Logics of Truth and Maximality

Johannes Stern
Department of Philosophy
University of Bristol
johannes.stern@bristol.ac.uk

Keywords: Truth, Logics of truth, Theories of truth, Norms for truth, Disquotation

Abstract

The paper develops a precise account of a *logic of truth* and, in particular, the logic of truth of a given truth theory. On the basis of this account maximality considerations are employed for comparing and evaluating different classical logics of truth. It is argued that, perhaps surprisingly, maximality considerations lead to a fruitful criterion for evaluating logics of truth. The paper provides two different routes for motivating the application of maximality considerations in the truth case. The first route employs the Maxim of Minimum Mutilation and suggests that we should vindicate as much of our basic truth intuition, i.e. the transparency intuition, as it is consistently possible. The second route argues that maximality considerations underlie much of our logical theorizing and its application in the truth case is just one particular such instance. On this view maximality considerations do not implicitly depend on the basic truth intuition but rather explain how the intuition arises and why it is not fully reliable.

1 Introduction

Research on truth in logic and philosophy frequently takes what we will label the Basic Truth Intuition (BTI) as a starting point, namely, the idea that φ and that $\ulcorner\varphi\urcorner$ is true are in some strong sense equivalent or, perhaps, that the statement that φ and the statement that $\ulcorner\varphi\urcorner$ is true express the same thought. Traditionally, this intuition has been spelled out by the following schematic principle known as the T-scheme:

$$\ulcorner\varphi\urcorner \text{ is true, if and only if, } \varphi.$$

While there seems to be wide ranging acceptance of BTI, there is a considerable amount of disagreement of how the equivalence between φ and $T\ulcorner\varphi\urcorner$ is best understood. One point of disagreement is whether BTI requires the truth predicate to be *transparent*, that is, whether φ and $T\ulcorner\varphi\urcorner$ are intersubstitutable in every non-opaque context or whether it is sufficient for φ and $T\ulcorner\varphi\urcorner$ to be merely materially equivalent. Clearly, these two understandings need not be equivalent. A second and for our purpose more important point of disagreement concerns the reliability of BTI: on the one hand there are theorists who take the intuition to be an irrevocable fact about the role of truth in reasoning. On the other hand there are others who conceive of BTI as a pretheoretic datum that will ultimately have to hold up against theoretical scrutiny.

These two different ways of conceiving of BTI are largely due to the infamous Liar paradox. The Liar paradox shows that if BTI is taken at face value, it cannot be squared with classical logic. One has to either reject classical logic or, if classical logic is assumed, one has to reject BTI at least in its general, unqualified form. In other words, if classical logic is assumed, then φ and $T\ulcorner\varphi\urcorner$ cannot be (materially) equivalent for every sentence φ. The schemata

(TB) $$T\ulcorner\varphi\urcorner \leftrightarrow \varphi$$

has to be rejected. This leaves us with the well-known dilemma that either classical logic can no longer be assumed as the default logic of the truth-functional connectives or we have to reject (TB) as a constitutive principle of truth in all contexts.

No matter which horn of the dilemma we opt for, we then face a choice between a variety of different possible options: if, on the one hand, we reject classical logic—let's call this strategy Non-classical—there will be several non-classical logics that can serve as the logic of the truth-functional connectives and it seems a difficult challenge to single out one such logic on principled grounds.[1] If, on the other hand, we reject (TB)—let's call this strategy Non-equivalence—there will be several alternative combinations of principles and rules of truth that, using some loose terminology, can serve as the *logic of truth* irrespective of whether we are in a paradoxical context or not. But again there is the difficult challenge to single out one such logic of truth on principled grounds. In both cases the problem seems to be that the naive approach led us to adopt the strongest possible logic available—classical logic for the truth-functional connectives and the logic characterized by the principle (TB) for the notion of truth—but that there are many different ways for how a logic can be weakened.

In this paper we will be ultimately concerned with the classical approaches, that is, Non-equivalence, but to start it will be helpful to examine both strategies in tandem, as from a methodological perspective, independently of which is our preferred strategy for answering the paradoxes, we seem to end up in parallel situations: there is no longer one preferred logic but many different ones to choose from. However, upon closer inspection it turns out that there is a significant asymmetry between the two cases. While the proponents of Non-classical compare different non-classical logics and discuss which of these logics are best suited for characterizing truth-theoretic reasoning in all contexts, the proponents of Non-equivalence typically work with theories rather than logics of truth. Theories, in contrast to logics, are not merely concerned with schemata and general patterns of truth, but also with which sentences one can prove true in the theory. For example, theorists are often interested in whether a given theory of truth is conservative over some other theory or, relatedly, whether a given theory of

[1] To be precise, proponents of Non-classical can also opt for so-called substructural logics (cf. Beall et al., 2018).

truth enables us to prove the consistency of, say, Peano arithmetic or some other theory. At least prima facie answering such questions does not seem relevant or helpful for determining the logic of truth, that is, the schematic principles and rules of truth the proponent of Non-equivalence proposes in lieu of (TB). Giving a precise account of how to determine the logic of truth associated with a truth theory and, even more importantly, to provide an exact definition of what a logic of truth is, seems necessary to this effect. Such an account would lay the ground for proponents of Non-equivalence to compare their favorite theories of truth relative to the specific account of truth-theoretic reasoning, that is the logic of truth, the theory gives rise to. Indeed, to provide such an exacting account of the notion of a logic of truth will be the first goal of the paper. As such the paper fills a lacuna in the truth literature where discussions of the "logic of truth" can frequently be found without a clear account of what such a logic is and how it relates to theories of truth. Rather the use of the term 'logic of truth' often remains on a metaphorical level (see, e.g., Kremer, 1988; McGee, 1991, 2010) and a precise account of the term is awaiting.

Our proposal is to identify logics of truths with certain modal (operator) logics. We shall call them *truth logics* (as opposed to *logics of truth*) to stress the fact that we are now working with a well defined term and provide a clear criterion of how the truth logic of a given theory of truth can be determined. The guiding idea of the proposal is to transfer existing work on the notion of provability to the truth case. In studying the notion of provability a number of modal logics, so-called provability logics, have been connected to arithmetical theories in a precise way and, in particular, to the properties of the provability predicate of these theories (cf. Solovay, 1976; Smorynski, 1985; Boolos, 1993). The basic idea underlying these results is to interpret the modal operator of a provability logic as the provability predicate of an arithmetical theory. By analogy we shall propose conceiving of a truth logic as a modal logic in which the modal operator can be interpreted as the truth predicate of some truth theory. Moreover, as we shall explain, results of Czarnecki and Zdanowski (2019); Standefer (2015) and Nicolai and Stern (2021) establish that such truth logics exist.

On the basis of this proposal we can again compare Non-classical and Non-equivalence in a systematic way. Both cases advocate revising some aspects of our default inferential apparatus, that is, they propose to replace what seems to be the default logic by a strictly weaker logic. Non-classical proposes replacing classical logic by some strictly weaker alternative, while the truth logic conceived of as a modal logic remains untouched: it will simply be the modal logic in which φ and $\Box\varphi$ can be replaced in every context—call this logic the Transparent Logic of Truth (TLT). In contrast, Non-equivalence takes classical logic to be sacrosanct and proposes revision of the truth logic. Both cases of revision are cases in which the logic under consideration is weakened—indeed they have to be for both classical logic and TLT cannot be strengthened in any meaningful way without trivializing the logic. In both cases there are many ways the logic can be weakened viz. the abundance of non-classical theories of truth based on a variety of different non-classical logics, but also the plethora of classical theories of truth that, as we shall see, have very different logics of truth. But then we may ask how we are to choose from the variety of different possible logics and what the criteria are that should inform such a revision?

One maxim that, at least tacitly, seems to influence many theorists in this respect is the desire to keep the changes to classical logic and, respectively, TLT as minimal as possible: a non-classical logic replacing classical logic needs to remain as close as possible to classical logic and, similarly, a truth logic needs to remain as close as possible to TLT. Non-classical logics and truth logics need to be maximal in this respect. Indeed, Hjortland (2021) argues that work on theories of truth in the Non-classical camp has been at least tacitly influenced by the Maxim of Minimum Mutilation (MMM) originating from Quine's (1970) work.[2] Similarly, it does not seem too far fetched to interpret the seminal formal work by Friedman and Sheard (1987) on classical theories of truth along these lines.

Focusing on Non-equivalence we shall pick up on this idea and investigate whether MMM can serve as a criterion guiding our choice of logic. To make this discussion precise we identify a

[2]Of course, Quine originally used MMM to argue against Non-classical (cf. Quine, 1970, p. 85).

truth logics with the set of its theorems and, as we shall explain in Section 3 propose to measure proximity to TLT via the subset relation. This yields a partial ordering of the truth logics and we can single out those truth logics that are as close to TLT as it is consistently possible. While there will be not one such *truth-maximal* logic, it turns out that, perhaps surprisingly, there are prominent classical truth theories that have truth-maximal logics. This shows that, applied to Non-equivalence, MMM amounts to a powerful criterion that can be used to inform our choice of theory. Interestingly, truth-maximal logics can be motivated via an alternative route that is independent of MMM and does not hinge on the acceptance of TLT as our default logic of truth. Rather, we argue, truth-maximal logics are motivated by a general methodological principle that is implicitly adopted by most philosophers and logicians. This account will not only provide us with a criterion guiding our choice of the appropriate truth logic, but it also explains why TLT and, as a consequence, BTI seem so appealing in the first place. The second principal goal of this paper is to make these remarks precise and to investigate the use of MMM and maximality consideration more generally within the framework of Non-equivalence.

Here is the plan of the paper. We will start by introducing *truth logics* and showing how the *truth logic* of a truth theory can be determined (Section 2). We then explain how MMM can be applied to inform the choice of an appropriate truth logic and introduce the idea of a maximal truth logic, that is, truth logics that cannot be strengthened without falling prey to the paradoxes of truth (Section 3). By relying on precise results on formal theories of truth, we show that maximal truth logics exist and that, indeed, some prominent theories of truth have a maximal truth logic. Next we motivate the idea that truth logics should be maximal independently of MMM, TLT, and BTI (Section 4). This alternative motivation of *truth maximality* provides us with a nice explanation of the intuitive appeal of BTI, yet it also explains why the intuition should not be considered to be an irrevocable fact about truth. We end the paper with a discussion of how truth maximality relates to alternative criteria that are supposed to inform our choice of an adequate truth theory (Section 5).

2 Truth Logics

Before we start discussing truth logics specifically, it may be helpful to recall some basic characteristics of logics more generally. Perhaps the most important feature of logics is that they are concerned "with the form rather than the content of the argument" (Mendelson, 2010, p. xv). In logic we ask which argument-forms yield valid arguments independently of the content of the particular sentences figuring in the premisses and conclusion of the argument. This is one aspect of the so-called *formality* of logic.[3] In logics we are, as a consequence, concerned with the sentential structure of premisses and conclusion and the general pattern of the argument that arises due to the structure of the sentences figuring in the argument. In propositional logic we analyze argument patterns with respect to the sentence's boolean or truth-functional structure. In modal logic we are, in addition, considering their modal structure, that is, we introduce modal operators to our language. In first-order logic we are also concerned with the quantificational structure of arguments. The focus of logic on argument-patterns or forms is illustrated by the rule of Uniform Substitution (US)

(US)
$$\frac{\varphi(p)}{\varphi(\psi)}$$

which distinguishes a logic from a mere collection of sentences, that is, a set of theorems. The rule of uniform substitution asserts that if we can show that φ holds for some propositional variable p, then we can show that φ will hold for every formula ψ. In propositional logic propositional variables act as placeholder for sentences that can then be further combined via sentential connectives to form new, complex sentences. (US) highlights that no particular assumption regarding the interpretation of the propositional variable p has entered the derivation of $\varphi(p)$. Hence it is legitimate to infer that φ holds for any sentence ψ.

In contrast, in theories we are usually concerned with the content of sentences. This means that in theories we focus on proving facts about specific sentences: using the above terminology

[3]See, e.g., MacFarlane (2000) and Dutilh Novae (2011) for more on the formality of logic.

we look at sentences that are provable under a given interpretation of a propositional variable. As a consequence the rule (US) will usually fail if applied to theories and this constitutes one of the challenges we face when we try to extract a logic for the *truth predicate* from a given truth theory. Indeed, in truth theories we face two distinct problems. First, we are usually constructing a theory of truth over some base theory, that is, a theory of arithmetics or syntax. This causes the problem we have already alluded to, that is, we usually prove facts concerning *particular* sentences as opposed to schematic claims of the kind we usually encounter in logics. Assuming such a base theory proves necessary to provide names for the sentences of the language and their syntactical operations and hence the basic tools for spelling out rules and principles of truth. But the introduction of names for sentences leads us straight to the second problem: within the scope of the truth predicate the sentences do not occur. They are only mentioned. Despite a perhaps contrasting impression the rules and principles of truth cannot, at least in a straightforward way, be treated as schemata of some propositional logic. Hence, even if we happen to solve the first problem, it seems as if there was little hope of obtaining a logic of the truth predicate, which is closed under (US).

It is at this point that a detour via the notion of provability will prove instructive. In contrast to truth, logics of provability are well understood and these logics are logics of the notion of provability in a very precise sense. Yet, just as for the notion of truth, provability is usually conceived of as a predicate of sentences in some theory and thus exactly the same problems arise as in the truth case. In the case of provability the solution to these problems was a shift of perspective: for the sake of studying the logic, provability was no longer treated as a predicate but as a modal operator. A provability logic is hence a modal logic in which the modal operator is understood as a provability operator. Clearly, this move would also solve both problems we encountered in the truth case: within a modal logic no particular interpretation is forced upon the propositional variables and the modal operator is applied to sentences rather than their names. Therefore no quotation contexts interfere with the meaningful application of the rule (US). Whilst conceiving of provability as a modal operator enables studying its logic, the question remains in what sense studying particular modal logics amounts to a study of the logic of the provability predicate of a theory, which, of course, was the initial target. The answer was given by Solovay's (1976) seminal completeness theorem: Solovay showed that the modal logic GL was precisely the logic of the provability predicate of PA in PA. GL captures exactly the logical properties of the notion of provability of PA as displayed in PA itself.

The idea underlying the result was to translate the modal operator language into the arithmetical language and, in particular, to translate the modal operator as the provability predicate of PA. Solovay proved that the theorems of GL translate precisely into those sentences of the arithmetical language which are theorems of PA independently of the way we translate the propositional variables of the modal operator language. Since this connection between a modal logic and a theory is somewhat crucial to our project let us give a bit more detail. First, we introduce the notion of a realization. A realization $*\colon \mathsf{At}_\Box \longrightarrow \mathsf{Sent}_{\mathcal{L}_{\mathsf{Ar}}}$ is a function from the propositional atoms of the modal operator language to the sentences of the arithmetical language. Now every realization induces a translation H, called *arithmetic interpretation*, from the modal operator language to the arithmetical language:

$$\mathsf{H}^*(\varphi) := \begin{cases} \varphi^*, & \text{if } \varphi \in \mathsf{At}_\Box \\ 0 = 1, & \text{if } \varphi \doteq \bot \\ \neg \mathsf{H}^*(\psi), & \text{if } \varphi \doteq \neg\psi \\ \mathsf{H}^*(\psi) \wedge \mathsf{H}^*(\chi), & \text{if } \varphi \doteq \psi \wedge \chi \\ \mathsf{Pr}_{\mathsf{PA}}\ulcorner \mathsf{H}^*(\psi) \urcorner, & \text{if } \varphi \doteq \Box\psi \end{cases}$$

The only really interesting aspect of H is the translation of the modal operator: it is translated by the provability predicate while the translation of the formula in the scope of the modal operator is moved into quotation marks—in this context typically understood in terms of Gödel corners[4]—and thus forms a name the provability predicate can then apply to.

[4]In other words $\ulcorner H^*(\psi) \urcorner$ is the numeral of the Gödel number (the numerical code) of the formula $\mathsf{H}^*(\psi)$.

Solovay's completeness theorem says that the theorems of GL are precisely the formulas of the modal operator language whose arithmetic interpretation can be proved in PA for every realization.

Theorem 1 (Solovay). *For all $\varphi \in \mathcal{L}_\Box$*

$$\mathsf{GL} \vdash \varphi \Longleftrightarrow \textit{for all realizations } * (\mathsf{PA} \vdash \mathsf{H}^*(\varphi)).$$

The quantification over all realizations accounts for the fact that GL is closed under the rule of uniform substitution while PA qua theory is not: we consider only the theorems of PA that are true solely because of their logical form and independent of the choice of particular sentences that are used to instantiate the particular form. Solovay's theorem hence presents us with a strategy for extracting a logic of provability from the logical properties of the provability predicate of an arithmetic theory. The result shows how to strip away all aspects of the theory that are inessential to questions pertaining to the logic of the provability predicate.

But this is exactly what we are looking for in the truth case: we seek to extract a truth logic by focusing on the logical properties of truth predicate, that is, those aspect of a theory of truth that are essential to the characterization of a logic. Fortunately, it proves rather straightforward to apply the strategy used for provability to the truth case. We only need to adjust the strategy to the fact that instead of the provability predicate we are concerned with the truth predicate within some truth theory. The fundamentals of the strategy remain the same: a truth logic will be a modal logic in which the modal operator is understood as a truth operator. However, we will no longer translate the modal operator language into the arithmetical language simpliciter. Rather the target language will be the arithmetical language (base language) extended by a one-place predicate, that is, the truth predicate.

As a consequence, a realization $*\colon \mathsf{At}_\Box \longrightarrow \mathsf{Sent}_{\mathcal{L}_\mathrm{T}}$ will now be a function from the propositional atoms of the modal operator language to the sentences of the language of the truth theory. Again every realization induces a translation $\mathfrak{I}$, now called *truth interpretation*, from the modal operator language to the language of the truth theory:

$$\mathfrak{I}^*(\varphi) := \begin{cases} \varphi^*, & \text{if } \varphi \in \mathsf{At}_\Box \\ 0 = 1, & \text{if } \varphi \doteq \bot \\ \neg\mathfrak{I}^*(\psi), & \text{if } \varphi \doteq \neg\psi \\ \mathfrak{I}^*(\psi) \wedge \mathfrak{I}^*(\chi), & \text{if } \varphi \doteq \psi \wedge \chi \\ \mathrm{T}\ulcorner\mathfrak{I}^*(\psi)\urcorner, & \text{if } \varphi \doteq \Box\psi \end{cases}$$

This then allows us to ask what precisely the truth logic of a given truth theory Σ is, by asking which modal logic would take over the role of GL in this case, i.e., which logic should assume the role of the question mark in the equation below. For all $\varphi \in \mathcal{L}_\Box$

$$? \vdash \varphi \Longleftrightarrow \text{for all realizations } * (\Sigma \vdash \mathfrak{I}^*(\varphi)).$$

Definition 2 (Truth Logic)**.** *Let L be a modal logic. Then L is called a truth logic iff there exists a consistent theory of truth* Σ *such that*

$$L \vdash \varphi \Longleftrightarrow \textit{for all realizations } * (\Sigma \vdash \mathfrak{I}^*(\varphi)).$$

Indeed, very recently a number of "Solovay-style" results have been proposed by (Czarnecki and Zdanowski, 2019), Standefer (2015) and Nicolai and Stern (2021) for a number of different theories of truth such as FS and KF.[5] This shows that using Solovay's strategy for extracting the logic of the truth predicate of a given truth theory we obtain a viable account for associating a truth theory with a logic that aims at modelling reasoning with truth. In other words, interesting truth logics exist.

The take-home message of this technical discussion is that talk about logics of truth and, in particular, the truth logic associated with a particular truth theory can be pushed beyond a

[5]See Halbach (2014) for a presentation of the various truth theories appealed to in this paper.

merely metaphorical level and given a precise meaning: truth logics are simply modal logics that can be linked to truth theories via Solovay-style results. This contrasts with the metaphorical uses often to be found in the literature where the "valid" truth-theoretic schemata are presented in a laundry-list or as axioms of a theory of truth. To this effect it is important to note that the schematic axioms and rules of a truth theory need not coincide with the axioms and rules of its logic, that is, the identification of the truth logic of a given theory is not a trivial matter.[6] Nonetheless once the truth logic of a theory has been identified we can ignore the complicated aspects of the truth theory and simply study the properties of the truth operator in the logic under consideration.

Before we take up the second theme of the paper and explore whether MMM can be of help in choosing a truth logic, we briefly inquire whether there is some unifying feature that distinguishes truth logics from some other family of modal logics, for example, provability logics. At this point we wish to remain fairly neutral and do not wish to rule out logics on the basis of debated or contested philosophical views. Still there are some minimal conditions that can be expected to hold for every truth logic, but that will clash with core properties of many provability logics like GL. These minimal conditions are *non-emptiness*, i.e., something is true and *non-triviality*, i.e., not everything is true. Availing ourselves to truth value-constants the former can be expressed by $\Box\top$ while the latter is expressed by $\neg\Box\bot$. The non-triviality condition already proves sufficient for distinguishing truth logics from many provability logics.[7] The reason for this is that the non-triviality condition translates into a consistency requirement, if the modal operator is read as provability. But due to Gödel's second incompleteness theorem, we know that PA cannot prove its own consistency. This fact is reflected in GL where, for the sake of consistency, we cannot prove the modal operator to be non-trivial, i.e., consistent.

[6] Löb's theorem provides us with an immediate example. We can derive the formalized version of Löb's principle for some predicate P in sufficiently strong arithmetic theories on the basis of the three Hilbert-Bernays derivability conditions formulated for P. However, we cannot derive formalized Löb's principle within a modal logic assuming the modal operator versions of the Hilbert-Bernays derivability conditions. So we need to add (the operator version of) the formalized Löb's principle to the axioms of the logic.

[7] It proves sufficient for distinguishing provability logics that are normal modal logics. The provability logic GLS, which is provability logic of the provability predicate of PA in the standard model satisfies both *non-emptiness* but also *non-triviality*.

Interestingly, taken together the non-triviality and the non-emptiness of the truth operator guarantee that truth logics will be sublogics of TLT formulated in classical logic. Indeed, TLT is the maximal modal logic which satisfies both non-emptiness and non-triviality, that is, a logic that cannot be consistently strengthened. Of course, on our terminology TLT itself is not a truth logic because we know that its truth operator cannot be interpreted as the truth predicate of a consistent theory of truth via the truth interpretation. Logics like TLT, which strictly speaking are not a truth logic, but satisfy non-emptiness and non-triviality, i.e., whose truth operator displays truth-like features, will sometimes be called logics of truth in alignment with the philosophical literature which uses 'logic of truth' in a rather loose sense. As we noted TLT is a rather special logic of truth, as it is the maximal logic of truth: there is no stronger consistent logic of truth. All (consistent) logics of truth, i.e., all modal logics that satisfy the non-triviality and the non-emptiness condition will be sublogics of TLT. But then a natural question arises: is there a maximal truth logic in the strict sense, that is, a truth logic for which no stronger truth logic exists?

3 Maximality and MMM

We closed the previous section by querying about the existence of maximal truth logics for which no stronger truth logic exists. For a maximal truth logic there is no stronger logic that is the logic of the truth predicate of a consistent truth theory, and such a logic would mutilate TLT in the most minimal way. If such logics exist, there is a straightforward way in which the Maxim of Minimum Mutilation (MMM) could be applied to the Non-equivalence strategy: MMM would advise adopting a maximal truth logic. Surprisingly, one can give rigorous formal results that establish the existence of maximal truth logics: there are truth logics for which no stronger truth logic exists.

To show this we need to properly introduce TLT. In classical logic TLT is simply the modal logic axiomatized by the modal operator version of (TB):[8]

$$(\mathsf{TB}_\Box) \qquad\qquad \Box\varphi \leftrightarrow \varphi.$$

Recall that all truth logics are expected to satisfy the non-triviality and non-emptyness conditions and, importantly, and all logics that meet these two conditions will be sublogics of TLT. It remains to specify when a truth logic truth is maximal, i.e., when it is a strongest possible truth logic that can be consistently maintained in the presence of self-referential, Liar-like propositions. To give an exacting definition of *truth maximality*, i.e., a definition of which logics are the strongest possible truth logics, one needs to appreciate what the strengthening of a logic amounts to and how different logics can be compared. Against the backdrop of classical logic we can conceive of a logic as a set of sentences of the language—the theorems of the logic—which is closed under (US). One can then meaningfully compare different logics via the subset relation.[9] If a set of logics $\mathbb{L}$ is partially ordered via the subset relation, a logic L will be maximal in $\mathbb{L}$ iff $L \in \mathbb{L}$ and there is no $L' \in \mathbb{L}$ such that $L \subsetneq L'$. It is important to keep in mind that in this case L' cannot be generated from L by simply adding an arbitrary sentence, say a propositional variable p, to L because the resulting set, though a superset of L, would not be closed under (US) and hence not be a logic. If we let $\mathbb{L}$ be the set of truth logics, i.e., the set of logics for which the modal operator can be interpreted as the truth predicate of a consistent truth theory via the truth interpretation, we obtain the following definition of truth maximality:

Definition 3 (Truth Maximality)**.** *L is truth-maximal iff L is a truth logic and there exists no truth logic L' such that $L \subsetneq L'$.*

By results of Czarnecki and Zdanowski (2019) and Nicolai and Stern (2021) we know that truth-maximal logics in the sense of Definition 3 exist. Indeed, these truth-maximal logics are not truth logics of obscure theories, but truth theories that are prominent in the philosophical

[8] In Hughes and Cresswell (1996) this logic is called Triv.

[9] Logics can be incomparable relative to the $\subseteq$-relation, but we can clearly say when one logic is stronger than another: a logic L' is strictly stronger than a logic L, if and only if, L' proves more theorems, i.e., if and only if $L \subsetneq L'$.

discussion. For example, the truth-maximal logic $\mathsf{KDD_c}$ is the logic of the theory FS and of the truth predicate of McGee's theory of truth (cf. McGee, 1991, p. 216). Similarly, the truth-maximal logic $\mathsf{M^n}$ is the truth logic of the theory $\mathsf{KF} + \mathsf{CN}$.[10] We take it that the existence of interesting truth-maximal logics shows that, if our understanding of MMM is accepted, MMM can be fruitfully applied in the context of Non-equivalence. It singles out specific truth logics that meet a precise condition, i.e., the logics have to be truth-maximal and these logics vindicate as much of BTI as is consistently possible without falling prey to the paradoxes of truth. Yet, MMM does not single out a unique truth logic. There will be many truth-maximal logics. These correspond to different ways in which TLT can be weakened. Or, in other words, there will be many different paths leading from the minimal truth logic to TLT and on each of these paths there will be a (distinct) truth-maximal logic (cf. Figure 1). It remains a philosophical and truth-specific task to single out the appropriate truth-maximal logic amongst the many. Nonetheless, it seems fair to say that in the context of Non-equivalence, MMM gives rise to an interesting and useful criterion to assess not only truth logics but also truth theories. Truth-maximality can serve as a necessary condition on truth logics and theories respectively.[11]

However, one may argue that the way we have applied MMM is confused and, indeed, that MMM is not applicable at all in the context of Non-equivalence. The idea is that in contrast to the case of Non-classical in which classical logic should indeed be considered as the received and established view, BTI and, as a consequence, TLT cannot assume a similar role in the context of Non-equivalence. Classical logic is the basic tool for reasoning in many sciences such as mathematics and (at least most areas of) physics to name only few. Hence, there are good arguments that are fully independent of the truth specific discussion in support

[10]We display the different truth-maximal logics in the appendix to this paper. The former result is implicit in Czarnecki and Zdanowski (2019) while the latter result is due to Nicolai and Stern (2021). KF is the theory Kripke-Feferman as presented in Halbach (2014) and CN stands for the axiom that asserts that (the interpretation of) the truth predicate is consistent.

[11]Let us, once more, be clear what the truth-maximality of a logic does not imply: it does not imply that if the logic of some theory Σ is truth-maximal, Σ is maximally consistent, or that Σ is the strongest possible recursively enumerable theory available. It tells us that strengthening the theory will not have an effect on the underlying truth logic, i.e., on the truth-theoretic fragment of the theory of truth that is closed under the rule of uniform substitution. As a consequence, a truth-maximal logic may be the logic of different and potentially even incompatible theories of truth. See Nicolai and Stern (2021, §5.2) for discussion.

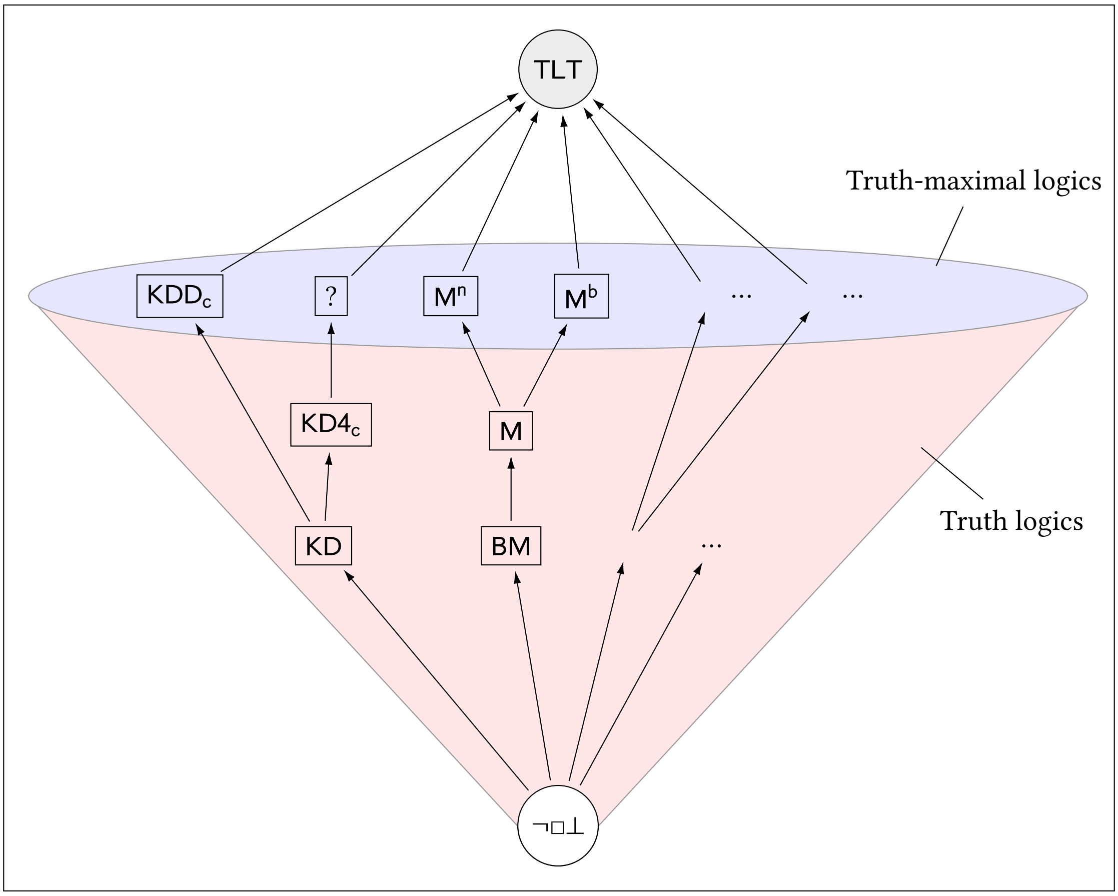

Figure 1: Logics of truth

of classical logic: classical logic is not to be given up lightly and should be conceived of as the default logic and received view. But TLT, at least the way we have presented things, was solely supported by BTI and it is one of the features of Non-equivalence to adopt a skeptical stance towards BTI. According to Non-equivalence, or so we argued, BTI is best understood in terms of a loose generalization rather than a firmly established position or view. If we take this idea seriously, so the argument would go, a further, independent argument in support of TLT seems necessary. In the absence of such independent support for TLT the application of MMM in the context of Non-equivalence seems to be, to say the least, in need of clarification. This, in turn, would undermine the argument in favor of truth-maximal logics. On this score truth-maximality should not be considered a necessary condition for logics of truth. We think that there is some credibility to this sort of argument. Yet, in practice, many classical truth theorist, e.g. Friedman and Sheard (1987); Feferman (2008); Leitgeb (2005); Schindler (2014); Picollo (2020) and Fujimoto and Halbach (2023), seem to, at least implicitly, endorse BTI and truth-maximality should be of some appeal to these theorists.

4 Truth-Maximality without MMM

If we take the argument against the application of MMM seriously, there seems to be no reason why the classical logician should favor truth-maximal logics to ordinary truth logics. In this section we explore the prospects of motivating truth-maximal logics in an alternative way that does not appeal to MMM and BTI. The alternative picture we propose depends on the acceptance of a general methodological maxim or principle that is intended to guide the formulation of logics more generally:

L-Maximality: The logic for notion X should be a strongest X-coherent logic possible.

At first glance L-Maximality may appear vague and implausibly strong, but we think that if suitably understood it is a fairly harmless and innocent maxim. To this effect it is important to appreciate that X-coherence is (typically) thought to be a much more encompassing and demanding requirement than mere consistency. Rather, X-coherence should be understood as requiring the logic to be philosophical adequate or acceptable. Indeed, the key idea underlying L-Maximality is that once all philosophical and logical desiderata regarding the logico-philosophical notion under consideration have been spelled out, we should opt for the strongest logic that fits these desiderata. For example, if the logico-philosophical desiderata pertaining to the notion of knowledge are compatible with the principle

(KK) $$\mathrm{K}\varphi \rightarrow \mathrm{KK}\varphi,$$

then (KK) should be a theorem of our logic: if there is no argument against adopting (KK), (KK) should be adopted. Of course, Williamson (2000) and others have given strong arguments against (KK), but the point is precisely that if these arguments are sound, the arguments and their conclusion will be part of the coherence criterion for knowledge. On this view, a logic that proves (KK), will not be a knowledge-coherent logic. If, to the contrary, the arguments against (KK) are fallacious, (KK) should be part of the logic of knowledge (ceteris paribus). This seems to be fairly standard, if implicit, methodology amongst philosophical logicians who will only veto a logic if the logic has counterintuitive consequences, i.e., a logic that contradicts some of the philosophical and formal desiderata put forth. Viewed in this light L-Maximality seems to be a fairly innocent principle that can be accepted without great cost.[12]

It may then come as a surprise that despite its innocuous character L-Maximality yields a straightforward argument in favor of truth-maximal logics. We know that truth-maximal logics are the strongest truth logics, that is, the strongest possible logics that can be understood as characterizing the notion of truth in classical logic. And all these logics will be sublogics of TLT formulated in classical logic. Choosing a truth-maximal logic from amongst the truth logic is thus precisely what the maxim of L-Maximality advises, i.e., to choose a strongest truth-coherent logic. To fully appreciate the role of L-Maximality in motivating truth-maximal logics it is important to understand the underlying notion of truth-coherence. To this end, recall that according to Definition 2 a truth logic is a modal logic for which the modal operator can be interpreted as the truth predicate of a consistent truth theory. Truth-coherence will thus depend on the adequacy conditions imposed on truth theories. A minimal adequacy condition on truth theories we already discussed at the end of Section 2 was that theories of truth ought to be non-empty (something is true) and non-trivial (not everything is true). This minimal adequacy condition and the resulting notion of truth-coherence led to truth logics and

[12]Notice that L-Maximality is not intended as a guideline in the pragmatic sense, that is, it does not mean that, in case of doubt, we should always opt for the stronger logic: we might not know, or be unsure about, the appropriate coherence-condition and prefer to err on the side of caution. In this case we may want to stick to a weaker logic, which we know to be "safe", that is, to be in conformity with the coherence-condition no matter how it is eventually spelled out. In other words, L-Maximality assumes complete information and is to be understood as an abstract theoretical principle rather than a principle guiding our choice of logic in practical applications.

truth-maximal logics as displayed in Figure 1. For sake of the discussion we shall call the latter logics *strongly truth-maximal* from now on. Strongly truth-maximal logics are logics that cannot consistently be strengthened if we allow for self-referential propositions. In contrast, once we introduce a stricter adequacy condition on theories of truth, there may, at least in principle, be truth-maximal logics that can be consistently strengthened albeit not *coherently*. Any extension of such truth-maximal logics violates the truth-coherence condition at play, but does not necessarily lead to plain inconsistency in presence of self-referential propositions. In other words, the question is whether L-Maximality *truncates* or *squeezes* (narrows down) the cone of truth logics displayed in Figure 1.

In the truncation case one might worry that truth maximality as a criterion for truth logics will be somewhat less interesting, as it would only make sense to determine the truth-maximal logics once the logico-philosophical account of truth has been fully developed, that is, once the relevant truth-coherence criterion has been fully specified.This would rule out truth-maximality as a pretheoretic criterion that can be used to narrow down the range of admissible truth logics before suitable adequacy conditions have been developed. Drawing such a conclusion is too hasty, however. Determining the truth-maximal logics will remain of interest for at least two reasons: for one, it allows us to specify the precise truth logics that match the logico-philosophical account under consideration. For another, it also provides us with a test of whether a logico-philosophical account of truth, i.e. the relevant truth-coherence condition, is sufficiently specific (at least from the logical perspective). If there exists precisely one truth-maximal logic according to the given notion of truth-coherence, then the proposed notion of truth-coherence is sufficiently specific. Otherwise the given logico-philosophical rationale underdetermines the way we reason with truth. In this case the given rationale would either need to be supplemented to the effect that a unique truth-maximal logic can be singled out, or a general explanation of this underdeterminacy needs to be developed. At any rate requiring truth logics to be truth-maximal remains an important and useful criterion in developing adequate theories of truth and understanding truth-theoretic reasoning.

However, perhaps a case for strongly truth-maximal logics can be mounted, that is, for the idea that L-Maximality will squeeze the cone of truth logics rather than truncate it. This would vindicate truth-maximality as pretheoretic criterion that can be applied prior to developing a philosophical account of truth. The idea is that if the notion of truth-coherence forces us to omit truth principles that can be consistently added, then we should expect a logico-philosophical justification of this omission and, more generally, an explanation of why the logic should not be further strengthened. Moreover, there seems to be at least some evidence that such a justification might be difficult to come by and that the logico-philosophical rationale in support of theories of truth whose truth logic is not strongly truth-maximal will remain underdeveloped. To illustrate the underlying idea it is instructive to consider the theory Kripke-Feferman (KF). By results of Nicolai and Stern (2021) we know that the modal logic BM is the truth logic of KF. Yet, this logic can be strengthened either by the modal principle D or the modal principle D_c to obtain the truth-maximal logics M^n and M^b respectively. The former is the truth logic of the theory KF + CN, while the latter is the truth logic of the theory KF + CM. It is well-known that both KF + CN and KF + CM have somewhat awkward features—at least from the perspective of BTI: while the former holds some of its theorems untrue, the latter refutes some sentences it holds true, i.e., for the Liar sentence λ

$$\mathsf{KF} + \mathsf{CN} \vdash \lambda \wedge \neg T\ulcorner \lambda \urcorner$$

and

$$\mathsf{KF} + \mathsf{CM} \vdash \neg\lambda \wedge T\ulcorner \lambda \urcorner.$$ [13]

KF avoids these awkward features as it is too weak to prove either of these conjunctions: KF does not hold any of its theorems to be untrue, nor does it refute any sentences it proves true. The reason is that KF cannot decide the disjunction $(\lambda\wedge\neg T\ulcorner \lambda \urcorner)\vee(\neg\lambda\wedge T\ulcorner \lambda \urcorner)$ (which it proves) and one might think that in this case the weakness amounts to an advantage. KF remains agnostic

[13] See, e.g., Reinhardt (1986); Bacon (2015), and Field (2008) for a discussion of these phenomena.

as to which disjunct should be adopted and thereby avoids the counterintuitive consequences of KF + CN and KF + CM respectively. At least *prima facie* one may thus be tempted to think that there is a good case for a truth-coherence criterion that singles out the modal logic BM. But, as Field (2008) convincingly argues, to be agnostic about the disjunction $(\lambda \wedge \neg T\ulcorner\lambda\urcorner) \vee (\neg\lambda \wedge T\ulcorner\lambda\urcorner)$ means that one is undecided which of the disjuncts to accept. To the contrary, if, as in the present case, one deems both disjuncts to be unacceptable, then one should also reject the disjunction, that is, $(\lambda \wedge \neg T\ulcorner\lambda\urcorner) \vee (\neg\lambda \wedge T\ulcorner\lambda\urcorner)$ and ultimately the truth theory KF. This suggests that an adequate and fully developed truth-coherence criterion that is to recommend a Kripke-Feferman-style theory will need to provide a convincing rationale of the asymmetry that arises in KF + CN or KF + CM, that is, in theories that have a strongly truth-maximal logic.[14] The task at hand is to provide a fully fledged out view that replaces BTI rather than to remain silent on the status of BTI. And one might take the moral of the case study to be that such a fully fledged out view will ultimately recommend strongly truth-maximal logics. Of course, this does not amount to an argument in support of strongly truth-maximal logics, but it shows that understanding truth-maximality in terms of strong truth-maximality should not be rejected out of hand. Moreover the example provides a stark reminder that a proponent of Non-equivalence must be careful not to implicitly rely on BTI (unless the latter has been independently motivated) in developing the logico-philosophical rationale of their account and should thus not reject potential truth logics on the basis of said intuition.

In light of their rejection of BTI the proponents of Non-equivalences face the challenge to explain why BTI and, for that matter, TLT are intuitively so compelling. Moreover, they also need to justify their frequent use of the T-scheme in philosophical debates in metaphysics and epistemology. Surprisingly, if our argument for truth-maximal logics is accepted, we possess an immediate rationale for the intuitive appeal of BTI: in absence of self-reference there will be a unique truth-maximal logic that meets the truth-coherence condition, namely, the logic

[14]Indeed, a philosophical defense and rationale for KF + CN has been developed by Maudlin (2004) and, to some extent, by Glanzberg (2004a,b, 2015).

TLT.[15] Arguably, the vicious forms of self-reference displayed by the Liar sentence and its kin is a fringe phenomenon in natural language, as well as in philosophical discourse. For most purposes we may then simply disregard the complications it engenders and work within a naive framework, that is, a framework in which TLT can be considered as the unique truth-maximal logic. Arguably, our intuitions regarding truth will form against the backdrop of this naive framework, i.e., a framework in which TLT is the logic of truth and it is thus no surprise that BTI is so intuitively compelling.[16] This suggest that explaining BTI will not in itself pose a challenge for the proponent of Non-equivalence. Rather such an explanation will come for free once a satisfactory logico-philosophical rationale for a classical truth theory has been provided. The challenge, of course, is to provide the latter (and one that we shall not address in this paper).

5 Conclusion

In the paper we developed a rigorous account of logics of truth, which led to our definition of *truth logics* (Definition 2). A truth logic characterizes the fragment of a truth theory which is closed under uniform substitution, i.e., the rule (US). This fragment consists precisely of the schematic truth rules and principles that apply independently of the particular content expressed by the sentences involved. With the precise definition of a truth logic at hand, one can compare different classical truth theories with respect to their underlying truth logic very much in the same way proponents of Non-classical compare their theories with respect to the non-classical logics the theories assume. In particular, different truth logic can be compared with respect to their proximity to the naive logic of truth, that is, the logic TLT. This led to the notion of a truth-maximal logic (Definition 3): a truth logic is a truth-maximal logic iff there

[15]That is, if the minimal truth-coherence criterion (non-triviality and non-emptyness) is assumed. Presumably, however, the same will hold for stronger truth-coherence criteria or, if not, the classical logician will indeed have quite a challenge in explaining BTI and the prevalent use of the T-scheme in many philosophical debates.

[16]The fact that in absence of self-reference "*all our unreflected intuition are preserved*" (cf. Gupta and Belnap, 1993, p. 218/19) is hardly a novel observation. See, e.g., Gupta and Belnap (1993); Gupta (1982), and Herzberger (1982) for discussion.

is no truth logic that properly extends it. A truth-maximal logic is as close to the naive logic of truth TLT, as it is coherently possible. We argued that truth-maximality can serve as an adequacy condition for truth logics and theories respectively.

Our arguments to this effect relied on normative criteria such as MMM or L-Maximality respectively. Viewed in this way our work thus supplements the list of norms for truth theories provided by Leitgeb (2007).[17] Indeed, according to our discussion Leitgeb's norm (d),

> "T-biconditionals should be derivable unrestrictedly."(Leitgeb, 2007, p. 279)

which cannot be satisfied if classical logic is assumed, should be replaced by the satisfiable norm

(d′) The logic of the truth theory should be truth-maximal.

The strength and theoretical purpose of (d′) will depend on whether truth-maximality is motivated via MMM or L-Maximality and, in the latter case if the relevant notion of truth-coherence can be expected to single out strongly truth-maximal logics, that is, truth logics that are as close to TLT as it is *consistently* possible. If either is the case we can conceive of (d') as a pretheoretic criterion that can be used to to narrow down the range of acceptable truth logics (and arguably that is how (d′) ought to be understood in the context of Leitgeb's discussion). Of course, in this case, as discussed in Section 3, there will be many candidate truth-maximal logic, so one cannot rely on (d') alone to single out a unique truth logic. Rather one needs to provide a logico-philosophical rationale to single out a unique truth-maximal logics amongst the many. Interestingly, some truth-maximal logics have already been defended on such principled grounds. For example, Maudlin (2004) develops a particular set of norms for assertion and denial that singles out the logic M^{n} while McGee (1991) provides a contrasting account in support of the truth-maximal logic $\mathsf{KDD}_{\mathsf{c}}$.[18] For a truth theorist this may come

[17] See also Sheard (2002); Halbach and Horsten (2015) and Terzian (2015) for a discussion of norms for truth theories.

[18] As a matter of fact, neither Maudlin (2004) nor McGee (1991) actually argue in favor of these logics qua logic. Rather they each argue in favor of a particular theory of truth whose truth logic is M^{n} and $\mathsf{KDD}_{\mathsf{c}}$ respectively. However, the arguments given in support of their favorite theory of truth are best reconstructed as arguments in favor of a particular truth logic.

as a surprise, as it is in stark contrast with the findings of McGee (1992). McGee showed that maximality considerations will not prove useful, if we attempt to block the paradox by restricting (TB) to sets of maximally-consistent instances of the scheme since, arguably, no such restriction can be motivated on principled grounds.[19]

If (d′) is motivated via L-Maximality and there is no expectation towards the relevant truth-coherence condition to single out strongly truth-maximal logics only, the theoretical role of the norm changes. In this case the norm can no longer be used as a pretheoretical selection criterion, but can be thought of as testing the logico-philosophical rationale in support of a given truth theory: the truth logic of the truth theory should be truth-maximal with respect to the relevant truth-coherence condition. On this understanding (d′) tests whether a theory of truth does justice to the logico-philosophical motivation that has been provided in its support and, in particular, whether the truth theory maximizes the motivation. As we mention in Section 4 truth-maximality also provides us with a test of whether the notion of truth-coherence at stake is sufficiently specific, that is, whether it singles out a unique truth-maximal logic. Otherwise, it seems fair to say that, in absence of a good explanation of this underdetermination of the principles of reasoning associated with truth, the philosophical account appears underdeveloped and needs to be enriched.

In conclusion, independently of whether (d′) and truth-maximality more generally are motivated via MMM or L-Maximality they amount to important truth-theoretic criteria and should be part of the proponent's of Non equivalence toolkit and the toolkit of any theorist working on classical theories of truth. Yet, the hard work for the truth theorist remains to provide a satisfactory logico-philosophical rationale, that is, as philosophical story in support of their theory and in this paper we have said little to help with this task. However, perhaps surprisingly, the hard task does not involve explaining why BTI seems so compelling despite

[19]McGee (1992) uses the fact that for every sentence of the language we can construct an equivalent T-sentence. In constructing a maximally-consistent set of T-sentences we thus need to decide whether to add the T-sentence that is equivalent to a truth-teller sentence or the T-sentence that is equivalent to the negation of the truth-teller sentence. But there simply seems no principled way of making these choices and, thus, even in combination with a philosophical story we cannot choose one such set over another in a principled way.

the fact that, according to the proponent of Non-equivalence, φ and $\mathrm{T}\ulcorner\varphi\urcorner$ will not be equivalent for all sentences φ. According to our view the intuition is simply a consequence of the fact that maximality considerations will lead to the Transparent Logic of Truth (TLT), if a naive framework is assumed, that is, a framework in which self-reference is disregarded. If the view is accepted the central role truth theorists have attributed to BTI and the T-scheme is misguided, and, indeed, an instance of *naivité*.

Acknowledgements

Most of the research that led to this paper was supported by the ERC Starting Grant *Truth and Semantics* (TRUST, Grant n. 803684). However, work on the theme of the paper started a long time ago, which makes it impossible to list all people that have contributed in some way or another to the present paper. I would like to explicitly thank Poppy Mankowitz, Theodor Nenu, Simone Picenni, and Thomas Schindler on helpful comments on a recent draft of the paper, as well as, Carlo Nicolai and Catrin Campbell-Moore for many helpful and stimulating discussions. I also thank two anonymous referees of this journal for helpful comments.

References

Bacon, A. (2015). Can the classical logician avoid the revenge paradoxes? *Philosophical Review*, 124(3):299–352.

Beall, J., Glanzberg, M., and Ripley, D. (2018). *Formal theories of truth*. Oxford University Press.

Boolos, G. S. (1993). *The Logic of Provability*. Cambridge University Press.

Czarnecki, M. and Zdanowski, K. (2019). A modal logic of a truth definition for finite models. *Fundamenta Informaticae*, 164(4):299–325.

Dutilh Novae, C. (2011). The different ways in which logic is (said to be) formal. *History and Philosophy of Logic*, 32:303–332.

Feferman, S. (2008). Axioms for determinateness and truth. *Review of Symbolic Logic*, 1:204–217.

Field, H. (2008). *Saving Truth from Paradox.* Oxford University Press.

Friedman, H. and Sheard, M. (1987). An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic*, 33:1–21.

Fujimoto, K. and Halbach, V. (2023). Classical determinate truth I. *Journal of Symbolic Logic*, pages 1–42.

Glanzberg, M. (2004a). A contextual-hierarchical approach to truth and the liar paradox. *Journal of Philosophical Logic*, 33:27–88.

Glanzberg, M. (2004b). Truth, Reflection, and Hierarchies. *Synthese*, 142(3):289–315.

Glanzberg, M. (2015). Complexity and hierarchy in truth predicates. In Achourioti, T., Galinon, H., Fujimoto, K., and Martínez-Fernández, J., editors, *Unifying the Philosophy of Truth*, pages 211–243. Springer.

Gupta, A. (1982). Truth and paradox. *Journal of Philosophical Logic*, 11:1–60.

Gupta, A. and Belnap, N. (1993). *The revision theory of truth.* The MIT Press.

Halbach, V. (2014). *Axiomatic Theories of Truth, 2nd Edition.* Cambridge University Press.

Halbach, V. and Horsten, L. (2015). Norms for theories of reflexive truth. In Achourioti, T., Galinon, H., Fujimoto, K., and Martiíez-Fernández, J., editors, *Unifying the Philosophy of Truth*. Springer.

Herzberger, H. (1982). Notes on naive semantics. *Journal of Philosophical Logic*, 11:61–102.

Hjortland, O. T. (2021). Theories of truth and the maxim of minimal mutilation. *Synthese*, 199:787–818.

Hughes, G. E. and Cresswell, M. (1996). *A New Introduction to Modal Logic.* Routledge, London.

Kremer, M. (1988). Kripke and the logic of truth. *Journal of Philosophical Logic*, 17(3):225–278.

Leitgeb, H. (2005). What truth depends on. *Journal of Philosophical Logic*, 34:155–192.

Leitgeb, H. (2007). What theories of truth should be like (but cannot be). *Philosophy Compass*, 2(2):276–290.

MacFarlane, J. (2000). *What does it mean to say that logic is formal?* PhD thesis, University of Pittsburgh.

Maudlin, T. (2004). *Truth and Paradox. Solving the Riddles.* Oxford University Press.

McGee, V. (1991). *Truth, Vagueness and Paradox.* Hackett Publishing Company, Indianapolis.

McGee, V. (1992). Maximal consistent sets of instances of Tarski's schema (T). *The Journal of Philosophical Logic*, 21:235–241.

McGee, V. (2010). Field's logic of truth. *Philosophical Studies*, 147(3):421–432.

Mendelson, E. (2010). *Introduction to Mathematical Logic.* CRC Press, 5. edition.

Nicolai, C. and Stern, J. (2021). The Modal Logic of Kripke-Feferman Truth. *The Journal of Symbolic Logic*, 86(1):362–395.

Picollo, L. (2020). Reference and truth. *Journal of Philosophical Logic*, 49:439,Äì474.

Quine, W. V. O. (1970). *Philosophy of Logic.* Harvard University Press.

Reinhardt, W. N. (1986). Some remarks on extending and interpreting theories with a partial predicate for truth. *The Journal of Philosophical Logic*, 15:219–251.

Schindler, T. (2014). Axioms for grounded truth. *Review of Symbolic Logic*, 7(1):73–83.

Sheard, M. (2002). Truth, provability, and naive criteria. In Halbach, V. and Horsten, L., editors, *Principles of Truth*, pages 169–181. Frankfurt a.M.

Smorynski, C. (1985). *Self-Reference and Modal Logic.* Springer Verlag.

Solovay, R. M. (1976). Provability interpretations of modal logic. *Israel Journal of Mathematics*, 25:287–304.

Standefer, S. (2015). Solovay-type theorems for circular definitions. *The Review of Symbolic Logic*, 8(3):467–487.

Terzian, G. (2015). Norms of truth and logical revision. *Topoi*, 34(1):15–23.

Williamson, T. (2000). *Knowledge and its limits.* Oxford University Press.

A The Modal Logics $\mathsf{KDD_c}$ and $\mathsf{M^n}$

The modal logics $\mathsf{KDD_c}$ and $\mathsf{M^n}$ are formulated in classical logic. That is, we assume the axioms of both modal logics to comprise a complete axiomatization of propositional logic in the language of modal logic. The syntax of the language is given by

$$\varphi \ ::= p \mid \bot \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box\varphi$$

where p is a propositional atom and $\bot$ the falsity constant. Throughout we shall take the remaining logical connective to be mere abbreviations and $\top$ to be defined by $\neg\bot$.

A.1 The normal modal logic $\mathsf{KDD_c}$

$\mathsf{KDD_c}$ is a normal modal logic which can be axiomatized by the following axioms and rules:

(D)	$\Box\neg\varphi \rightarrow \neg\Box\varphi$
($\mathrm{D_c}$)	$\neg\Box\varphi \rightarrow \Box\neg\varphi$
(K)	$\Box(\varphi \rightarrow \psi) \wedge \Box\varphi \rightarrow \Box\psi$
(Nec)	$\dfrac{\varphi}{\Box\varphi}$

A.2 The non-congruental modal logics $\mathsf{BM}, \mathsf{M}, \mathsf{M}^{\mathsf{b}}$ and M^{n}

The logic governing the scope of the modal operator of the non-congruental modal logic is a non-classical logic and, as a consequence, the modal operator is not closed under (classical) logical equivalence. The logic can be given by the following axioms:

Definition 4 (Modal logic BM). *The modal logic* BM *extends classical propositional logic with:*

$(\top)$	$\Box\top$
$(\bot)$	$\neg\Box\bot$
$(\neg)$	$\Box\varphi \leftrightarrow \Box\neg\neg\varphi$
$(\wedge 1)$	$\Box(\varphi \wedge \psi) \leftrightarrow \Box\varphi \wedge \Box\psi$
$(\wedge 2)$	$\Box\neg(\varphi \wedge \psi) \leftrightarrow \Box\neg\varphi \vee \Box\neg\psi$
$(\vee 1)$	$\Box(\varphi \vee \psi) \leftrightarrow \Box\varphi \vee \Box\psi$
$(\vee 2)$	$\Box\neg(\varphi \vee \psi) \leftrightarrow \Box\neg\varphi \wedge \Box\neg\psi$
$(\Box 1)$	$\Box\varphi \leftrightarrow \Box\Box\varphi$
$(\Box 2)$	$\Box\neg\varphi \leftrightarrow \Box\neg\Box\varphi$
$(\textsc{faith}^{\Box})$	$\Box\varphi \wedge \neg\Box\neg\varphi \longrightarrow \varphi$

Definition 5 (The logics M, M^{n}, M^{b}). *The modal logic* M^{n} *extends* BM *with (D); the modal logic* M^{b} *extends* BM *with* ($\mathrm{D_c}$) *and the modal logic* M *extends* BM *with*

$$(\mathsf{DD_c}) \qquad (\neg\Box\neg\varphi \vee \neg\Box\varphi) \vee (\Box\neg\psi \vee \Box\psi).$$