way that affects its rationality features. "Construction" means the purpose-constrained but otherwise freely taking apart, studying, amending, producing or (re-) assembling representations of intellectual entities, including what the tradition named "logical construction", such as e.g. Frege-Russell's logical definition of number (terms).

Let us therefore call the replacement of *X* by *X'* as described previously a *logical reconstruction in the narrow sense*, and let a *logical reconstruction in the broad sense* be anything that is either a mere logical analysis or a logical reconstruction in the narrow sense (in which logical analysis is but the first step).[21]

Now that the terminology is sorted, I will sketch what, accordingly, may be called *logical reconstructionism*: a methodological proposal for how to carry out logical reconstructions in the broad sense within the area of philosophical logic. It extrapolates, amongst others, from Tarski's explications described in Section 3, as a comparison of Section 3 to the maxims below will confirm. It is not necessarily meant to govern each and every activity in philosophical logic but at least many paradigmatic ones; its aim is not to police logicians but to support them; and it not supposed to exclude any of the traditional topics of philosophical logic. If it constitutes one useful methodology for philosophical logic next to others and, at the same time, clarifies our methodological self-image just a bit, my mission will be more than accomplished.

These are the maxims of the *logical reconstructionist* proposal:

(1) *Logical analysis:* Logically analyse an *X* of logical interest. That is: (i) collect logically relevant examples or empirical studies (if there are such) of the usage or practice of *X*, consequences of that usage relevant to logic, and pre-theoretic logical "intuitions" (if there are such) about *X*; (ii) on that basis, describe *X*'s logical structure, taking into account the history of the logical study of *X*. *X* may be pre-theoretic or belong to a previous stage of logical inquiry, and it may be a concept, an inferential practice, an argument (pattern), a rule, a definition, a theory, a language (fragment), a method of logical representation or application, a semantics, a proof system,…

(2) *Problem determination:* Determine problems affecting *X*, based on (1). Such problems might be: unclarity; confusion; vagueness; lack of systematicity; inconsistency; lack of fruitfulness; narrowness; deductive weakness; expressive weakness; discontinuity with successful theories/areas; excessive theoretical or computational complexity; inelegance;... For different purposes *p* for which *X* might be used, determine which problems would become especially

[20] See Leitgeb and Carus (2022, Supp. D).

[21] I include also logical constructions (without "re-") of *completely novel X'*. But since these are rare in philosophy, I will put them to one side here.

detrimental if *X* were used for any such purpose. *"p"* might refer to a semantic purpose, a proof-theoretic purpose, a methodological purpose, a purpose of philosophical or scientific application,… or to the "sum" of all such purposes taken together (logic as an "all purpose" device[22]).

(1)(i) suggests to begin a logical analysis by collecting "(quasi-)data" about *X*. While the methodological role of these "data" has much in common with that of empirical data in science[23], there is also an important difference. If scientific data are known to have been collected correctly and undistortedly, they are scientifically "sacrosanct": while abstraction, idealization, and extrapolation may be applied in the transition from these data to hypotheses and models, every deviation from the data should evoke bad conscience, as the task of science is to describe, predict, and explain what the world is like, and data are meant to capture what it is like. In contrast, even when the actual (performance or competence) usage of, e.g., the truth predicate is known to have been recorded correctly and undistortedly, that usage itself may still be confused or inconsistent or… (see (2)), which is why Tarski aimed *to improve the existing usage*. Similarly, logical "intuitions" merely serve as nonreflective starting points and can and should be just as much revised by systematic rigorous theorizing as initial mathematical intuitions ("all infinite collections are of the same size") or physical intuitions ("heavier objects always fall more quickly"). Strawson's (1963) criticism of Carnapian revisionary explication as changing the original subject matter misses the mark: e.g., the point of Tarski's material adequacy condition was to ensure there was enough resemblance between the ordinary concept of truth and Tarskian truth, and indeed no modern logician would think that Tarski's definition did not concern *truth* anymore. And merely diagnosing persistent inconsistent usage without "treating" it would simply be irresponsible. Finally, externalist worries whether any such "treatment" by conceptual engineering is possible at all, due to our lack of control over our representational devices (Cappelen 2018), lose their bite in the abstract realm of logic and mathematics.[24] E.g., I take it to be clear that Tarski *did* have control over the concept of *logical consequence* as used by subsequent logicians/philosophers: after all, whole generations of logicians followed his lead.[25]

(3) *Transition to rejection or to narrow logical reconstruction:* If no serious problems are found in (2), describe the results of (1) and stop. If there are problems in (2) that are too severe to be treated, explain why that is, propose to reject/dismiss *X*, and stop. Else, continue with (4).

(4) *Logical reconstruction in the narrow sense:* Choose one (or several) of the purposes *p* for which *X* might be used and for which *X* was found to be problematic in (2). Propose, by definition or axioms or rules…, a logical replacement/improvement *X*' of *X* for purpose *p* and argue that (i) *X*' resembles *X* in function and use concerning *p* (for this, use (1), and, if useful and available,

[22]But note that even two "all-purpose" logical systems may still be such that one of them is better suited, say, for semantic purposes while the other one does better for proof-theoretic purposes.

[23]See Martin and Hjortland (forthcoming) for more on evidence in logic.

[24]Compare mathematics: would anyone think that mathematicians do not have control over the notion of continuous function, such that the term "continuous" would still refer to the ancient *can be drawn without lifting the pen from the paper* rather than the modern *pointwise* ε-δ*-continuity* or *uniform continuity* or *continuity in a topological space*?

[25]What is going on here is that logical constants do *not* refer to worldly objects to whose baptisms the usage of logical constants is causally linked, or the like. Rather, logical constants are *merely expressive devices*: they help us express propositions by structuring thought without contributing worldly reference. Different logical systems may structure thought differently, but however the world maps to that structure, the corresponding logical truths will be satisfied. The obvious way of determining the meaning of such logical constants in logical systems is by laying down inferential or semantic rules (or both), and such rules can be altered by non-externalist means.

empirical methods) and that (ii) *X'* does better than *X* concerning problems in (2) relevant to *p* and/or concerning desiderata relevant to *p* (see (5) below).

The dismissal of a logically flawed *X* mentioned in (3) may well happen: e.g., some argument might be too bad to be even approximately saved, and there are cases of "dialectical logics" that simply cannot be rescued from logical flames. As stated in (1) and (4)(i), empirical science and experimental philosophy may be useful in logic, too, though their roles will be restricted.

(5) *Desiderata for narrow logical reconstruction:* Understand "does better… concerning desiderata relevant to *p*" in (4)(ii) as referring to the greater satisfaction of (perhaps weighted) *p*-relevant defeasible desiderata amongst: being clearer; more precise; more systematic; less threatened by inconsistency; more fruitful in consequences or applications; more general or unifying; deductively stronger; more expressive; more continuous with other successful theories or areas; theoretically simpler or of less computational complexity; more elegant;...

(6) *No facts of the matter:* Do not presuppose, without further argument, that there must be facts of the matter that would make the definition or axioms or rules or… for *X'* in (4) true or false.

See, e.g., Leitgeb (2007), Halbach and Horsten (2015), and Terzian (2016) for specific norms for formal theories of truth and paradox, some of which derive from more general desiderata as in (5). These desiderata may work against each other: e.g., there may be *multiple pairwise distinct* maximal satisfiable sets of such desiderata on truth (cf. Leitgeb 2007). Or take expressive vs deductive power: a more expressive logic might drive a wedge between formulas *A* and *B* by not proving their equivalence, while *A* and *B* might become logically indistinguishable in a deductively stronger theory that proves their equivalence. If, e.g., the task is to offer a system of deontic logic as a common logical background for arguments between proponents of different moral systems, a weak logic offering fine-grained deontic-modal distinctions and minimal deontic commitments might be the preferred choice. Not so, if one merely intends to develop one's own moral position with maximal logical strength and convenience.[26] Or take intuitionistic logic: if compared to classical logic "naively" via the identity map on formulas, the system of intuitionistic logic is weaker than that of classical logic. But for the same reason it can be extended by principles (e.g. continuity axioms of intuitionistic mathematics) that would be inconsistent with classical logic. Hence, for the purpose of developing mathematics on constructivist foundations, the weaker system is preferable, while classical logic is of course the default choice for classical mathematics.[27]

Clearly, the desiderata for logical-reconstruction choice resemble those of recent anti-exceptionalist methodologies for logic which liken theory-choice in logic to abductive theory-choice or inference-to-the-best-explanation in science.[28]

But there is also a potential *dissimilarity*: even if logical "data" had the same status as scientific data (which, recall, they don't), the "abductive" transition from them to logical reconstructions may differ from that to scientific hypotheses or models. In the scientific case, data may underdetermine theories, and abduction is meant to fill that epistemic gap, but *the default is to assume that there is some (at that point unknown) fact of the matter* whether

[26] Shouldn't one always prefer the stronger system that derives more (logical) truths? I will turn to this question briefly in footnote 58.

[27] See Shapiro (2014) for more on how pluralism about logic relates to pluralism about foundations of mathematics.

[28] See Hjortland (2017) and Martin and Hjortland (2022), who also discuss relevant work by Maddy, Priest, Williamson, Sher, Russell, and others.

one's abductive conclusion is true. While (6) suggests *not* to make that default assumption for logical reconstructions in the narrow sense: first of all, *X*' may not be introduced by truth-apt linguistic expressions at all (but, e.g., by *rules*). Secondly, even if introduced by declarative sentences, the introduction of *X*' may be more like the stipulative definition of a new term for which one would not like to say the definition "describes a fact of the matter" — except perhaps *post hoc*, if and when the definition may have "ossified" to an extent that has turned it into something like a fact. (E.g., Tarski's definition of logical consequence as truth-preservation over all classical interpretations with varying domains has almost become such a "fact".) While the quasi-data from which logical reconstruction starts in (1) surely have an anchoring effect on what the outcome of (4) may be like, they can be overridden, and they may be vague: just as all usage facts concerning a vague predicate, taken together with all non-linguistic facts, may not determine (*pace* epistemicism) which member of a set of precise admissible interpretations is "the objectively right one", all potential "data" concerning *X* may not determine whether a logical reconstruction of *X* is "the objectively right one" either.[29] The emphasis is on "may": the proposal is not to presuppose that there are no such objective correctness facts for *X*' but just *not to presuppose there are*. It would seem risky at best, and utterly wrong at worst, to do otherwise: for how likely is it that there are *facts* by which, say, logical notions are not characterised by invariance under permutation (as suggested by Tarski 1986) but "really" under isomorphism between two domains or partial isomorphism or potential isomorphism or homomorphism…[30] or characterised proof-theoretically etc.? Or how likely is it that are there facts by which one branch in Fig. 1 corresponds to the "actual" solution to the Liar paradox? There does not seem any indication from the practice of philosophical logic that there are such facts, nor are there good arguments for their existence from other areas (e.g. metaphysics).[31] Which does not mean that one could rule out their existence completely — as one should not have ruled out in the past either that some areas called "philosophical" would become sciences later — just that their existence is unlikely and that it is therefore wise to remain cautious.[32]

Clearly, this should not stop logic from being *scientific* (as Tarski 1986, p. 145 urged) in the sense of applying scientific methods and deriving truths: e.g. proving mathematical theorems about a (classical, intuitionistic, relevantist,…) logical system *L* with a set of logical truths in its object language, deriving metalinguistically that the logical truths of *L* are true-in-*L*[33], such that everyone who would decide to reason according to the system would become committed to them, and more. And none of this rules out *logical progress* either: it is just that, other than their resemblance and problem-solving parts, logical reconstructions may not necessarily always improve in the epistemic sense in which scientific theories get closer to the truth but perhaps "merely" in the pragmatic sense in which tools become more effective, more widely

[29]This kind of semantic indeterminacy is compatible with classical logic (classical logical laws, rules, and metarules) and truth-conditional semantics: it is just that *classical truth may outrun the facts* in the presence of semantic indeterminacy (see Leitgeb, forthcoming).

[30]See Feferman (2010) for a survey, including references to Sher's and McGee's seminal work done on this.

[31]One might try to establish an analogy between a "Platonic" realism about logic with the more familiar "Platonic" realism about mathematics. But the analogy would be less than perfect: every model-theoretically inclined logician would grant that the logical constants of a logical system are defined to express features of some given structures. However, the actual question from the viewpoint of the logician is *which* such structures, or which components thereof, are expressed by the logical constants, and there might well be no fact of the matter which such (components of) structures are "the right ones".

[32]And if it turns out there are such facts, the best place for describing, predicting, and explaining the factual phenomena underlying *X* might not be philosophical logic anymore but, perhaps, linguistics or some other area.

[33]But see Field (2009: Section 5.2) for an argument that not all logical truths are truths, assuming the presence of certain Curry-type self-referential sentences.

usable, simpler to handle,… (without any "guiding to the truth" connotations).[34] Indeed, a lack of relevant underlying facts of the matter would explain why the development of philosophical logic (part of which is depicted by Fig. 2) does not exhibit the same theoretical convergence one expects of the natural sciences: the reason might be that there is not enough "logical reality" to rule out logical reconstructions. The upshot is: if (4) includes instances of "abduction", they might be merely pragmatic rather than properly epistemic inferences-to-the-best-explanation of existing phenomena; it is the similarity of the theoretical virtues involved that makes it easy to confuse them. To what extent this amounts to a difference between logical reconstructionism and contemporary anti-exceptionalism about logic depends, ultimately, on how firmly anti-exceptionalists want to draw the analogy between logical reconstruction and scientific inference to the best explanation.

(7) *Transparency:* Be transparent about the purpose of your reconstruction, the choices you made when you constructed *X*', and the parts of your metatheory (e.g. second-order logic, a fragment of set theory, a theory of syntax,…) that were presupposed by that construction.

(8) *Preparation for alternatives and continuations:* List potential worries about *X*' and the outcomes of (7) as possible grounds for the logical reconstruction of *X* by an alternative *X*'' and/or the further improvement of your own *X*' by some *X*''.

Amongst others, (8) motivates philosophical logicians to exploit, in their logical reconstructions, the great plurality of logical systems that are available or constructable: for logical reconstruction is open-ended, both synchronically (by rival simultaneous transitions from *X* to *X*' and from *X* to *X*'') and diachronically (by subsequent transitions from *X* to *X*' and from *X*' to *X*''). And because it is unlikely that there are facts of the matter by which definitions or axioms for logical reconstructions are made true or false, it is likely that there are multiple ways in which (4) can be satisfied even for one and the same purpose *p*, corresponding to the many paths one may choose towards the goal of satisfying *p*. Compare the task of building a bridge over a gorge: various descriptive-factual matters will enter into the considerations, such as the spatial extension of the gorge and the laws of physics. But this will still leave the builders with lots of *free choices* concerning the type of the bridge, its shape and material, its structure and parts, and so forth. The same holds for choices of logical representation, formal language, logical constants, semantics, deductive system, and the like: a general logical pluralism.

In particular, this holds when *X* is pre-theoretic/informal *consequence*, which, if Tarski is right, is unclear, vague, and subject to mutually "contradictory tendencies" (recall Section 3.1), and which therefore leaves ample space for being clarified, sharpened, and made consistent, even for one and the same purpose, in distinct ways that all resemble (more or less) ordinary usage and satisfy (more or less) the desiderata in (5). This amounts to a Beall-Restall-type of *logical pluralism*:

> "*Logical pluralism* is the claim that at least two different instances of GTT provide admissible precisifications of logical consequence" (Beall and Restall 2006: 29)

where GTT is the "Generalised Tarski Thesis", according to which logical consequence is given by truth preservation in all "cases", and where "case" can be precisified classically, intuitionistically, relevantistically,…. This matches Carnap in *Meaning and Necessity* in which

[34] Dunn (2016: xxxii) expresses such an engineering ethos, when he describes himself "more as an engineer of logics – a maker of tools".

he presents his semantic definition of logical consequence as resulting from the explication of customary imprecise consequence[35], which, just as every other explication, is governed by similarity, exactness, fruitfulness, and simplicity, but not by objective facts of correctness,[36] and which allows for one and the same *X* to have more than one explication[37]. Except for GTT, which Carnap did not state but would have agreed with, Beall&Restall's logical pluralism just *is* Carnap's logical pluralism in *Meaning and Necessity* (although Carnap also allows for corrections of "data" while Beall&Restall only mention the indeterminacy left by the "data").

When Beall and Restall (2006: Section 7.3) say their logical pluralism differs from Carnap's, they mean the additional *pluralism about languages* that was proposed in his *Logical Syntax of Language*, which can be defended or rejected independently. But another plausible pluralism it is: Carnaps regards the meaning of logical constants in logical systems to be given by their syntactic rules of inference and/or, in his later work, by their semantic rules in these systems. Thus, when rules of inference and/or semantic rules differ in relevant respects from one logical system to the next, the meanings of the logical constants in the respective logical systems will differ, too.[38] And this will be so even when the object languages of two such systems are syntactically identical, when the two systems are defined in one and the same interpreted metalanguage, and when, e.g., the "→" in one (say, classical) system and the "→" in another (say, intuitionistic) system are used to logically represent the very same if-then sentences from natural or mathematical language. I suppose that many contemporary philosophical logicians would agree with Carnap on these points at least for certain logical systems and hold that, e.g., classical and intuitionistic logical systems assign different interpretations to "→"; putting my own cards on the table, I certainly do so. (Note that this is a claim about *constructed* logical systems, not natural language, which is why the worries in Quine 1951 about analyticity/synonymy for natural language are irrelevant.)[39] In any case:

[35]"the concept of L-truth thus defined… is an adequate explicatum for logical truth. […] *L-implication* is meant as explicatum for logical implication" (Carnap 1956: 11).

[36]"In a problem of explication the datum, viz., the explicandum, is not given in exact terms… if a solution for a problem of explication is proposed, we cannot decide in an exact way whether it is right or wrong. Strictly speaking, the question whether the solution is right or wrong makes no good sense because there is no clear-cut answer" (Carnap 1950: 3–4).

[37][About Frege's and Russell's explication of number terms:] "other logicians have proposed other explicata for the same explicandum" (Carnap 1956: 8).

[38]See Leitgeb and Carus (2022: Supp. G and H) for more on Carnap's logical inferentialism.

[39]Recent work in the philosophy of logic has been interested in the so-called "collapse" worry about logical pluralism according to which the epistemic goal of truth might favour choosing deductively stronger logical systems for the simple reason that they derive more truths from a given set of premises (see e.g. Steinberger 2019a). From the viewpoint of logical reconstructionism, the choice of logical systems may lack underlying facts of the matter, which is why epistemic considerations may simply run dry. And if one adds the view that, at least for certain logical systems, the meaning of the logical constants in these systems is determined by rules of inference or semantic rules (or both) of these systems, the comparison between such systems in terms of the "naive" identity map is a complete non-starter: for a formula *A* in one such system may not mean the same as *A* does in another such system. Indeed, logicians often do *not* compare classical and intuitionistic logic via the identity map but relative to a more complex double-negation or modal translation for precisely the reason that doing otherwise would be like comparing apples with oranges. Or reconsider the examples I gave in my discussion of the desiderata in (5) potentially working against each other: a weaker system of deontic logic might successfully underdetermine the meaning of the *O*(ught)-operator to the effect that different moral philosophers may regard their own distinct and more contentful conceptions of "ought" as being compatible with it. (Much like conditional definitions in mathematics are compatible with different kinds of strengthening.) Accordingly, a stronger system of deontic logic might not derive more truths for a pre-determined maximally specific *O* but just constrain the meaning of a merely partially interpreted *O* more severely.

logical reconstructionism urges philosophical logicians to explore *whatever* kind of logical pluralism there is.[40]

This completes the maxims of *logical reconstructionism*. Other than the fact that they resulted by extrapolating from the methodology that governed the two successful Tarskian paradigm case instances of work in philosophical logic from Section 3, one may expect these maxims to be justified pragmatically by promoting the successful development of philosophical logic in the future. If that expectation should turn out to be false, *logical reconstructionism* ought to be abandoned.

Let me conclude by sketching what this logical reconstructionist proposal has to say about the normativity of logic.

For a start, (1)-(8) are methodological maxims that may themselves be regarded as norms. But they are norms specifically for *logical work done by logicians*, and their normativity derives from the general normativity of methodology in the same way in which, e.g., the normativity of maxims for how scientists ought to conduct scientific experiments does. But what does logical reconstructionism have to say about the traditional debate about the normativity of logic for *reasoning by human agents*, that is, about whether logic "instructs us about how we ought or ought not to think or reason" (Steinberger 2022)?[41]

So far as the possible outputs of complying with the logical reconstructionist proposal are concerned, if adhering to (1)-(4) does not just result in the description of what the logical analysis of *X* has brought to light but in the verdict that *either X ought to be rejected or X ought to be logically reconstructed in the narrow sense for a certain purpose*, then that outcome is normative again. And since there is not necessarily a fact of the matter which such logical reconstruction is "the objectively correct one", this normativity does not necessarily reduce to merely descriptive-factual matters in the same manner in which, ultimately, the normativity of scientific theory-choice does. How the phrase "logical replacement/improvement *X*′ of *X* for purpose *p*" in (4) is to be understood exactly, depends on *X* and *p*: e.g., if *X* is already the outcome of a logician's construction and *p* is an academic purpose in logic, then the definition, axioms,… for *X*′ are simply recommended to replace those of *X* for purpose *p* in that logical work; and the corresponding instrumental norm has the simple and precise form "it ought to be that the definition, axioms,… for *X*′ replace(s) the definition, axioms,… of *X* for purpose *p*". E.g., for truth-theoretic and paradox-related purposes, Kripke (1975) may be understood as recommending his three-valued fixed-point definition(s) of the grounded extension of a type-free truth predicate to replace Tarski's original definition of a typed truth predicate for an object language in an essentially stronger metalanguage. And, in logical research about formal theories of truth and paradox, it mostly did.

However, if *X* is pre-theoretic and informal, and when *p* is an everyday purpose or one that concerns informal philosophical work outside of logic, the "normative power"[42] of *X*′ should be thought of differently: in such cases, *X*′ will rather serve as a *normative paradigm* (role model, pattern to be copied) that is subject to a *vague* norm of the form "*the original logical use/practice concerning X ought to be "reorientated towards" X′ for purpose p*". What I mean by "normative paradigm" can be understood by analogy to Newton's mechanics being a paradigm (in the sense of Kuhn's exemplars) for 18th and 19th century physicists, or a person being a moral paradigm for those who try to emulate that person's moral behaviour, or Rafael

[40]See Matti Eklund's contribution to this handbook for more on different versions of logical pluralism.

[41]See Filippo Ferrari's and Ulf Hlobil's joint contribution to this handbook for more on the normativity of logic.

[42]Whether concerning directives, evaluations or appraisals: see Steinberger (2019b).

Nadal and his groundstrokes and professional attitude serving as a role model for tennis players. E.g., even non-logicians may follow Tarski's definition of truth by accepting 'snow is white' as true just in case snow is white, regarding truth as compositional, expecting the truth predicate to add to the expressiveness of language, paying attention to the object/ metalanguage distinction when talking about truth, and so on. When they do so, they orientate their pre-theoretic and informal use of the truth predicate to Tarski's formal theory in an approximate and multi-faceted manner. The corresponding vague "reorientation" norm is thus already satisfied whenever one makes one's use of the truth predicate *(more) similar* to Tarski's in certain respects and contexts. Carnap had urged similarity to hold between an informal explicandum *X* and its more precise explicatum *X*', but now we find that it should also hold between a precise *X*' and the new *informal practice* to which *X*' should give rise when *p* is an everyday informal purpose. Accordingly, the precise logical consequence relation of, say, classical logic may be said to inform us how we ought or not ought to reason whenever we make our reasoning practices *(more) similar* to it in the same approximate and multi-faceted ways in which teaching logic serves its pedagogical role in philosophical education (recall Section 2).[43]

This latter normativity of logical reconstructionism does not so much derive from the "logical" but the "reconstructionism" aspect thereof and would be shared, e.g., by rational reconstructions in mathematics or science. Is there anything else that makes rational reconstruction in *logic* normative in a way in which rational reconstruction in mathematics or science is not?

I think there is, which brings me to my final, and more speculative, point. Assume human rationality to have two sources: on the one hand, truth in virtue of empirical facts, and, on the other, what we ourselves bring to the table independently of empirical matters of fact. The latter crucially involves logical components: for logical concepts play a constitutive role in the rationality postulates that govern the coherence of degrees of belief independently of what these beliefs are about,[44] and the necessity and formality of logic secure its independence of empirical matters of fact.[45] Call *reason* the part of human rationality that rationally reflects on, criticizes, and improves *the rationality of what we bring to the table*. If logical reconstructions have their purposes set to such world-independent "rationality-engineering" ones, their normativity will be practical but not "merely" instrumental: for they will contribute

[43]Thinking of logical consequence as a normative paradigm governed by a vague reorientation norm might explain why it is so difficult to formulate a comprehensive set of strict, simple, and universal bridge norms between logic and epistemology, as investigated by MacFarlane, Steinberger, and others (cf. Steinberger 2019b). The usual Harman-type worries about the existence of such norms are not much of a problem once the coordination between logical consequence and reasoning in everyday contexts is not claimed to be strict, simple or universal. (Which also takes pressure off the closely related so-called "logical omniscience problem".) Perhaps the normativity of logic in such contexts just amounts to this: *you ought to make your reasoning (more) similar to logical consequence in the respects that are important for your purposes.*

[44]If degree-of-belief assignments P are assumed to apply to sentences, the classical axioms of subjective probability have it that (i) $P(A) = 1$ for all *logically true A* (where 1 is the maximal degree), and (ii) $P(A \vee B) = P(A) + P(B)$ for all A and B that *logically exclude each other*. If degrees of belief are assigned to propositions, analogous axioms holds for the *logical structure (Boolean algebra)* of propositions.

[45]Putnam famously argued (taking this back later) that quantum logic was true in virtue of empirical fact and that "Logic is as empirical as geometry" (Putnam 1968: 184). However, quantum theory can be developed based on classical logic, too, and there are good reasons to think that *any* empirical understanding of quantum logic is bound to fail: "Since space and time are not just in our minds, our theories of space-time structure can get negative feedback from experience. But logical structure is not like space-time structure: there is no physical counterpart. As a result, physics cannot be simplified or improved by changing logic" (Maudlin 2022: 205).

to reason's business of setting the standards of rationality themselves[46] (while rational reconstructions in mathematics and science only serve their specifically mathematical or scientific purposes). Every such logical reconstruction will then amount to a proposal for how to become *more rational independently of what the empirical facts of the matter* are like — a constitutive-normative logical endeavour that seems genuinely philosophical.[47]

Acknowledgements: I am very grateful to André Carus, Filippo Ferrari, Norbert Gratzl, Levin Hornischer, Alessandra Marra, Paul Weingartner, and two anonymous reviewers for their comments on an earlier version of this chapter.

References:

Adamson, P. (2014), *Classical Philosophy. A History of Philosophy Without Any Gaps* (Oxford University Press).

Adamson, P. (2015), *Philosophy in the Hellenistic & Roman Worlds. A History of Philosophy Without Any Gaps* (Oxford University Press).

Aiello, M., Pratt-Hartmann, I., and van Benthem, J. (2007), *Handbook of Spatial Logics* (Springer).

Attridge, N., Aberdein, A., and Inglis, M. (2016), "Does Studying Logic Improve Logical Reasoning?", in C. Csíkos, A. Rausch, and J. Szitányi (eds.), *Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 2, PME, 27–34.

Barnard, R., and Ulatowski, J. (2016), "Tarski's 1944 Polemical Remarks and Naess' 'Experimental Philosophy'", in *Erkenntnis* 81: 457–77.

Beall, JC, and Restall, G. (2006), *Logical Pluralism* (Oxford University Press).

Van Benthem, J. (2007), "Logic in Philosophy", in D. Jacquette (ed.), *Philosophy of Logic* (Elsevier), 65–99.

Benzmüller, C. (2020), "A (Simplified) Supreme Being Necessarily Exists, says the Computer: Computationally Explored Variants of Gödel's Ontological Argument", in D. Calvanese, E. Erdem, and M. Thielscher (eds.), *Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning*, KR 2020, IJCAI, 779–89.

Cappelen, H. (2018), *Fixing Language*. *An Essay on Conceptual Engineering* (Oxford University Press).

Carnap, R. (1950), *Logical Foundations of Probability* (University of Chicago Press).

Carnap, R. (1956), *Meaning and Necessity: A Study in Semantics and Modal Logic*. Second edition (University of Chicago Press).

Cobreros, P., Egré, P., Ripley, D., and van Rooij, R. (2012), "Tolerant, Classical, Strict", in *Journal of Philosophical Logic* 41: 347–85.

Dunn, J.M. (2016), "An Engineer in Philosopher's Clothing", in K. Bimbó (ed.), *J. Michael Dunn on Information Based Logics* (Springer), xvii–xxxviii.

Dutilh Novaes, C. (2012), *Formal Languages in Logic* (Cambridge University Press).

Feferman, S. (2010), "Set-Theoretical Invariance Criteria for Logicality", in *Notre Dame Journal of Formal Logic* 51/1: 3–20.

Field, H. (2009), "Pluralism in Logic", in *The Review of Symbolic Logic* 2/2: 342–59.

Halbach, V., and Horsten, L. (2015), "Norms for Reflexive Theories of Truth", in D. Achourioti et al. (eds.), *Logic, Epistemology, and the Unity of Science*, Vol. 30 (Springer), 263–80.

Hjortland, O.T. (2016), "Anti-Exceptionalism about Logic", Philosophical Studies 174: 631–58.

Holliday, W.H., Hoshi, T., and Icard, T.F. (2011), "Schematic Validity in Dynamic Epistemic Logic: Decidability", in H. van Ditmarsch, J. Land, and S. Ju (eds.), LORI 2011, *Lecture Notes in Computer Science* 6953: 87–96.

[46]None of this conflicts with logical pluralism: necessity and formality apply to non-classical logics, too, and rationality postulates for degrees of belief can be based on non-classical logics as well (see Williams 2016).

[47]See Steinberger (2017) for more on such a Carnapian take on constitutive logical norms.

Kripke, S. (1975), “Outline of a Theory of Truth”, in *Journal of Philosophy* 72/19: 690–716.
Leitgeb, H. (2007), “What Theories of Truth Should Be Like (But Cannot Be)”, in Blackwell *Philosophy Compass* 2/2, 276–90.
Leitgeb, H. (2013), “Scientific Philosophy, Mathematical Philosophy, and All That”, in *Metaphilosophy* 44/3: 267–75.
Leitgeb, H. (forthcoming), “Ramsification and Semantic Indeterminacy”, in *The Review of Symbolic Logic*.
Leitgeb, H., and Carus, A. (2022), "Rudolf Carnap", in E.N. Zalta and U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), <https://plato.stanford.edu/archives/fall2022/entries/carnap/>.
Mancosu, P. (2017), *Abstraction and Infinity* (Oxford University Press).
Martin, B., and Hjortland, O.T. (2022), ”Anti-Exceptionalism about Logic as Tradition Rejection”, in *Synthese* 200/2: 1–33.
Martin, B., and Hjortland, O.T. (forthcoming), “Evidence in Logic”, in M. Lasonen-Aarnio and C.M. Littlejohn (eds.), *Routledge Handbook of the Philosophy of Evidence* (Routledge).
Maudlin, T. (2022), “The Labyrinth of Quantum Logic”, in S. Chakraborty and J.F. Conant (eds.), *Engaging Putnam* (De Gruyter), 183–206.
Quine, W.V.O. (1951), “Two Dogmas of Empiricism”, *The Philosophical Review* 60/1: 20–43.
Patterson, D. (2012), *Alfred Tarski: Philosophy of Language and Logic* (Palgrave Macmillan).
Priest, G. (2018), *The Fifth Corner of Four: An Essay on Buddhist Metaphysics and the Catuskoti* (Oxford University Press).
Scharp, K. (2013), *Replacing Truth* (Oxford University Press).
Schlick, M. (1930), “Die Wende der Philosophie”, in *Erkenntnis* 1 (1930/31): 4–11. Translated as “The Turning Point in Philosophy” by D. Rynin, in M. Schlick, *Philosophical Papers*, Vol. II, ed. by H. L. Mulder and B. F. van de Velde-Schlick (D. Reidel) 1979, 154–60.
Shapiro, S. (1998), “Logical Consequence: Models and Modality”, in M. Schirn (ed.), *Philosophy of Mathematics Today* (Oxford University Press), 131–56.
Shapiro, S. (2014), *Varieties of Logic* (Oxford University Press).
Sher, G. (1996), “Did Tarski Commit ‘Tarski’s Fallacy’”, in *The Journal of Symbolic Logic* 61/2: 653–86.
Steinberger, F. (2017), “Frege and Carnap on the Normativity of Logic”, in *Synthese* 194/1: 143–62.
Steinberger, F. (2019a), “Logical Pluralism and Logical Normativity”, in *Philosophers’ Imprint* 19/12: 1–19.
Steinberger, F. (2019b), “Three Ways in Which Logic Might Be Normative”, in *Journal of Philosophy* 116: 5-31.
Steinberger, F. (2022), “The Normative Status of Logic”, *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition), in E.N. Zalta and Uri Nodelman (eds.), <https://plato.stanford.edu/archives/win2022/entries/logic-normative/>.
Strawson, P.F. (1963), “Carnap’s Views on Constructed Systems versus Natural Languages in Analytical Philosophy”, in P.A. Schilpp (ed.), *The Philosophy of Rudolf Carnap* (Open Court), 503–18.
Tarski, A. (1936), “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), 1–11. Translated as “On the Concept of Logical Consequence” by J.H. Woodger, in Tarski (1983), 409–20. The Polish version, “O pojęciu wynikania logicznego”, *Przegląd Filozoficzny* (1936) 39: 58–68, was translated as “On the Concept of Following Logically” by M. Stroińska and D. Hitchcock, *History and Philosophy of Logic* 23: 155–96.
Tarski, A. (1944), “The Semantic Conception of Truth: And the Foundations of Semantics”, in *Philosophy and Phenomenological Research* 4: 341–76.
Tarski, A. (1983a), *Logic, Semantics, Metamathematics*: *Papers from 1923 to 1938*, 2nd edition, edited by J. Corcoran (Hackett Publishing Company).
Tarski, A. (1983b), “The Concept of Truth in Formalized Languages”, in Tarski (1983a), 152–278.
Tarski, A. (1986), “What Are Logical Notions?”, ed. by J. Corcoran, in *History and*

Philosophy of Logic 7: 143–54.
Terzian, G. (2016), "Norms of Truth and Logical Revision", *Topoi* 34: 15–23.
Williams, J.R.G. (2016), "Probability and Nonclassical Logic", in A. Hájek and C. Hitchcock (eds.), *Oxford Handbook of Probability and Philosophy* (Oxford University Press), 248–76.
Williamson, T. (2017), "Semantic Paradoxes and Abductive Methodology", in B.P. Armour-Garb (ed.), *Reflections on the Liar* (Oxford University Press), 325–46.
Zalta, E.N. (1983), *Abstract Objects: An Introduction to Axiomatic Metaphysics* (D. Reidel).
Zalta, E.N. (2011), "Logic and Metaphysics", in A. Gupta and J. van Benthem (eds.), *Logic and Philosophy Today*, Vol. 2, Studies in Logic: Vol. 30 (College Publications), 153–82.