THE REVIEW OF SYMBOLIC LOGIC
Volume 8, Number 2, June 2015

AXIOMATIZING SEMANTIC THEORIES OF TRUTH?

MARTIN FISCHER
MCMP, LMU München

VOLKER HALBACH
University of Oxford

JÖNNE KRIENER
Birkbeck College, London

and

JOHANNES STERN
MCMP, LMU München

Abstract. We discuss the interplay between the axiomatic and the semantic approach to truth. Often, semantic constructions have guided the development of axiomatic theories and certain axiomatic theories have been claimed to capture a semantic construction. We ask under which conditions an axiomatic theory captures a semantic construction. After discussing some potential criteria, we focus on the criterion of ℕ-categoricity and discuss its usefulness and limits.

§1. Introduction. In recent years formal theories of truth have seen increased attention within the philosophical community. These formal theories of truth can be divided into two camps. Axiomatic theories of truth, on the one hand, attempt to characterize a concept of truth by stating axioms and rules for the truth predicate. Semantic theories of truth, on the other hand, attempt to characterize truth by defining a suitable interpretation of the truth predicate in a semantic metalanguage. These two camps, although very different in character, are intimately linked. Axiomatic principles are often used to motivate and justify semantic theories and semantic constructions have often guided the development of axiomatic theories. In particular, it is sometimes claimed that a certain axiomatic theory of truth captures or axiomatizes a semantic theory of truth. Such a claim, however, is in need of clarification because most semantic theories of truth characterize a set of sentences that is not recursively enumerable, i.e. they characterize a nonaxiomatizable set of sentences. In this paper we provide three possible explanans of the claim that an axiomatic theory axiomatizes a semantic theory of truth.

The paper is structured as follows. We start by explaining how the interplay between the axiomatic and the semantic approach has shaped formal work on truth (Section §2). Then, we introduce and discuss three different criteria as to when an axiomatic theory may be said to capture a given semantic theory (Section §3). The first criterion is based on the idea of structural similarity between the axiomatic and the semantic theory of truth.

Received: November 21, 2013.

doi:10.1017/S1755020314000379

The second criterion requires the axiomatic and the semantic theory to be of the same proof-theoretic strength. Our last criterion is based on a specific notion of categoricity to be specified. In Section §4 we focus on this last criterion, and discuss its merits and limitations in connection with Kripke style theories of truth. Finally, in the last section of this paper we briefly discuss the philosophical relevance of our results.

We assume the reader to be broadly familiar with axiomatic theories of truth. As to terminology and notation, we will largely follow Halbach (2014).

§2. The interplay of axiomatics and semantics. Since the inception of modern formal truth theory in Tarski's *Concept of Truth*, axiomatic and semantic theories have been methodologically linked and have therefore often been developed in tandem. This claim is not only intended as a systematic claim, but also as a historical observation: Axiomatic theories have played a crucial role in the development of semantic theories, and semantic theories have motivated many axiomatic theories. To substantiate the historical claim, we begin with the origin of modern formal theories of truth in Tarski. We merely highlight the relevant aspects of his account. Details are given in Halbach (2014, chap. 3) and Patterson (2012).

Tarski's main objective is to define truth of object-language sentences in an essentially richer metalanguage.[1] Before he sets out to do this, however, he specifies an adequacy condition, which is known as *Convention T*: A definition of 'T' as a predicate of truth is declared *adequate*[2] if and only if certain sentences follow from the definition: first, the sentences $T\ulcorner\phi\urcorner \leftrightarrow \phi$ (known as the 'T-sentences'), for every object language sentence ϕ; second, a sentence stating that only sentences of the object language are true.[3]

Tarski's definition of truth initiated the semantic approach to truth, and has become its paradigm. However, at a closer look, Tarski's perspective is not wholly semantic, but incorporates elements of the axiomatic approach, too. The starting point of his discussion are the T-sentences. These in turn are introduced like axioms. More precisely, Tarski takes the T-sentences as evidently correct, in the same way as, say, mathematicians take the axioms of PA as evidently correct. After all, Tarski does not see a need to argue for them. Thus, Tarski employs postulates or axioms in order to motivate and justify his semantic theory of truth.

However, he does not see his semantic theory merely as a technical vehicle for carrying out the reduction of truth to a mathematical theory. Instead, he uses his semantic theory to assess axiomatizations of truth. So he reverses the relation of axiomatics and semantics: The semantic theory of truth is used to evaluate the axiomatization.

This happens when later in the *Concept of Truth*, Tarski considers taking truth as a primitive notion and using the principles from Convention T as axioms. However, he rejects such an axiomatic approach. The reason he gives for this is puzzling (Tarski, 1956, p. 257).

[1] Notice that Tarski used the term "metalanguage" in a way that differs from the modern terminology. For Tarski a language does not only specify the vocabulary but also contains axioms and rules. To this extent, Tarski's usage of 'language' conflates language and theory. For more on this see Halbach (2014, p. 16f).

[2] We omit the specification 'material' as it has mislead readers of the English translation. For discussion, see Patterson (2012).

[3] Our presentation of Tarski is not intended to be a historically precise account and we skip many details. For instance, Tarski famously used structurally descriptive names for sentences and $\ulcorner\phi\urcorner$ should be understood accordingly.

Tarski observes that his definition of truth satisfies desirable principles not stated in Convention T. These are general principles such as the law of contradiction or the law of excluded middle. The observation that they should be consequences of a good theory of truth but are not derivable from the principles laid down in Convention T should have prompted Tarski to reconsider Convention T and ask whether it needs augmentation with those desirable principles. He merely considers them briefly as candidate axioms and as such evaluates them against his semantic definition. However, because the resulting system does not fully capture his semantic definition, Tarski quickly abandons this axiomatic approach Tarski (1956, p. 258).[4]

Taking a step back, we find Tarski's methodology to be characteristic of much work on formal truth theories, in the following sense. Semantic and axiomatic theories are developed in close connection. Axiomatic theories lead to semantic theories as Tarski's T-sentences guided his semantic definition. Conversely, semantic theories also lead on to further axiomatic theories. In Tarski this is the step just considered, when he realizes certain generalizations to be consequences of the semantic theory, but not to be proved from the T-sentences. Axiomatic theories are evaluated against semantics and semantic theories are judged by the truth-theoretic principles, that is, axioms and rules, that they validate.

The dialectics of Tarski's approach on which semantic and axiomatic theories are developed together, has shaped much work on truth to the present day.

Various authors have developed semantic theories; nowadays the semantic constructions are usually carried out in set theory, although they can be formalized in much weaker theories for arithmetic languages. Perhaps the most prominent semantic theory is due to Kripke (1975). As a matter of fact, Kripke does not specify a single semantic theory but rather a general method for obtaining semantic theories of truth. In particular, Kripke considered different logics for dealing with truth-value gaps and did not commit himself to a particular way of carrying out his construction.

Different axiomatic theories have been developed that are supposed to capture different instantiations of Kripke's theory. Most notably, Feferman (1991) proposed an axiomatization of the fixed points of Kripke's semantic theory with Strong–Kleene logic in its 'closed-off' variant, which has come to be called the *Kripke-Feferman* theory, or KF for short. 'Closed-off' here means the following. KF's intended model is classical. In it, 'T' is interpreted by a Strong–Kleene fixed point. All other sentences are declared not true, those false in Kripke's model as well as those without truth value. Cantini (1990) tried to do the same for supervaluational Kripke fixed points.

Building on Feferman's work, Burgess (2009) suggested an axiomatization of the minimal fixed point with Strong Kleene logic.[5,6] All these axiomatic theories suffer from the problems diagnosed by Tarski for his own theory: They are supposed to capture a semantic theory, but they obviously are deductively too weak to prove all sentences that are validated

[4] The account here is very condensed. For more details see Halbach (2014, chap. 3). However, the tension between Tarski's use of Convention T and the rejection of the axiomatic approach deserves to be discussed in more detail than, to our knowledge, has been done so far.

[5] There have also been attempts to axiomatize versions of Kripke's theory directly without 'closing off' the partial model. To this end Kremer (1988), Halbach & Horsten (2006), and others have tried to provide axiomatization of Kripke's theory in nonclassical logics.

[6] The research programme initiated by Field in his 2008 monograph *Saving Truth From Paradox* is also well understood along those lines. Although most work on Field's programme for languages containing an additional conditional focuses on semantic constructions, he himself seems inclined to view an "effectively generable" set of principles as the ultimate goal (Field, 2008, p. 277).

by the semantic theory. So the question arises in which way these axiomatic theories are supposed to capture the semantic constructions.

The main competitors of Kripke-style semantic theories come from revision semantics, as advanced by Herzberger (1982), Gupta (1982), and Gupta & Belnap (1993). It proved much harder to devise axiomatizations of these theories and 'capture' stable or nearly stable truth and related constructions. An attempt by Turner (1990b) proved unsound.[7] The theory of Halbach (1994) is meant to axiomatize revision semantics for finite levels. Horsten et al. (2012) advanced theories supposed to capture stable and nearly stable truth.

As we have argued, truth theorists go back and forth between axiomatic and semantic theories. We do not intend to discuss or defend this methodology any further. We have outlined it in order to explain why authors have come up with claims that a certain deductive theory axiomatizes a semantic construction and we believe, that such axiomatization claims gain their importance from the methodology of developing semantic and axiomatic theories of truth in connection with each other. In what follows, we are concerned solely with this latter aspect of the back and forth methodology. We will focus on the transition from a semantic to an axiomatic theory. In particular, we would like to ask in what sense an axiomatic theory of truth can be said to capture a semantic theory.

Tarski wanted an axiomatic theory to prove those general principles of truth, such as the law of excluded middle, which are consequences of his semantic theory. It mattered for him that the axiomatic theory thus matches the semantic theory or is complete with respect to it. Later authors have been much more explicit: theories such as Feferman's KF have been explicitly advocated as axiomatizating some given semantics. In the remainder of this paper we will investigate what exactly could be meant by such axiomatization claims and whether, once made precise, they are correct. We will then focus on one criterion which we think particularly suits the case of theories of truth: $\mathbb{N}$-categoricity. Moreover, we will restrict our attention to theories of truth inspired by Kripke (1975) and apply the criterion to them.

§3. Criteria. In this section we provide different criteria for determining whether an axiomatic theory axiomatizes a semantic theory of truth. Semantic theories can be conceived in an extensional or intensional way. On the extensional understanding, a semantic theory is merely a specific class of models of the language of truth. On an intensional understanding, the definition of the class forms part of the semantic theory, that is, the specific way the class is determined belongs to the semantic theory. For the main Criterion 3.3 an extensional understanding will suffice. In some cases, for example Criterion 3.1, the intensional understanding will also be of interest.

Of course, an axiomatic theory of truth can only be said to axiomatize a certain semantic theory, i.e.a class of models, if it is sound with respect to these models. That is, the axiomatic theory must be true in any model of the semantic theory. Soundness ensures consistency, and ω-consistency if, as it often is the case, the semantic theory is based on the standard model of arithmetic. However, soundness is at best a necessary criterion for an axiomatic theory to capture a given semantic construction, even if the axiomatic theory of truth is proof-theoretically strong. To convince oneself that soundness alone is insufficient it is enough to consider examples like the minimal fixed of the Kripke construction with the strong Kleene evaluation scheme. There are a variety of axiomatic theories, such as

[7] See Cantini (1996, p. 394).

PUTB, that are sound in the minimal fixed point, however they are not considered good axiomatizations of the minimal fixed point.

In this section we will seek a way of supplementing soundness by a sufficient condition. We will introduce and discuss criteria other than mere soundness, as to when an axiomatic theory of truth captures a semantic theory.

3.1. Similarity. One idea that may guide the theorist in axiomatizing a semantic theory is similarity between the axiomatic and the semantic characterization. Recall that in the present context, a semantic theory is a class of models. Having said that, in most cases these semantic theories are definable in some set theory or second order arithmetic. Although there are different possibilities to define a class of models, in some cases one definition appears particularly natural. In these cases, we assume a canonical definition.

CRITERION 3.1. *An axiomatic theory Σ is an adequate axiomatization of a semantic theory $\mathcal{M}$ if and only if the axioms resemble the canonical definition of $\mathcal{M}$.*

Of course, as it stands, the criterion is as imprecise as the notions of resemblance. However, at least in some cases it can be explicated in terms of *translations*. One example is Tarski's definition of truth and its axiomatic counterpart CT.[8]

According to Donald Davidson, an early proponent of the axiomatic approach, the clauses of Tarski's definition should be turned into axioms (Davidson, 1996, p. 277). Davidson suggested to follow Tarski's inductive definition but omit the last step of turning it into an explicit definition. Although he did not explicitly commit himself to a specific formal axiomatic theory, arguably CT would have been a viable candidate for him if the language of arithmetic were used as object language.[9]

As is well known, Tarski's definition of truth for first-order arithmetic can be carried out in second-order arithmetic. Following Davidson's suggestion, we can consider it as an inductive definition turned into an explicit definition. An inductively defined set is the least set closed under a certain monotone operator. In many cases, however, there are other such fixed point sets. In second-order arithmetic we can single out the least fixed point, our inductive set, by universal quantification over all sets closed under the operator. This can be understood as turning the inductive definition into an explicit definition.

In the case of arithmetical truth the closure conditions can be formulated by an arithmetical formula $\Phi(X, y)$, that is a formula without second-order quantifiers.[10] Hence, $\Phi(X, y)$ can be translated into the language of truth by replacing all occurrences of X by T. This allows us to derive axioms for truth. We only need to observe that the conjunction of the CT axioms is equivalent to the following, universally quantified sentence in the language of truth.

$$\forall y(T(y) \leftrightarrow \Phi(T, y)) \tag{3.1}$$

Thus, the axiomatic theory CT can be viewed as the result of translating Φ from the language of second-order arithmetic into the language of first-order arithmetic supplemented

[8] For a precise definition of CT, and other axiomatic theories of truth referred to in this paper, consult Halbach (2014).

[9] Keep in mind that Davidson was of course interested in languages much more comprehensive than that of arithmetic.

[10] In Tarski's original inductive definition X does not only occur positively, but it is possible to define the set of arithmetical truths via some X-positive formula $\Phi(X, y)$ and we take this to be the canonical definition. Compare also Footnote 11.

by a truth predicate. It is in this precise sense that Tarski's definition and the axioms of CT resemble one another.[11] CT satisfies the Criterion 3.1.

The theory TB, however, does not satisfy Criterion 3.1 because its axioms, the T-sentences, do not resemble the compositionality of Tarski's definition. Arguably, this is the correct outcome. As we pointed out in the previous section, Tarski himself observed that his T-sentences do not suffice to prove generalizations which ought to hold in a Tarskian theory of truth.

So far we have given two examples, one in which the criterion is satisfied and one in which it is not. Is it possible to apply the similarity criterion, regimented in terms of translation, to other cases? We could apply the criterion to candidate axiomatizations of Tarski's theory partly because this semantic theory can be given by an arithmetical formula Φ, as in Equation 3.1. In fact, this formula Φ determines the extension of the truth predicate uniquely – there is only one fixed point over the standard model that satisfies Tarski's closure conditions: the *Tarski truth set*.[12] Therefore, the minimality condition in the explicit definition is not necessary.

If we consider an inductive definition as for example in the case of Kripke's fixed point construction based on Strong Kleene, we still have a positive inductive definition by means of an arithmetical formula $\Psi(X, y)$. Replacing 'X' in $\forall x(T(x) \leftrightarrow \Psi(X, y))$ with 'T' we arrive at an axiom that can easily be broken down into the usual axioms of Feferman's theory KF (p. 259 above).[13] This gives us a translation between Kripke's semantic clauses and KF, just like that between Tarski's clauses and CT. Thus, we can apply the similarity criterion. However, in this case refraining from a minimality condition does make a difference. The inductive definition $\Psi(X, y)$ has several fixed points and therefore several possible extensions. The minimality condition, however, picks out a unique solution, namely the least fixed point. Since minimality is neglected in the proposed translation, none of the KF axioms expresses that the intended interpretation of 'T' is the least fixed point. Hence, KF satisfies the similarity criterion, but only with respect to the semantic theory of *arbitrary* Strong Kleene fixed points and not with respect to the least fixed point theory.

Thus, we have found two instances for which the criterion of similarity seems to work fine. In general, however, the criterion is rather limited. One serious problem is that the criterion only applies to theories that are characterized by an arithmetically definable operator, and therefore not allowing any second-order quantifiers in the (positive) inductive definition. This restriction enabled a translation into the language of truth and thus to sharpen the similarity criterion. However, already in the case of Kripke's theory based on supervaluation the operator is defined in terms of a universal second-order quantifier (see p.269 below) and as we will see the operator is no longer arithmetically definable. This basic feature of Kripke's supervaluational theories blocks the translation of their semantic clauses into the language of truth. Hence, we can no longer sharpen the imprecise notion of similarity as we could in the simpler cases of CT and KF. Moreover, there are semantic

[11] In fact, the case of CT is somewhat more involved. For a positive inductive definition of the set of arithmetical truths we have to use different closure conditions that differ slightly from Tarski's original version Halbach (2014). A formulation of the positive inductive definition is found in McGee (1991, p. 109). However, we also have a positive inductive definition of the complement. Combining these we get the axioms of CT.

[12] Again, see McGee (1991, p. 109).

[13] See Halbach (2014).

theories are not even positive inductively definable and for those theories, such as Revision theory, the criterion is not applicable.

3.2. Proof-theoretic strength. Axiomatic theories of truth with Peano arithmetic as base theory have often been compared to extensions of Peano arithmetic. In particular, there are many results that show that a certain truth-theoretic system is intertranslatable or proof-theoretically equivalent in some other sense, such as proving the same arithmetical sentences or proving the same amount of transfinite induction with a system of inductive definitions or a subsystem of second-order arithmetic. It might be asked whether proof-theoretic strength could also function as a useful criterion for axiomatizing semantic theories of truth. The guiding idea would be that the axiomatic theory should be at least as strong as a theory sufficient to carry out the semantic construction.

Before we try to render precise the criterion let us take a look at a motivating example. Kripke's fixed-point construction is a special case of an inductive definition. However there are "formal theories which directly axiomatize the crucial properties of i.d. [inductively definable] classes of natural numbers",[14] such as the theory ID_1 and $\widehat{\mathrm{ID}}_1$. The first axiomatizes the *minimal* fixed point and the latter *arbitrary* fixed points for arithmetically definable positive monotone operators. These theories are formulated in an expansion of the arithmetical language by new predicates. For each X-positive arithmetical formula $F(X, x)$ (without second order quantifiers), a new predicate $P_F(x)$ is introduced.

For ID_1 we have two different kinds of axioms. The first of those is $\forall x(F(P_F, x) \rightarrow P_F(x))$ and states that $I_F(x)$ is closed under F. The second is the scheme $\forall x(F(G, x) \rightarrow G(x)) \rightarrow \forall x(P_F(x) \rightarrow G(x))$ stating the minimality of $P_F(x)$. In contrast $\widehat{\mathrm{ID}}_1$ has only the characteristic axioms $\forall x(P_F(x) \leftrightarrow F(P_F(x), x))$ and unlike ID_1 it contains no minimality claim. $\widehat{\mathrm{ID}}_1$ is a proper subtheory of ID_1. Moreover $\widehat{\mathrm{ID}}_1$ is a predicative theory and thus proof-theoretically significantly weaker than the impredicative theory ID_1.

Accepting our guiding idea that the axiomatic theory should be at least as strong as a theory sufficient for the semantic construction motivates that an axiomatization of a minimal fixed point of Kripke should be able to interpret ID_1. Let us evaluate KF from a proof-theoretic perspective. Cantini (1989) observed that KF interprets $\widehat{\mathrm{ID}}_1$. However KF is just a special case of $\widehat{\mathrm{ID}}_1$ and therefore interpretable in it. This shows that KF does not interpret ID_1. For Burgess 2009, this limited proof-theoretic strength of KF is the main reason for rejecting it as an adequate axiomatization of the *minimal* fixed point.[15] So KF is proof-theoretically not sufficiently strong to axiomatize the Strong Kleene minimal fixed point theory, but it is adequate as an axiomatization of arbitrary fixed points.

An extension of KF that interprets ID_1 has been proposed recently by John Burgess (2009). His theory KFB extends KF by a scheme to the effect that the intended interpretation of 'T' is minimal with respect to the KF axioms.[16] KFB is proof-theoretically equivalent to ID_1. From a proof-theoretic perspective, therefore, KFB is a more adequate axiomatization of Kripke's minimal fixed point theory based on Strong Kleene.

We take the example as supporting a proof-theoretic criterion. However, it is not obvious how precisely to formulate such a criterion. A problem we encounter is that there is not always directly a theory axiomatizing the semantic construction, as in the case of ID_1.

[14] See Feferman & Sieg (1981, p. 33).

[15] Compare Burgess (2009, p. 17).

[16] KFB extends the well-known theory KF (Halbach, 2014, chap. 15, 17) by a schema saying that whenever a formula $\phi(x)$ is 'upwards' closed under the KF-axioms, then '$\forall x(T(x) \rightarrow \phi(x))$'.

For example, for monotone operators that are not arithmetically definable we have no obvious corresponding axiomatic theory. In those cases we would have to look for a more general way to make sense of the criterion. The guiding idea could be to consider theories that guarantee that the semantic theory is nonempty, i.e. that there is a set in the relevant class of models.

Usually some set theory is used to build models for the semantic theories and we could ask for our axiomatic theory of truth to be proof-theoretically as strong as a set theory sufficient to construct a model.[17] However, in many cases we can conceive of a semantic theory as a class of models $\mathcal{M}$ that can be characterized by a formula $\Psi_{\mathcal{M}}$ of second-order arithmetic. One simple example are the fixed points for a monotone operator Γ given by the formula $\Gamma(X) = X$. Then, we determine the means necessary to carry out the semantic construction by asking for the weakest fragment of second-order arithmetic that guarantees the existence of some set satisfying $\Psi_{\mathcal{M}}(X)$.[18] Our axiomatic theory of truth should then be proof-theoretically at least as strong as this fragment of second-order arithmetic. As for our motivating example we have that ID_1 and KFB are proof-theoretically equivalent to $\Pi^1_1\text{-CA}^-_0$, i.e. the second-order arithmetic extending ACA_0 by parameterfree Π^1_1 comprehension.[19] $\Pi^1_1\text{-CA}^-_0$ is sufficient to prove the existence of minimal fixed points for arithmetically definable monotone operators.

This suggests to complement soundness with respect to the models in $\mathcal{M}$ by the following criterion of proof-theoretic strength,

CRITERION 3.2. *An axiomatic theory Σ is an adequate axiomatization of a semantic theory $\mathcal{M}$ if and only if a theory that proves the existence of some extension satisfying $\Psi_{\mathcal{M}}$ is reducible to Σ.*

Of course this criterion is still vague. In particular, we should be more specific on what *reducible* means. As a first suggestion we take it to mean that the proof-theoretic ordinal of the theory proving the existence of an extension is smaller or equal to the proof-theoretic ordinal of the axiomatic theory. However, other notions of reducibility may be used as well, such as proof-theoretic reducibility or *relative interpretability*. In the latter case, the additional requirement may be imposed that the arithmetical vocabulary is not affected by the interpretation. It might well be that the plausibility of the motivating example relies on this specific notion and is closer in spirit to the similarity criterion.

One of the positive aspects of the criterion of proof-theoretic strength is that its range of applicability is wide. In contrast to the two other criteria, it is not only applicable to relatively simple semantic theories but also to more complex cases. On the negative side, the criterion is too coarse grained.

Some theories satisfy Criterion 3.2 with respect to a given semantic theory and are sound, but intuitively fail to axiomatize it. Consider the theory PUTB of positive type-free Tarski-biconditionals (Halbach, 2014, p. 276ff). PUTB is proof-theoretically equivalent to KF, in the sense that both theories prove the same arithmetical statements. This implies that $\widehat{\mathrm{ID}}_1$ and the relevant second-order arithmetic are not only reducible to KF but also to PUTB.

[17] Compare Cantini's remark after establishing that VF_p is proof-theoretically as strong as ID_1: 'Conversely, VF_p has a model in a set theory, which is proof-theoretically equivalent to ID_1' (Cantini, 1996, p. 356).

[18] The existence of such a set guarantees that $\mathcal{M}$ is nonempty.

[19] For a proof see Pohlers (2009, p. 351).

Moreover, PUTB is sound with respect to Kripke's theory in the same sense as KF; in fact PUTB is a subtheory of KF. So PUTB should be an adequate axiomatization of Kripke's theory in the same way that KF is. Intuitively, though, PUTB does not axiomatize Kripke's theory of truth with respect to the Strong Kleene evaluation scheme. The axioms of PUTB do not reflect the compositional structure of Kripke's theory with the strong Kleene schema; in fact one can prove that the compositional clauses, for instance the axiom that truth commutes with conjunction, cannot be proved in PUTB. Even worse, PUTB does not prove the truth of a single nonpositive sentence. But it is an important feature of Kripke's theory that it can also handle nonpositive sentences, for example $\neg T(\ulcorner 0 = 1 \urcorner)$. Thus, KF seems much better at capturing the main traits of Kripke's theory, even if it has the same proof-theoretic strength as PUTB. We conclude that proof-theoretic strength is at best a necessary condition of an axiomatization, but not a sufficient condition.

3.3. Categoricity. Taking a step back, the claim that an axiomatic theory of truth captures a semantic construction is naturally phrased in terms of categoricity: An axiomatic theory Σ of truth captures a class $\mathcal{M}$ of models if and only if the models of Σ are exactly those in $\mathcal{M}$, at least up to isomorphism. However, no recursively axiomatized theory of truth will ever capture an interesting class of models based on the standard model for arithmetic for obvious reasons: Any first-order theory that has infinite models, has models of different cardinalities.

The impossibility of a categorical syntax or base theory, however, should not cause us to abandon the categoricity approach. It does not rule out that the theory of *truth* may capture a certain semantic theory. This is obvious if the base theory contains contingent vocabulary and sentences such as *Snow is white*. Even if it does not decide such contingent sentences, an axiomatic theory may still fully capture, e.g., Tarski's semantic theory of truth. Equally, the failure of categoricity of truth theories based on Peano arithmetic should not be taken to mean that such an axiomatic theory cannot capture Tarski's semantic theory.

There are different ways to separate the problem of categoricity for the base theory, from the problem of categoricity of the truth theory. A straightforward way which we shall adopt consists in keeping the standard model fixed. More precisely, we focus on structures that extend the natural number structure $\mathbb{N}$ by an interpretation of the truth predicate S. Let $\mathcal{L}_{\mathrm{PA}}$ be the language of arithmetic, and let $\mathcal{L}_{\mathrm{PAT}}$ be $\mathcal{L}_{\mathrm{PA}}$ augmented by a one-place predicate T. For every set $S \subseteq \omega$ there is an $\mathcal{L}_{\mathrm{PAT}}$-model $(\mathbb{N}, S)$ that is obtained from the standard model $\mathbb{N}$ by interpreting T as S. In analogy to subsystems of second-order arithmetic (Simpson, 2009, p. 244), we call such a structure an '$\mathbb{N}$-model'. Restricting the categoricity requirement to $\mathbb{N}$-models, we arrive at the following adequacy criterion:

CRITERION 3.3. *Let Σ be an axiomatic and $\mathcal{M}$ a semantic $\mathcal{L}_{\mathrm{PAT}}$-theory. Σ is an adequate axiomatization of $\mathcal{M}$ if and only if for all $S \subseteq \omega$*

$$(\dagger) \qquad (\mathbb{N}, S) \models \Sigma \Leftrightarrow (\mathbb{N}, S) \in \mathcal{M}$$

Thus, an axiomatic theory Σ is deemed adequate if it is categorical with respect to $\mathbb{N}$-models and, in this case, we call Σ an $\mathbb{N}$-categorical axiomatization of $\mathcal{M}$.[20]

[20] The criterion as it stands has two assumptions. On the one hand we assume classical axiomatic theories. The criterion might be adapted to nonclassical settings, compare Section 4.3. On the other hand we assume a standard model for the base theory, which might be more problematic for nonarithmetical base-theories.

For example, Feferman's theory KF (p. 259 above) is $\mathbb{N}$-categorical with respect to the class of Strong Kleene fixed point models (Theorem 4.2 below). In fact, its $\mathbb{N}$-categoricity appears to underlie the acceptance of KF as an axiomatization of this version of Kripke's semantic theory (McGee, 1991, p. 93).

There are other examples of axiomatic theories $\mathbb{N}$-categorical with respect to their target semantic theory of truth. Recall from Section 3.1 the theory CT. It is true in a model $(\mathbb{N}, S)$ just in case S is closed under the clauses of Tarski's definition of truth, that is, just in case S is the Tarski truth set (p. 262 above). In fact, already TB, the theory of T-sentences, is $\mathbb{N}$-categorical with respect to Tarski's theory. Some remarks of Davidson's can be interpreted as motivating the T-sentences in Convention T from this fact.[21]

Due to this important role of $\mathbb{N}$-categoricity we will henceforth focus on this criterion and subject it to further examination. For this, we restrict our attention to one specific family of semantic theories, and ask for $\mathbb{N}$-categorical axiomatizations of Kripke's theory of truth. We will rehearse well-known positive results, most prominently the fact that KF is $\mathbb{N}$-categorical with respect to the Strong Kleene fixed point models. Our focus, however, will be on limitative results that have not yet been discussed in detail. We will show that for a wide range of interesting semantic theories of truth an $\mathbb{N}$-categorical axiomatization cannot be found.

§4. Axiomatizing Kripke's theory of truth. We concentrate on axiomatizations of semantic theories of truth which were laid out in Kripke's *Outline of a Theory of Truth*. There are at least two reasons for focusing on Kripke's theories. First, various axiomatic theories have been developed with the explicit intention of capturing one of Kripke's semantic theories. Some of the axiomatic theories have been advocated as important and promising axiomatic theories of truth. Second, as will become apparent, only some versions of Kripke's semantic theory of truth allow for an axiomatization in the sense of Criterion 3.3 whereas others do not.[22]

In his paper Kripke defines models for a language with type free truth. Kripke expands a model of the base language $\mathcal{L}$ to a model of the language $\mathcal{L}_T$ with truth predicate. For our purpose it will be sufficient to take the natural number structure $\mathbb{N}$ as the base model and to assume $\mathcal{L}_{PA}$ to be the base language. If we are given a suitable set of sentences that have been declared as true—possibly the empty set—Kripke showed how to transform this initial model of $\mathcal{L}_{PAT}$ to a model where the truth predicate has certain desirable properties. Kripke's idea was to work with partial models for the language $\mathcal{L}_{PAT}$, that is models in which an extension and an antiextension of the truth predicate is provided and where a sentence 'Tt' is true iff the denotatum of 't' is in the extension of 'T' and false if the denotatum is in the antiextension of 'T' though undefined otherwise. As a consequence of allowing so-called truth value gaps into the picture one has to give an account of how one can compute the truth value of complex propositions, that is one has to say, e.g., whether the conjunction of a sentence with undefined truth value and, say, a false

[21] Thus, in Davidson (1990, p. 299) he speaks of '[...] the key role of convention-T in determining that truth, as characterized by the theory, has the same extension as the intuitive concept of truth[...]'.

[22] There is further reason for focusing on Kripke's theories of truth which is that only few axiomatization have been suggested for alternative semantic theories of type-free truth such as the revision theory.

sentence is itself false or undefined. Although Kripke used the Strong Kleene scheme, he remained neutral as to which evaluation scheme is preferable. What matters to Kripke's proposal is that an evaluation scheme e gives rise to a monotone operator Γ_e, the "Kripke jump".[23]

DEFINITION 4.1 (Kripke Jump). *Let* $(\mathbb{N}, S^+)$ *be a model for* $\mathcal{L}_{\mathrm{PAT}}$*,* e *an evaluation scheme where* $(\mathbb{N}, S^+) \models_e \phi$ *denotes that* ϕ *is true in the model* $(\mathbb{N}, S^+)$ *according to* e*. The operation* $\Gamma_e : P(\omega) \to P(\omega)$ *such that*

$$\Gamma_e(S^+) = \{\#\phi : (\mathbb{N}, S^+) \models_e \phi\}$$

is called a Kripke jump iff it is monotone, i.e. for all S, S'

$$S \subseteq S' \Rightarrow \Gamma_e(S) \subseteq \Gamma_e(S')$$

The monotonicity of the Kripke jump guarantees the existence of fixed points, that is sets S for which $\Gamma_e(S) = S$ and, moreover, the existence of a *least* fixed point I_{Γ_e}. A Γ_e-fixed point S contains all sentences true according to the evaluation scheme e and thus counts as a suitable interpretation of the truth predicate in Kripke's theory of truth based on the evaluation scheme e. Kripke's theory of truth may therefore be taken to advocate the fixed points of the Kripke jump as the suitable interpretations of the truth predicate in the standard model. Accepting all the fixed points of Kripke's construction for a given scheme e means that Kripke's theory $\mathcal{M}_e$ consists of those standard models $(\mathbb{N}, S)$ for which $\Gamma_e(S) = S$. Reconsidering our Criterion 3.3 we therefore say that a recursively enumerable theory Σ axiomatizes Kripke's theory of truth spelled out using a scheme e, if and only if for all $S \subseteq \omega$:

$$(\mathbb{N}, S) \models \Sigma \Leftrightarrow \Gamma_e(S) = S$$

Instead of working with all fixed points, one may also restrict attention to some designated fixed points. The minimal fixed point stands out in this respect as it is thought to single out all and only the sentences whose truth value is grounded. If we understand Kripke's theory in this sense, then a recursively enumerable theory Σ will axiomatize Kripke's theory based on a scheme e according to Criterion 3.3 if and only if Σ uniquely singles out the least fixed point I_{Γ_e}, i.e. if $(\mathbb{N}, I_{\Gamma_e})$ will be the only standard model of Σ.

After this short introduction to Kripke's theory of truth we will ask for the prospects and limitations of axiomatizing Kripke's semantic theory of truth based on some scheme e in the sense of Criterion 3.3. We begin with Kripke's theory based on the Strong Kleene evaluation scheme *sk*.

4.1. Axiomatizing strong Kleene truth. As already mentioned, Kripke himself concentrated on the Strong Kleene scheme when discussing his theory of truth; and the Strong Kleene jump operation is a Kripke jump according to our Definition 4.1. By a well-known result of Feferman we know that the axiomatic theory of truth KF is an $\mathbb{N}$-categorical axiomatization of the Strong Kleene fixed points:[24]

[23] To simplify matters, we suppress mention of the truth predicate's antiextension. It can be recovered from the extension by negation: $S^- = \{\#\phi : \#\neg\phi \in S^+\}$ where $\#\phi$ denotes the Gödel number of ϕ.

[24] For a proof of the following theorem see e.g. Halbach (2014). Note that in contrast to Halbach we think of KF as including the axiom $\forall x(Tx \to Sent(x))$.

THEOREM 4.2 (Feferman). *For all $S \subseteq \omega$*

$$(\mathbb{N}, S) \models \mathrm{KF} \Leftrightarrow \Gamma_{\mathrm{sk}}(S) = S$$
$$(\mathbb{N}, S) \models \mathrm{KF} + \mathrm{Cons} \Leftrightarrow \Gamma_{\mathrm{sk}}(S) = S \,\&\, S \textit{ is consistent}$$
$$(\mathbb{N}, S) \models \mathrm{KF} + \mathrm{Comp} \Leftrightarrow \Gamma_{\mathrm{sk}}(S) = S \,\&\, S \textit{ is complete}$$

S is called consistent iff for every ϕ it is not the case that $\#\phi \in S$ and $\#\neg\phi \in S$, and complete iff for every ϕ, $\#\phi \in S$ or $\#\neg\phi \in S$.

KF axiomatizes the theory of the Strong Kleene fixed points in the sense of Criterion 3.3, because the fixed point property $S = \Gamma_{\mathrm{sk}}(S)$ guarantees that $(\mathbb{N}, S)$ is amongst the models of Kripke's theory.

This positive result contrasts with our negative findings if we ask for an $\mathbb{N}$-categorical axiomatization of the least fixed point. In this case Kripke's theory requires the axiomatic theory Σ to uniquely determine the interpretation of the truth predicate on the standard model to be the least fixed point $I_{\Gamma_{sk}}$. It turns out that there will be no axiomatic theory Σ satisfying Criterion 3.3.

To see this recall that the least fixed point $I_{\Gamma_{\mathrm{sk}}}$ of the Strong Kleene jump is Π^1_1-complete.

FACT 4.3 (Kripke). *$I_{\Gamma_{\mathrm{sk}}}$ is Π^1_1-complete.*

THEOREM 4.4. *There is no recursively enumerable first-order theory Σ such that*

$$(\mathbb{N}, S) \models \Sigma \Leftrightarrow S = I_{\Gamma_{\mathrm{sk}}}$$

Proof. The classical satisfaction relation is Δ^1_1 in the parameter S. Assume, for contradiction, that Σ axiomatizes $I_{\Gamma_{\mathrm{sk}}}$ in the sense of Criterion 3.3. In the remainder of the proof we use this assumption to provide a Σ^1_1-definition of $I_{\Gamma_{\mathrm{sk}}}$—which is absurd by Fact 4.3. Notice that since the classical satisfaction relation is Δ^1_1 we know that there is a quantifier free arithmetical formula Ψ such that we can rewrite $(\mathbb{N}, S) \models \Sigma$ as

$$\forall x (Pr_\sigma(x) \rightarrow \exists S' (\Psi(S', S, x)))$$

where σ is a standard representation of the axioms of Σ. Now S' does not occur in the antecedent of the conditional and we can therefore pull the existential quantifier in front. But by a theorem of Kleene[25] we also know that there is a quantifier free arithmetical formula Φ such that

$$\forall x \exists S' (Pr_\sigma(x) \rightarrow \Psi(S', S, x))) \Leftrightarrow \exists S' \forall x (\Phi(S', S, x))$$

Given this equivalence we can define $I_{\Gamma_{\mathrm{sk}}}$ as follows

$$y \in I_{\Gamma_{\mathrm{sk}}} \Leftrightarrow \exists S (\exists S' \forall x (\Phi(S', S, x)) \wedge y \in S).$$

However, since S' does not occur in the second conjunct we can pull the existential quantifier in front of the conjunction and thereby obtain a Σ^1_1-definition of $I_{\Gamma_{\mathrm{sk}}}$. □

Thus if we focus on the least fixed point there will be no $\mathbb{N}$-categorical axiomatization of Kripke's theory. An $\mathbb{N}$-categorical axiomatization is impossible because the semantic theory of truth provides a unique interpretation for the truth predicate, which is too complex for a characterization via the classical satisfaction relation. This shows that the theory KFB

[25] See, for example, Rogers (1967, p. 375, Theorem III).

proposed recently in Burgess (2009) (see Halbach, 2014, p. 17) cannot be an axiomatization of the least fixed point in the sense of Criterion 3.3 although KFB is clearly intended to be an axiomatization of the least fixed point.

Before we move on to Kripkean theories based on supervaluational evaluation schemes, we note that under certain qualifications everything we just said with respect to Kripke's theory of truth based on the Strong Kleene scheme carries over to the Weak Kleene scheme. First, there exists an $\mathbb{N}$-categorial axiomatization of the fixed points of Kripke's theory based on the Weak Kleene scheme: the system WKF.[26] Moreover, if the base language has symbols for certain recursive functions, the least Weak Kleene fixed-point is also Π^1_1 complete and we know that there cannot be an $\mathbb{N}$-categorical axiomatization of it.[27]

4.2. Axiomatizing supervaluational truth. We now turn to versions of Kripke's theory of truth based on supervaluational evaluation schemes. The idea behind supervaluation is that given an interpretation S^+ of the truth predicate, we consider arbitrary extensions S of S^+, each of which induces a classical model $(\mathbb{N}, S)$. Then we determine which sentences come out true in all these models and fix the new interpretation of the truth predicate to consist of (the codes of) these sentences.

Of course, the more extensions S we consider, the less agreement will there be between the models $(\mathbb{N}, S)$. Usually, therefore, further conditions are imposed on the range of extensions S considered. In the literature, various such *admissiblity* conditions have been discussed. We focus on the following supervaluation schemes:

DEFINITION 4.5 (Supervaluational Evaluation). *Let* $(\mathbb{N}, S^+)$ *be an* $\mathcal{L}_{\mathrm{PAT}}$ *model where* S^+ *is consistent, i.e. there exists no* $\phi \in Sent_{\mathcal{L}_{\mathrm{PAT}}}$ *such that* $\#\phi \in S^+$ *and* $\#\neg\phi \in S^+$. *Then for all* $\phi \in Sent_{\mathcal{L}_{\mathrm{PAT}}}$

$$(i)\qquad (\mathbb{N}, S^+) \models_{\mathrm{sv}} \phi :\Leftrightarrow \forall S(S^+ \subseteq S \Rightarrow (\mathbb{N}, S) \models \phi)$$

$$(ii)\qquad (\mathbb{N}, S^+) \models_{\mathrm{vb}} \phi :\Leftrightarrow \forall S(S^+ \subseteq S \,\&\, S \cap S^- = \emptyset \Rightarrow (\mathbb{N}, S) \models \phi)$$

$$(iii)\qquad (\mathbb{N}, S^+) \models_{\mathrm{vc}} \phi :\Leftrightarrow \forall S(S^+ \subseteq S \,\&\, S \textit{ is consistent} \Rightarrow (\mathbb{N}, S) \models \phi)$$

$$(iv)\qquad (\mathbb{N}, S^+) \models_{\mathrm{mc}} \phi :\Leftrightarrow \forall S(S^+ \subseteq S \,\&\, S \textit{ is maxcons} \Rightarrow (\mathbb{N}, S) \models \phi)$$

S^- *in item (ii) is the antiextension of the model and is given by* $\{\#\phi : \#\neg\phi \in S^+\}$. *$S$ is called 'maxcons' iff S is consistent and complete.*

It is straightforward to verify that all these supervaluation schemes give rise to a Kripke jump in the sense of Definition 4.1. The restriction to consistent S^+ proves necessary for $e \in \{\mathrm{vb}, \mathrm{vc}, \mathrm{mc}\}$. To see this, note that for inconsistent S^+ there will be no admissible superset S. Then the definiens is trivially satisfied, i.e. for all $\#\phi \in Sent_{\mathcal{L}_{PAT}}$, $(\mathbb{N}, S) \models_e \phi$ for $e \in \{\mathrm{vb}, \mathrm{vc}, \mathrm{mc}\}$.[28]

[26] For more on the theory WKF see Feferman (1991) and Fujimoto (2010). To our knowledge no proof of the $\mathbb{N}$-categoricity of WKF has been published so far, but a minor modification of the proof of Feferman's Theorem 4.2 yields the desired result.

[27] As Cain & Damnjanovic (1991) show, in the absence of certain function symbols from the base language $\mathcal{L}_{PA}$ the least fixed point may already be attained at ω and therefore be simpler than Π^1_1. We thank Thomas Schindler for pointing out to us that things are not as simple as we had expected.

[28] The scheme vb was discussed by Burgess (1986) and our results heavily rely on his work. Yet, Burgess does not assume the consistency of S^+ but requires S^+ not to intersect with the

The restriction to consistent sets S^+ does not make the resulting theory of truth less general. For every supervaluational scheme e with admissibility condition Ψ given by

$$\forall S(S^+ \subseteq S \,\&\, \Psi(S) \Rightarrow (\mathbb{N}, S) \models \phi).$$

the fixed points of Γ_e will always be generated starting from consistent sets S^+. In other words, the supervaluational Kripke jump can only be sound with respect to consistent sets S^+.[29] For suppose there was an inconsistent S^+ such that $S^+ \subseteq \Gamma_e(S^+)$. Then there must be a sentence ϕ such that

$$\forall S(S^+ \subseteq S \,\&\, \Psi(S) \Rightarrow (\mathbb{N}, S) \models \phi \wedge \neg\phi)$$

which is absurd. Consequently, every fixed point of Γ_e can be reached starting from a consistent set S^+ — except for, of course, the fixed point which arises from the trivialization of the satisfaction relation of a given supervaluation scheme e.

The following lemma collects some useful facts concerning the supervaluational Kripke jumps.

LEMMA 4.6. *Let e, f be supervaluation schemes and F_e the set of Γ_e-fixed points $\{S^+ : \Gamma_e(S^+) = S^+\}$. Moreover let $\Gamma_e \subset \Gamma_f$ denote that for all $S \subseteq \omega$, $\Gamma_e(S) \subset \Gamma_f(S)$. Finally, let $\lambda \in Sent_{\mathcal{L}_{\mathrm{PAT}}}$ such that* PAT $\vdash \neg T\ulcorner\lambda\urcorner \leftrightarrow \lambda$.

(i) *if $e, f \in \{\mathrm{sv}, \mathrm{vb}, \mathrm{vc}, \mathrm{mc}\}$ with $e \neq f$, then $F_e \cap F_f = \emptyset$;*

(ii) $\Gamma_{\mathrm{sv}} \subset \Gamma_{\mathrm{vb}} \subset \Gamma_{\mathrm{vc}} \subset \Gamma_{\mathrm{mc}}$*;*

(iii) *let $e, f \in \{\mathrm{sv}, \mathrm{vb}, \mathrm{vc}, \mathrm{mc}\}$, $\Gamma_e \subset \Gamma_f$ with $e \neq f$ and $S \subset \omega$ such that $S \subset \Gamma_e(S)$, then $\Gamma_e(S) \subsetneq \Gamma_f(S)$;*

(iv) *let $e \in \{\mathrm{sv}, \mathrm{vb}, \mathrm{vc}, \mathrm{mc}\}$ and S a consistent subset of ω then $\#\lambda \notin \Gamma_e(S)$.*

Proof. We only give a proof of item (i). The remaining items are straightforward consequences of the definition of a supervaluational Kripke jump Γ_e or follow from the previous items. For (i) we need to show that there is no S^+ such that $S^+ \in F_e$ and $S^+ \in F_f$. Suppose $S^+ \in F_{\mathrm{mc}}$. Then $\#\forall x(Sent(x) \rightarrow (Tx \vee T\underset{\cdot}{\neg}x)) \in S^+$ but for $X^+ \in F_f$ with $f \in \{\mathrm{sv}, \mathrm{vb}, \mathrm{vc}\}$ we have $\#\forall x(Sent(x) \rightarrow (Tx \vee T\underset{\cdot}{\neg}x)) \notin X^+$ as we always consider supersets S which are not maximal.

Now suppose $S^+ \in F_{vc}$. Then $\#\forall x(Sent(x) \rightarrow (\neg Tx \vee \neg T\underset{\cdot}{\neg}x)) \in S^+$ but for $X^+ \in F_f$ with $f \in \{\mathrm{sv}, \mathrm{vb}\}$ we have $\#\forall x(Sent(x) \rightarrow (\neg Tx \vee \neg T\underset{\cdot}{\neg}x)) \notin X^+$ as we always consider inconsistent supersets S.

Finally, suppose $S^+ \in F_{\mathrm{vb}}$. Then $\neg T\ulcorner 0 = 1\urcorner \in S^+$ but for $X^+ \in F_{\mathrm{sv}}$ we have $\neg T\ulcorner 0 = 1\urcorner \notin X^+$ for we always consider supersets S such that $\#0 = 1 \in S$. □

Notice that Lemma 4.6 implies that the minimal fixed point of the scheme sv, I_{sv}, is a subset of every supervaluational fixed-point. We now use a recent result of Welch (2014) to show that all supervaluational fixed-points are Π^1_1-hard.[30]

antiextension S^-. This equally prevents triviality. By the reasoning in the main text Burgess' scheme generates the same fixed points as our variant. vc is the scheme used by Cantini (1990).

[29] For a given set S the jump operation Γ_e is sound with respect to S if and only if $S \subseteq \Gamma_e(S)$. S is called a Γ_e -sound set.

[30] Welch's argument is devoted to showing that Leitgeb's dependence operator (cf. Leitgeb, 2005) is Π^1_1. He attributes the observation that his argument generalizes to the supervaluation case to Toby Meadows.

LEMMA 4.7 (Welch). *Let $X \in \Pi^1_1$. Then there exists a 1-1 recursive function f such that for every Γ_e-sound set $S \subset \omega$, $e \in \{sv, mc\}$*

$$n \in X \leftrightarrow f(n) \in \Gamma_e(S).$$

Proof. Save some minor tweaks the proof is due to Welch (2014). Let X be Π^1_1 and Seq the set of codes of finite sequences. Then we know that there exists a recursive relation $R(u,n) \subseteq Seq \times \mathbb{N}$ such that

$$(\dagger) \qquad n \in X \leftrightarrow \forall f \in {}^{\omega}\omega\, \exists k_0\, \forall k \geq k_0\, (\neg R((f \upharpoonright k), n)),$$

where $(f \upharpoonright k)$ is short for the code of the finite sequence $(f(0), ..., f(k))$.[31] Since the set $\Gamma_e(S)$ is supposed to consist of Gödel numbers of sentences we now introduce an alternative coding of finite sequences. A sequence $u = (u_0, \dots, u_n)$ will now be coded by

$$\#((\underbrace{\lambda \wedge \dots \wedge \lambda}_{u_0+1\text{-times}}) \vee \dots \vee (\underbrace{\lambda \wedge \dots \wedge \lambda}_{u_n+1\text{-times}}))$$

where λ denotes a standard liar sentence. We denote this deviant coding by *. The resulting set of sequence numbers Seq^* remains recursive and we can replace R by a recursive relation R^* in (†) such that

$$(\ddagger) \qquad n \in X \leftrightarrow \forall f \in {}^{\omega}\omega\, \exists k_0\, \forall k \geq k_0\, (\neg R^*((f \upharpoonright k)^*, n))$$

Next let σ_n be the $\mathcal{L}_{PAT}$-sentence expressing

$$[\exists u \in Seq^* \cap \mathrm{Tr} \wedge \forall u, v \in \mathrm{Tr}(u, v \in Seq^* \rightarrow$$
$$((u \subseteq v \vee v \subseteq u) \wedge \exists u' \in Seq^* \cap \mathrm{Tr}(u \subset u')))] \rightarrow \exists u \in Seq^* \cap \mathrm{Tr}(\neg R^*(u,n))$$

where Tr stands for the interpretation of the truth predicate. We now show that

$$(C) \qquad n \in X \leftrightarrow \#\sigma_n \in \Gamma_e(S)$$

For the left-to-right direction we assume $n \in X$ and $S \subseteq S'$ for some set S'. We need to show $(N, S') \models \sigma_n$. If S' does not contain the codes of the finite initial segments of some infinite sequence then there is nothing to show because the antecedent of σ_n will be false. Assume otherwise, i.e. for some function f, $(f \upharpoonright k)^* \in S'$ for infinitely many k. By $n \in X$ and (‡) it follows that there must be an $k \in \omega$ such that $\neg R^*((f \upharpoonright k)^*, n)$. This establishes the left-to-right direction for $e \in \{sv, mc\}$.

For the converse direction assume $n \notin X$. We need to show that $\#\sigma_n \notin \Gamma_e(S)$. We have to construct a set S' such that

$$S \subseteq S' \wedge S' \text{ is maxcons } \wedge (N, S') \not\models \sigma_n$$

From $n \notin X$ we know that there exists a function f such that $\forall k R^*((f \upharpoonright k)^*, n)$. Define $S_0 = S \cup \{(f \upharpoonright k)^* : k \in \mathbb{N}\}$. The set S_0 is consistent since neither the liar sentence λ nor any sentence equivalent to λ can be a member of a Γ_e-sound set. Since S_0 is consistent

[31] This follows from the following normal-form theorem by Kleene (cf. Rogers, 1967, §16.1, Corollary V). Let A be Π^1_1-set and A a recursive relation. Then

$$n \in A \Leftrightarrow \forall f \in {}^{\omega}\omega\, \exists k (R((f \upharpoonright k), n)).$$

See Welch (2014) for further explanation.

we can extend it to a maximally consistent set S' for which by construction $S \subseteq S'$ and $(\mathbb{N}, S') \not\models \sigma_n$. This establishes the converse direction for $e \in \{sv, mc\}$ and hence completes our proof. □

Welch's result in combination with Lemma 4.6 allows us to determine the lower bound of the complexity of the fixed points of a wide range of supervaluation schemes, namely for all supervaluation schemes between sv and mc.[32] As mentioned by Welch his argument establishes that already one application of the operator Γ_e to a Γ_e-sound set results in a Π^1_1-hard set.

THEOREM 4.8. *Let e be an evaluation scheme with* $\Gamma_{sv} \subseteq \Gamma_e \subseteq \Gamma_{mc}$ *and* $S \subset \omega$. *If* $\Gamma_e(S) = S$ *then S is* Π^1_1*-hard.*

Proof. We need to show that for all $X \in \Pi^1_1$ there exists a recursive 1-1 function such that

$$n \in X \Leftrightarrow f(n) \in S$$

By Lemma 4.6 we know that there exists an S' such that $\Gamma_{mc}(S') = S'$ and

$$I_{sv} \subseteq S \subseteq S'$$

By Lemma 4.7 we obtain

$$n \in X \Rightarrow f(n) \in I_{sv} \Rightarrow f(n) \in S \Rightarrow f(n) \in S' \Rightarrow n \in X$$

which establishes the claim. □

As a consequence of Theorem 4.8 there will be no $\mathbb{N}$-categorical axiomatization of any fixed point of the supervaluational schemes under consideration. In particular there will be no $\mathbb{N}$-categorical axiomatization of the least fixed point.

THEOREM 4.9. *Let e be an evaluation scheme with* $\Gamma_{sv} \subseteq \Gamma_e \subseteq \Gamma_{mc}$. *Then there is no recursively enumerable first-order theory* Σ *such that*

$$(\mathbb{N}, S^+) \models \Sigma \Leftrightarrow S^+ = I_{\Gamma_e}$$

Proof. By Theorem 4.8, following the outlines of the proof of Theorem 4.4. □

This theorem is analogous to Theorem 4.4, which establishes that no $\mathbb{N}$-categorical axiomatization of the minimal Strong Kleene fixed point is possible. In the case of the Strong Kleene scheme this negative finding contrasts with the positive result that KF satisfies Criterion 3.3 (Theorem 4.2). It is often stated that Cantini's theory VF

> [...] stands to the Supervaluation construction as Feferman's theory [i.e. KF] stands to the Strong Kleene model.(Leitgeb, 2007, p. 287)

Thus apparently VF is considered as an axiomatization of Kripke's theory based on the evaluation scheme vc. But we will show that Kripke's theory of truth based on (most) supervaluational evaluation schemes cannot be axiomatized in the sense of Criterion 3.3 and consequently VF does not stand in the same relation to the supervaluation construction

[32] This result might suggest that the different supervaluational schemes we consider coincide. But Lemma 4.6 shows that this is not the case. Indeed none of the four schemes we consider share a single fixed point. Moreover, even the unions over the lattice of fixed points of the respective supervaluation schemes do not coincide as the proof of Lemma 4.6 shows.

as KF stands to the Strong Kleene model. The reason is that the complexity of the fixed-point property, i.e. $\Gamma_e(S) = S$, of the supervaluational schemes is Π^1_1 in a parameter S (and not Σ^1_1). But if there were an $\mathbb{N}$-categorical axiomatization, we would have a Δ^1_1 definition of this property. The upper bound of the fixed-point property can be directly computed from the definition of the Kripke jump Γ_e. For the lower bound we shall appeal to a Basis theorem by Gandy.[33] Again our findings are robust in the sense that they hold for any supervaluation scheme in between sv and mc.

LEMMA 4.10. *Let e be an evaluation scheme with $\Gamma_{sv} \subseteq \Gamma_e \subseteq \Gamma_{mc}$. Then $\Gamma_e(S) = S$ is not Σ^1_1.*

Proof. By a corollary of a theorem of Gandy[34] we know the class of all functions of hyperdegree less than the hyperdegree of T, i.e. the set of all index numbers of the characteristic functions of a finite path tree,[35] forms a basis for Σ^1_1. For our case this means that every nonempty Σ^1_1-set of functions contains a function of hyperdegree less than the hyperdegree of T.[36] Now suppose that we have a Σ^1_1-definition of $\Gamma_e(A) = A$, then the set $F = \{c_A : \Gamma_e(A) = A\}$ is Σ^1_1 where c_A is the characteristic function of A. Then, by the Basis theorem there exists some $c_A \in F$ which is of hyperdegree less than T. However by Theorem 4.8 we know that each supervaluational fixed-point X is Π^1_1-hard. And if X is Π^1_1-hard then the hyperdegree of c_T is less or equal to the hyperdegree of c_X since T is hyperarithmetical in X, contradicting our assumption. □

We thus know that the fixed point property $\Gamma_e(S) = S$ is Π^1_1 (and not Σ^1_1) and hence not Δ^1_1. This implies that there will be no $\mathbb{N}$-categorical axiomatizations of Kripke's theory based on most supervaluation schemes.

THEOREM 4.11. *Let e be an evaluation scheme with $\Gamma_{sv} \subseteq \Gamma_e \subseteq \Gamma_{mc}$. Then there is no recursively enumerable first-order theory Σ such that*

$$(\mathbb{N}, S) \models \Sigma \Leftrightarrow \Gamma_e(S) = S$$

Proof. Assume for reductio that there exists a Σ such that for all $S \subseteq \omega$

$$(\mathbb{N}, S) \models \Sigma \Leftrightarrow \Gamma_e(S) = S$$

But $(\mathbb{N}, S) \models \Sigma$ is Δ^1_1 in S and we could thus give a Δ^1_1-definition $\Gamma_e(S) = S$ which contradicts Lemma 4.10. □

In particular, Cantini's theory VF is not $\mathbb{N}$-categorical with respect to Kripke's theory of truth based on the scheme vc.[37]

[33] A set of functions F is a basis for a family of sets of functions $\mathcal{P}$ if for all $P \in \mathcal{P}$ the following holds: there is a function f in P iff there a function f with $f \in F$ and $f \in P$. Thanks to Sean Walsh for drawing our attention to Gandy's theorem.

[34] See Rogers (1967, p. 421, Corollary XLIII(a)).

[35] For the definition of T see Rogers (1967, p. 395).

[36] Note that we are now talking about a Σ^1_1-set of functions and not of numbers.

[37] This specific observation can be obtained independently of our theorem. Cantini (1996, p.400) showed that the set of stable truths of a revision sequence with Herzberger's limit rule and the empty hypothesis (Gupta & Belnap, 1993, 5C) is a suitable interpretation of the truth predicate of VF in the standard model. Yet, due to an argument of (Burgess, 1986, p. 673) we know that this set is no fixed point of Γ_{vc}. While this shows that VF is not an $\mathbb{N}$-categorical axiomatization of Kripke's theory of truth based on vc, our observation implies that there cannot be an $\mathbb{N}$-categorical

4.3. Axiomatizing Kripke's theory of truth in partial logic. So far we have shown that there is no hope for providing an $\mathbb{N}$-categorical axiomatization of the fixed points of the supervaluational Kripke jump, at least for the more reasonable schemes. In a nutshell the reason for this failure was that the complexity of the fixed point property for the supervaluational schemes e with $\Gamma_{\mathrm{sv}} \subseteq \Gamma_e \subseteq \Gamma_{\mathrm{mc}}$ was nonreducible Π^1_1 in a parameter whereas the classical satisfaction relation is hyperelementary, that is Δ^1_1 in a parameter. In general this implies that fixed point theories of truth can only be axiomatized in classical logic if the corresponding fixed point property is at most Δ^1_1.

Now in the case of Kripke's theory of truth the complexity of the fixed point property seems to depend on the complexity of the satisfaction relation of the evaluation scheme under consideration. The Strong Kleene and the Weak Kleene satisfaction relation are of complexity Δ^1_1, which allowed for a characterization of the fixed property using the classical satisfaction relation. The supervaluational satisfaction relation is Π^1_1 and no $\mathbb{N}$-categorical axiomatization of the fixed-point property was possible. This analysis points towards a solution that the friend of the supervaluational approach might explore, if she is interested in an $\mathbb{N}$-categorical axiomatization of her semantic theory. Instead of axiomatizing Kripke's theory of truth in classical logic she might try to axiomatize the theory in partial logic. In particular, she might explore axiomatic theories based on a supervaluational logic. In this case our criterion of $\mathbb{N}$-categoricity for the special case of Kripke's theory of truth would boil down to the following requirement. Let e be an evaluation scheme. A theory Σ axiomatizes Kripke's theory of truth using the scheme e, if and only if, for all $S \subseteq \omega$

$$(\mathbb{N}, S) \models_e \Sigma \Leftrightarrow \Gamma_e(S) = S$$

Note that spelled out in this way the argument we have used to show that there is no $\mathbb{N}$-categorical axiomatization of Kripke's supervaluational theory of truth breaks down. The supervaluational satisfaction relation is Π^1_1 in a parameter S and thus of the same complexity as the fixed point property $\Gamma_e(S) = S$. Surprisingly, this also implies that an $\mathbb{N}$-categorical axiomatization of the least fixed point version of Kripke's supervaluational theory of truth is no longer ruled out by our complexity considerations because the complexity of the least fixed point is also Π^1_1, hence of the same complexity as the supervaluation satisfaction relation. Of course, this does not show that there exists an $\mathbb{N}$-categorical axiomatization of Kripke's supervaluational theory in partial logic but at least such an axiomatization is not ruled out in principle.

In the case of Strong Kleene evaluation scheme such an $\mathbb{N}$-categorical axiomatization of Kripke's theory of truth in partial logic has been provided, namely the theory PKF developed by Halbach & Horsten (2006).

THEOREM 4.12. *For all $S \subseteq \omega$*

$$(\mathbb{N}, S) \models_{sk} \mathrm{PKF} \Leftrightarrow \Gamma_{sk}(S) = S$$

Halbach and Horsten only show the soundness of the theory PKF with respect to Strong Kleene fixed point models, but the converse direction proves to be straightforward.[38] Note,

axiomatization of Kripke's theory based on a wide range of supervaluation schemes. We thank Toby Meadows and Philip Welch for the pointer.

[38] A small qualification, however, is necessary. As far as we can see it is not sufficient to assume that all theorems of PKF are true in $(\mathbb{N}, S)$ according to the Strong Kleene scheme, but the model also needs to preserve truth with respect to all PKF derivable sequents.

however, that there will not be an $\mathbb{N}$-categorical axiomatization of the least fixed point version of Kripke's theory in Strong Kleene logic. Such an axiomatization is excluded by an argument parallel to the one we used in the classical case (Theorem 4.4). The Strong Kleene satisfaction relation is of complexity Δ^1_1. Thus, if there were an $\mathbb{N}$-categorical axiomatization of the least fixed point version of Kripke's theory, we would have a Δ^1_1 definition of the least fixed point – this, however, is impossible by Fact 4.3.

§5. Discussion. Tarski's work on truth initiated two paradigms: the semantic and the axiomatic approach to truth. However, already in his *Concept of Truth* we find that they stimulate one another (Section §2). We believe that generally, formal work on truth has benefited from this interplay of semantics and axiomatics. In the present paper we focused on one aspect of it and asked when an axiomatic theory captures a semantic theory of truth. More precisely, under which conditions has a theorist succeeded in axiomatizing a given semantic theory?

Without attempting to give a conclusive answer we examined three possible such criteria of adequacy. Although initially plausible, both similarity (3.1) and proof theoretic strength (3.2) proved to have their limitations. In the remainder of the paper we therefore concentrated on $\mathbb{N}$-categoricity as a plausible candidate (3.3). We applied it to the question of axiomatizing Kripke's semantic theory of truth (Section §4) and collected positive, as well as, limitative results. On the one hand, we found that no $\mathbb{N}$-categorical first-order axiomatization can be obtained for all evaluation schemes we have discussed (Theorems 4.4 and 4.9) if we are only interested in the minimal fixed point. On the other hand, we have found that the supervaluational theories of truth do not allow for an $\mathbb{N}$-categorical axiomatization, even if all we ask for is truth in *some* fixed point (Theorem 4.11).

What are we to make of our findings? We expect that the conclusions drawn will be guided by an author's broader outlook on truth, and consider some possible responses. First, if $\mathbb{N}$-categoricity is accepted as a necessary condition for an adequate axiomatization, the negative results of Section §4 are problematic for some theories of truth. On the axiomatic side, friends of KFB or VF are put under pressure. On the semantic side, authors who think truth ought to be *grounded*, in the sense of Kripke, or authors who prefer a supervaluational approach to truth, may find it troublesome that there is no adequate axiomatization of their preferred semantic theory.[39]

Second, $\mathbb{N}$-categoricity may be considered to be a merely desirable feature. From this point of view, our results shows that KF, WKF and PKF are more closely linked to their corresponding semantic theory than the theory VF is to the supervaluational fixed points, or KFB to the least Strong Kleene fixed point.

Finally, even for authors who do not sympathize with the criterion of $\mathbb{N}$-categoricity our findings may still be found interesting, to the extent that they distinguish between the various existent axiomatic systems each of which claims to axiomatize a semantic theory of Kripkean design. Whereas we do not claim that grounded truth or supervaluational fixed points cannot be axiomatized nor to have shown that VF and KFB fail to capture their intended models, we think that our findings shift the burden of proof and require the friend of these axiomatic theories to explicate in what sense an axiomatization is supposed to capture the semantic theory.

[39] Note that Leitgeb (2005) also identifies the collection of grounded sentences with the least fixed point of his *dependence* jump. As noted by Leitgeb, this least fixed point is Π^1_1-complete, such that our limitative result 4.9 carries over (Leitgeb, 2005, p. 190).

Yet, these positive aspects of $\mathbb{N}$-categoricity are relativized by the fact that it needs supplementation by other criteria of adequacy. To realize that $\mathbb{N}$-categoricity is not a sufficient criterion it is enough to recall from p. 266 that the theory TB is $\mathbb{N}$-categorical with respect to the Tarski truth set, but was considered inadequate with respect to his semantic theory by Tarski himself (see p. 258).

Independently of whether $\mathbb{N}$-categoricity is the best criterion of adequacy available, we believe that the results of Section §4 shed light on the interplay of axiomatics and semantics. If we see ourselves in the tradition outlined in Section §2, and move back and forth between a semantic and an axiomatic theory of truth, we should seek balance between the axiomatic theory and the semantic construction. In the light of our findings we propose to view $\mathbb{N}$-categoricity as one way of rendering precise this intuitive thought. In a nutshell, our findings show that for an $\mathbb{N}$-categorical axiomatization to be possible the complexity of the satisfaction relation and the complexity of the specific semantic property of the intended interpretations of the truth predicate according to the semantic theory, e.g. the property of being a fixed-point, have to agree. If we apply this idea of a balance between the axiomatic and the semantic theory to this very observation, our semantic theory should not be too complicated. Moreover, if we understand $\mathbb{N}$-categoricity as a criterion that guarantees balance in complexity between the axiomatic and the semantic theory, then the criterion seems to compatible with the other criteria we discussed because balance of complexity alone cannot be the only criterion for a successful axiomatization. As matter of fact, this seems to be a desirable outcome given that all the criteria are insufficient considered in isolation.

We conclude by outlining what we believe to be promising routes of future work, keeping in mind the idea that axiomatic and semantic approaches should be in balance. One way to achieve balance may be by restricting the complexity of the model constructions, for example by seeking Kripke jumps in a supervaluational spirit that allow for a Δ^1_1 definition. This line of research would also lead us back from axiomatizations to semantic constructions and thus stand in the tradition outlined in Section §2. However, it may well be that the agreement of complexity has not to be exactly at the level indicated by $\mathbb{N}$-categoricity. So another way to achieve balance also for supervaluational or minimal fixed points may be to lift the complexity of the satisfaction relation, possibly by using a nonclassical satisfaction relation.

Laying the basis for such future research, we hope that the present paper contributes to a better understanding of the interplay of axiomatic and semantic work on truth.

§6. Acknowledgements. We would like to thank Timo Behringer, Graham Leigh, Dag Normann, Thomas Schindler and two anonymous referees for their helpful comments. Martin Fischer's and Johannes Stern's work was supported by the DFG research project "Syntactical Treatments of Interacting Modalities" and the Alexander von Humboldt Foundation. Volker Halbach's work was supported by the Arts & Humanities Research Council AH/H039791/1. Jönne Kriener's work was supported by the *European Research Council* and the University of Oslo.

BIBLIOGRAPHY

Burgess, J. P. (1986). The truth is never simple. *The Journal of Symbolic Logic*, **51**(3), 663–681.

Burgess, J. P. (2009). Friedman and the axiomatization of Kripke's theory of truth. Unpublished manuscript, delivered at a conference in honour of the 60th birthday of Harvey Friedman at the Ohio State University, 14–17 May 2009.

Cain, J., & Damnjanovic, Z. (1991). On the weak Kleene scheme in Kripke's theory of truth. *The Journal of Symbolic Logic*, **56**(4), 1452–1468.

Cantini, A. (1989). Notes on formal theories of truth. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, **35**(2), 97–130.

Cantini, A. (1990). A theory of truth formally equivalent to ID_1. *Journal of Symbolic Logic*, **55**, 244–258.

Cantini, A. (1996). *Logical Frameworks for Truth and Abstraction*. Amsterdam: Elsevier Science Publisher.

Davidson, D. (1990). The structure and content of truth. *The Journal of Philosophy*, **87**(6), 279–328.

Davidson, D. (1996). The folly of trying to define truth. *The Journal of Philosophy*, **93**, 263–278.

Feferman, S. (1991). Reflecting on incompleteness. *The Journal of Symbolic Logic*, **56**, 1–47.

Feferman, S., & Sieg, W. (1981). Inductive definitions and subsystems of analysis. In Dold, A. and Eckmann, B., editors. *Iterated Inductive Definitions and Subsystems of Analysis: Recent Proof-Theoretical Studies*, *Lecture Notes in Mathematics*, Vol. 897. Berlin: Springer, pp. 16–77.

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

Fujimoto, K. (2010). Relative truth definability of axiomatic truth theories. *Bulletin of Symbolic Logic*, **16**(3), 305–344.

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

Gupta, A., & Belnap, N. (1993). *The Revision Theory of Truth*. Cambridge, MA: MIT Press.

Halbach, V. (1994). A system of complete and consistent truth. *Notre Dame Journal of Formal Logic*, **35**, 311–327.

Halbach, V. (2014). *Axiomatic Theories of Truth* (second edition, revised ed.). Cambridge: Cambridge University Press.

Halbach, V., & Horsten, L. (2006). Axiomatizing Kripke's theory of truth. *The Journal of Symbolic Logic*, **71**, 677–712.

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

Horsten, L., Leigh, G. E., Leitgeb, H., & Welch, P. (2012). Revision revisited. *The Review of Symbolic Logic*, **5**(04), 642–664.

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

Kripke, S. (1975). Outline of a theory of truth. *The Journal of Philosophy*, **72**, 690–716.

Leitgeb, H. (2005). What truth depends on. *Journal of Philosophical Logic*, **35**, 155–192.

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

McGee, V. (1991). *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth*. Indianapolis, Cambridge: Hackett.

Patterson, D. (2012). *Alfred Tarski: Philosophy of Language and Logic*. Palgrave New York: Macmillan.

Pohlers, W. (2009). *Proof Theory: The First Step into Impredicativity*. Berlin: Springer.

Rogers, H. (1967). *Theory of Recursive Functions and Effective Computability*. New York: McGraw-Hill.

Simpson, S. G. (2009). *Subsystems of Second Order Arithmetic* (Second edition). Perspectives in Logic. Cambridge: Cambridge University Press.

Tarski, A. (1956). The concept of truth in formalized languages. In Corcoran, J., editor. *Logic, Semantics, Metamathematics* (1983, second edition). Indianapolis: Hackett Publishing Company.

Turner, R. (1990b). *Truth and Modality*. London: Pitman.

Welch, P. D. (2014). The complexity of the dependence operator. *Journal of Philosophical Logic*. Online First DOI 10.1007/s10992-014-9324-8.

MARTIN FISCHER
LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
MUNICH CENTER FOR MATHEMATICAL PHILOSOPHY
GESCHWISTER-SCHOLL-PLATZ 1
80539, MÜNCHEN, GERMANY
E-mail: M.Fischer@lrz.uni-muenchen.de

VOLKER HALBACH
NEW COLLEGE
OXFORD, OX1 3BN, UK
E-mail: volker.halbach@philosophy.ox.ac.uk

JÖNNE KRIENER
DEPARTMENT OF PHILOSOPHY, BIRKBECK COLLEGE
LONDON, WC1E 7HX, UK
E-mail: jkriener@runbox.com

JOHANNES STERN
LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
MUNICH CENTER FOR MATHEMATICAL PHILOSOPHY
GESCHWISTER-SCHOLL-PLATZ 1
80539, MÜNCHEN, GERMANY
E-mail: Johannes.Stern@lrz.uni-muenchen.de