Supervaluation-Style Truth Without Supervaluations[*]

Johannes Stern
Department of Philosophy
University of Bristol

johannes.stern@bristol.ac.uk

October 11, 2017

Abstract

Kripke's theory of truth is arguably the most influential approach to self-referential truth and the semantic paradoxes. The use of a partial evaluation scheme is crucial to the theory and the most prominent schemes that are adopted are the strong Kleene and the supervaluation scheme. The strong Kleene scheme is attractive because it ensures the compositionality of the notion of truth. But under the strong Kleene scheme classical tautologies do not, in general, turn out to be true and, as a consequence, classical reasoning is no longer admissible once the notion of truth is involved. The supervaluation scheme adheres to classical reasoning but violates compositionality. Moreover, it turns Kripke's theory into a rather complicated affair: to check whether a sentence is true we have to look at all admissible precisification of the interpretation of the truth predicate we are presented with. One consequence of this complicated evaluation condition is that under the supervaluation scheme a more proof-theoretic characterization of Kripke's theory becomes inherently difficult, if not impossible. In this paper we explore the middle ground between the strong Kleene and the supervaluation scheme and provide an evaluation scheme that adheres to classical reasoning but retains many of the attractive features of the strong Kleene scheme. We supplement our semantic investigation with a novel axiomatic theory of truth that matches the semantic theory we have put forth.

1. Introduction

In his seminal *Outline of a Theory of Truth*, Kripke (1975) proposed a semantic account of truth, which remains one of the most prominent approaches to self-referential truth and the semantic paradoxes. Kripke's idea was to work with models for an arithmetical language with the truth predicate, $\mathcal{L}_T$, in which the truth predicate was only partially defined.[1] The starting point would be a model of the base language together with a (possibly empty) set of sentences that have already been declared true. Then using an appropriate partial

[*] Forthcoming in *Journal of Philosophical Logic*.

[1] The choice of an arithmetical base language and theory is not essential to the proposal. We could work in an alternative framework and language as long as a sufficiently rich theory of syntax is available. Throughout the paper we assume $\mathcal{L}_T$ to be a standard first-order arithmetical language that contains an additional unary predicate—the truth predicate. $\mathcal{L}_T$ may also, in addition to S, $+$ and $\times$, contain further function symbols for certain primitive recursive functions.

evaluation scheme we collect sentences that are true under this initial interpretation of the truth predicate. These sentences then make up the new interpretation of the truth predicate. If we have chosen our initial interpretation of the truth predicate wisely the initial interpretation will be a subset of the new interpretation of the truth predicate. We may then continue this procedure until we eventually reach a point in the process in which no new sentences enter the interpretation of the truth predicate. Such an interpretation of the truth predicate is called a fixed point and contains all the sentences that are true under the very interpretation itself, given the chosen evaluation scheme. These fixed points display some very attractive, truth-like features and are thought to be suitable (candidate) interpretations of the truth predicate. However, depending on which partial evaluation scheme is assumed, these fixed points—and ultimately the resulting notion of truth—will have very different characteristics. Among the suitable partial evaluation schemes, two schemes stand out: the strong Kleene scheme and the supervaluation scheme.[2,3]

The strong Kleene scheme provides us with truth tables for three truth values and thereby allows us to compute the truth value of a complex sentence by examining the truth value of all its subsentences: once we know the the truth values of all the subsentences of a sentence we know the truth value of the sentence itself. In this sense the scheme is fully compositional.[4] The compositionality of the strong Kleene scheme gives the process by which we arrive at suitable interpretations of the truth predicate a constructive and transparent flavor since, in general, we move from simple sentences in the interpretation of the truth predicate to more complex ones. Moreover, due to the compositionality of the strong Kleene scheme the construction process will also lead to a compositional notion of truth, that is, whether a sentence is true in the object-linguistic sense, only depends on whether its subsentences are true or false (or neither) in the object-linguistic sense.[5] Since compositionality is thought to be a key feature of the notion of truth this outcome is highly desirable. More generally, we take it that the evaluation of a sentence should be as transparent as possible and should proceed, as far as possible, via the compositional structure, that is the built-up, of a sentence. Within the context of Kripke's theory of truth the strong Kleene scheme is the strongest available partial scheme that fully subscribes to this maxim.

However, the strong Kleene scheme has one major drawback. Logical truths or, more generally, *penumbral truths*, that is, sentences that are true because of the logical relations that hold among the sentences of the language, will not alway be true under the strong Kleene

[2] As a matter of fact there is not one particular supervaluation scheme but rather a family of different such schemes. For the sake of this introduction we ignore this complication and treat them as one. For the most important supervaluation schemes for theories of truth see our Section 3.1, McGee (1991),Burgess (1986) or Field (2008).

[3] Further partial evaluation schemes have been discussed as well but never really caught on to the same extent. See, e.g., Martin and Woodruff (1975), Fujimoto (2010) and Field (2008).

[4] Compositionality is sometimes taken to imply that truth, or the satisfaction relation, commutes with all logical connectives. However, since the strong Kleene scheme assumes three truth values but is formulated in a classical metatheory this will not generally hold for this scheme. While the scheme commutes with conjunction, disjunction and the quantifiers it does not commute with negation: if a sentence is not true in the metalinguistic sense according to the strong Kleene scheme this does not imply that its negation is true in the metalinguistic sense according to the strong Kleene scheme. The sentence may be neither true nor false. From this point of view the strong Kleene scheme is perhaps not fully compositional. But this notion of compositionality, which ties the idea of compositionality to the commutation of truth with the logical connectives, is at best a derived notion, which seems to be acceptable for classical logic but deeply misguided when applied to partial evaluation schemes.

[5] Throughout this paper we say that a sentence is false iff its negation is true.

scheme.[6] This is problematic because there seems to be a strong intuition that a disjunction such as *"all of Nixons assertions about Watergate are false or it is not the case that all of Nixons assertions about Watergate are false"* is true, but in the strong Kleene scheme if neither disjunct receives a classical truth value, this disjunction will not be true, instead it will lack a classical truth value. In contrast, penumbral truths will always be true under the supervaluation scheme and this feature was one of the motivations for introducing the scheme.[7] The supervaluation scheme is not based on truth tables for three truth values. Rather, it considers classical extensions of a partial model, so-called precisifications. Precisifications are models in which more semantic information is provided than in the partial model we have started out from. In the case of truth this means that we consider classical models in which the interpretation of the truth predicate extends the interpretation of the truth predicate in the partial model. Since we are only considering classical precisifications we are guaranteed that all logical truths will be true under the supervaluation scheme. More generally, all sentences that are a classical consequence of sentences that are true in all the precisifications that are considered will also be true under the evaluation scheme, independently of whether their subsentences always receive the same classical truth value. Moreover, despite the abstract sounding truth conditions, there is an intuitive rationale to the supervaluation scheme: a sentence is true according to the supervaluation scheme iff it is true according to all (classical) ways of making our current understanding of the truth predicate more precise. However, this way of evaluating sentences is also rather complicated and highly intransparent: instead of computing the truth value of a sentence by appeal to the truth value of its subsentences the scheme looks at all candidate interpretations of the truth predicate to determine whether a sentence is true. It is thus not surprising that the process by which we arrive at suitable interpretations of the truth predicate is highly intransparent under the supervaluation scheme and also less constructive than in the case of the strong Kleene scheme. Moreover, since the evaluation of a sentence does not proceed via its compositional structure, the notion of truth Kripke's theory of truth based on the supervaluation scheme gives rise to is non-compositional.

The complicated nature of the supervaluation scheme also makes it very difficult to provide proof-theoretic characterizations of Kripke's theory of truth based on the supervaluation scheme. For the strong Kleene case there are interesting proof-theoretic characterizations both in classical but also in strong Kleene logic: for classical logic we have the theory KF (Kripke-Feferman) whereas for strong Kleene logic this would be the theory PKF (Partial Kripke-Feferman).[8] While it is impossible to axiomatize a given fixed point directly, it is possible to characterize the lattice of fixed points relative to the natural number structure: the interpretations of the truth predicate of KF (PKF) are exactly the fixed points of Kripke's theory of truth based on the natural number structure. This criterion is called $\mathbb{N}$-categoricity in Fischer et al. (2015) where it is also shown that there can be no $\mathbb{N}$-categorical axiomatization of Kripke's theory of truth based on the supervaluation scheme in classical logic. Nor is it possible to provide a proof-theoretic characterization of Kripke's supervaluational theory of truth in some "supervaluational logic", at least if understood in a straightforward way: as Kremer and Kremer (2003) and Kremer and Urquhart (2008) show there is no such thing

[6] The term *penumbral connections* was introduced by Fine (1975) in connection to vagueness.

[7] See van Fraassen (1968); Van Fraassen (1969) and Fine (1975) for discussion in connection to non-denoting singular terms and vagueness. See McGee (1991) for an explicit appeal to this type of motivation in the case of truth.

[8] See Halbach (2011) for an exposition of both theories.

as a "supervaluational logic" matching the supervaluation scheme at play. This absence of interesting proof-theoretic characterizations supports our claim that the supervaluation scheme is intransparent and to a certain extent even mysterious: it cannot be illuminated by a proof theory that goes alongside.[9]

So we find ourselves in some sort of dilemma: on the one hand we have Kripke's theory of truth based on the strong Kleene scheme, which is based on a simple and transparent evaluation scheme, gives rise to a compositional notion of truth and allows for proof-theoretic characterizations but fails to declare logical truths to be true. On the other hand, we have Kripke's theory based on the supervaluation scheme, which declares all penumbral truths and, in particular, all logical truths to be true. But this version of the theory is based on a rather complicated scheme, assumes a non-compositional notion of truth and does not allow for a proof-theoretic characterization. So both versions of Kripke's theory of truth have some very desirable features and thus the immediate question arises whether we can compromise between the two. Unfortunately, there is a limit to what we can do since there is a genuine tension between compositionality and the intuition that penumbral truths should be true under the evaluation scheme at stake. If we admit the latter, then a disjunction in which both disjuncts do not have a classical truth value can either be true or indeterminate.[10] As a consequence no scheme that allows for classical logical truths can compute the truth value of a sentence by appeal to the truth value of its subsentences and therefore the scheme cannot be compositional. While we cannot have both compositionality and penumbral truths, it is possible, as we shall see, to narrow the gap between compositional schemes and schemes that allow for penumbral truths. In particular, it is possible to provide evaluation schemes that allow for penumbral truths while being much simpler and more transparent than the supervaluation scheme. Such an alternative evaluation scheme will work like the strong Kleene scheme in most cases, that is, it will usually evaluate sentences by appeal to its subsentences—except for cases of penumbral truths. For the cases of penumbral truths, the scheme should exploit the logical relations that hold among the relevant sentences, as it is done in the supervaluation scheme. However the idea would be that this can be done in a much simpler and transparent way, so to minimize the gap between the compositional strong Kleene scheme and this scheme. As a consequence of the newly gained simplicity the scheme should lend itself to a version of Kripke's theory of truth that allows for a neat proof-theoretic characterization. The aim of this paper will be to provide such an alternative evaluation scheme together with a list of principles of truth, which jointly axiomatize Kripke's theory of truth based on this novel scheme.

1.1 Structure of the Paper

In the next section we take a fresh look at the supervaluation scheme and investigate how it is precisely that the scheme accounts for penumbral truths or, more generally, penumbral connections between sentences. As we shall see, the supervaluation scheme accounts for the penumbral connections between sentences by some form of second-order consequence

[9] The theory VF developed by Cantini (1990) is intended to capture aspects of Kripke's supervaluational theory of truth. However, the theory cannot distinguish supervaluational truth from revision theoretic truth and in this sense falls short from providing a proof-theoretic account of Kripke's supervaluational theory of truth.

[10] See, e.g., Fine (1975) for remarks along these lines. Note that a sentence is indeterminate iff it does not receive a classical truth value. For a true disjunction with indeterminate disjuncts the Nixon-example above does the trick. For the indeterminate case replace one disjunct of the Nixon-example by, e.g., the sentence *"this sentence is not true"*.

relation. We argue that this is too strong and that we should instead account for penumbral connections between sentences using the first-order consequence relation. This will lead to a simpler and more transparent scheme in which penumbral truths come out true as it was intended. In the section thereafter we provide a more rigorous discussion of the different evaluation schemes at play, that is, the supervaluation schemes, the strong Kleene scheme and the novel supervaluation-style truth schemes. Section 4 shows that the supervaluation scheme and the supervaluation-style truth scheme lead to the same notion of grounded truth, that is, both schemes have the same minimal fixed point. The result is established by appeal to a characterization of the minimal supervaluation fixed point via an infinitary sequent calculus introduced by Cantini (1990). In the remainder of the paper we turn to more proof-theoretic aspects. In Section 5 we introduce the truth theory of Inductive Truth (IT), which is intended to capture salient aspects of Kripke's theory of truth based on the supervaluation-style truth scheme. To substantiate this claim we show IT to be an $\mathbb{N}$-categorical axiomatization of the aforementioned version of Kripke's theory of truth, i.e., the fixed points of the novel supervaluation-style truth scheme are exactly the suitable interpretations of the truth predicate of IT relative to the standard model. Section 6 ends our proof-theoretic investigations by establishing, again using a result by Cantini (1990), that IT is proof-theoretically equivalent to ID_1. The paper closes with a short summary of our results and their relevance.

2. Consequence, Supervaluation and Penumbral Truths

In the introduction we motivated the supervaluation scheme chiefly by its ability to account for the truth of sentences that are true simply because of the logical relations that hold between their subsentences, or their logical relation to the sentences in the interpretation of the truth predicate. For example, the *tertium non datur* will always be true since it is a disjunction of a sentence and its negation. Similarly, if we allow for vague predicates, 'it is not the case that this coffee mug is blue and green' is true because the two color predicates are, by definition, mutually exclusive. The truth value of the complex sentence is hence independent of the truth value of its subsentences since the two color ascriptions of the sentence might not receive a classical truth value. So whereas the strong Kleene scheme only uses the compositional structure of sentences, that is their build-up, for their evaluation, the supervaluation scheme also uses the logical relations that hold among different sentences, that is, the supervaluation scheme also uses what one might call the *logical structure* a sentence is embedded in. This raises the obvious question of what we take this *logical structure* to be and how we should explicate the notion of *logical relation*. However, before we enter this discussion we should provide a precise definition of the supervaluation scheme (sv) in the context of Kripke's theory of truth:

$$(\mathbb{N}, S) \models_{\mathsf{sv}} \phi :\Leftrightarrow \forall S' \left(S \subseteq S' \,\&\, \Phi(S') \Rightarrow (\mathbb{N}, S') \models \phi \right)$$

$(\mathbb{N}, S)$ is a model for the language $\mathcal{L}_T$, where $\mathbb{N}$ is the natural number structure and thus interprets the arithmetical fragment of the language, while S is a set of (codes of) sentences that serves as the interpretation of the truth predicate. In partial evaluation schemes such as the supervaluation scheme it is common to provide not only an extension but also an antiextension for interpreting the truth predicate. The extension tells us which sentences are true whereas the antiextension tells us which sentences are false. However, we can omit explicit mention of the antiextension by taking it to consist of the negations of the sentences in the truth predicate's extension and we shall adopt this approach throughout this paper independently of the partial evaluation scheme under consideration.[11] In this version of the supervaluation scheme we do not evaluate sentences with respect to every precisification but only those that are *admissible*, i.e., which meet some *admissibility condition* $\Phi(S)$. For example, one basic admissibility condition requires a precisification not to contradict the sentences in the truth predicate's extension. As a matter of fact we shall only consider admissibility conditions that enforce the minimal requirement of the basic admissibility conditon, which we shall discuss in more detail when we discuss the supervaluation scheme vb in Section 3.1.

The above definition of the supervaluation scheme makes it difficult to understand our previous remarks that the supervaluation scheme uses the compositional and the logical structure in evaluating sentences. Let us consider a simple disjunction $T\ulcorner\phi\urcorner \vee \neg T\ulcorner\psi\urcorner$ to illustrate these two components of the supervaluation scheme.[12] According to the supervaluation scheme there are two possible scenarios under which this disjunction is true in a given interpretation S: either (i) $S \models_{\mathsf{sv}} T\ulcorner\phi\urcorner$ or $S \models_{\mathsf{sv}} \neg T\ulcorner\psi\urcorner$, or (ii) $S \not\models_{\mathsf{sv}} T\ulcorner\phi\urcorner$ and $S \not\models_{\mathsf{sv}} \neg T\ulcorner\psi\urcorner$ but $S \models_{\mathsf{sv}} T\ulcorner\phi\urcorner \vee \neg T\ulcorner\psi\urcorner$. Assuming some reasonable admissibility condition (i) will hold iff either $\#\phi \in S$ or $\#\neg\psi \in S$. In case (ii), the interpretation S together with the admissibility condition Φ has to imply that in all admissible precisifications S': $\#\psi \in S'$ implies $\#\phi \in S'$. Now in case (i) the evaluation of the disjunction proceeds via the compositional structure of the sentence. Indeed case (i) provides exactly the standard truth conditions of $T\ulcorner\phi\urcorner \vee \neg T\ulcorner\psi\urcorner$ under the strong Kleene scheme. In contrast, in case (ii) the compositional structure of the sentence plays at best an indirect role in the evaluation of the sentence. Whether the disjunction will be true depends on which precisifications we consider and, ultimately, what logical relation the disjunction bears to the members of S. From this perspective the supervaluation scheme seems to be somewhat disjunctive in character: on the one hand the scheme checks whether a given sentence is true because of is compositional structure. On the other hand, it checks whether the sentence is true due to the logical relations it bears to the members of S. These admittedly rather vague observations and remarks generalize and can be made fully precise by providing an alternative, equivalent definition of the supervaluation scheme, which we will explain informally after explaining some terminology:[13]

$$S \models_{\mathsf{sv}} \phi :\Leftrightarrow \exists G\Big(\forall\psi \in G(S \models_{\mathsf{sk}} \psi)\,\&\,\forall S'\big(S \subseteq S'\,\&\,\Phi(S') \Rightarrow (\forall\psi \in G(S' \models \psi) \Rightarrow S' \models \phi)\big)\Big)$$

[11] More precisely, the antiextension of the truth predicate S^- is defined relative to the extension S in the following way:

$$S^- := \{\#\psi : \exists\phi(\phi \doteq \neg\psi\,\&\,\#\phi \in S\} \cup \{\#\neg\psi : \#\psi \in S\,\&\,\neg\exists\phi(\psi \doteq \neg\phi)\}.$$

By $\#\phi$ we denote the Gödel number of a sentence ϕ. In the context of Kripke's theory of truth there is no loss of generality by taking the antiextension to be defined because for all customary evaluation schemes the Kripkean process of constructing the interpretation of the truth predicate guarantees the extension and the antiextension to be interdefinable in the above way.

[12] $\ulcorner\phi\urcorner$ is a name of the sentence ϕ. Throughout this paper we take $\ulcorner\phi\urcorner$ to be the numeral of the Gödel number of ϕ. The Gödel number of ϕ will be denoted, as mentioned, by $\#\phi$.

[13] The two definiens are equivalent for consistent sets S only but the supervaluation scheme is usually only defined for consistent sets. So there is no loss of generality. The argument works by showing that for all S' such that $S \subseteq S'$ and $S^- \cap S' = \emptyset$ and all sentences ϕ of $\mathcal{L}_T$:

$$S \models_{\mathsf{sk}} \phi \Rightarrow S' \models \phi.$$

Since we only work in the standard model of arithmetic we omit from now on explicit mention of the natural number structure when denoting the models of $\mathcal{L}_T$. That is, $S \models \phi$ is short for $(\mathbb{N}, S) \models \phi$. Accordingly, $S \models_{\mathsf{sk}} \phi$ is short for $(\mathbb{N}, S) \models_{\mathsf{sk}} \phi$, that is, ϕ is true under the strong Kleene scheme in the model $(\mathbb{N}, S)$.[14] At first sight this alternative definition of the sv-scheme might seem terribly opaque but it can be summarized in a simple way: a sentence ϕ is true under the interpretation S in the supervaluation scheme iff ϕ follows in the admissible precisification from the sentences that are true under the interpretation S in the strong Kleene scheme (i.e. the sentences in G). More precisely, ϕ is true according to the supervaluation scheme in a given interpretation S iff the sentences that are true in S according to the strong Kleene scheme jointly imply ϕ in the class of classical (standard) models $\mathcal{M}_S = \{S' : S \subseteq S' \,\&\, \Phi(S')\}$. It is worth noting that if a sentence is true according to the strong Kleene scheme it will be true according to the supervaluation scheme simply because it will always imply itself in $\mathcal{M}_S$. With this remark in mind it is now clear how the compositional structure of a sentence and the logical structure the sentence is embedded in play together in the supervaluation scheme: either a sentence is true according to its compositional structure, that is, the strong Kleene scheme, or it is implied in $\mathcal{M}_S$ by the sentences that are true according to the strong Kleene scheme under the interpretation S.

It is this second condition, which guarantees that penumbral truths will always be true under the supervaluation scheme independently of the truth values of its subsentences. This tells us that this second condition, i.e. truth preservation in the model class $\mathcal{M}_S$, is a sufficient condition for accounting for penumbral truths. But is it a necessary condition? To answer this question recall that penumbral truths are sentences that are true simply because of the *logical relations* that hold between the sentences of the language. Our analysis of the supervaluation scheme then suggests that the relevant notion of *logical relation* is truth preservation in $\mathcal{M}_S$, that is, truth preservation in a class of standard models with varying interpretations of the truth predicate. But is this the right explication of *logical relation* or does the supervaluation scheme take us beyond penumbral truths? This depends on how strict an understanding of *logical* is employed. The consequence relation at play, that is truth preservation with respect to the class $\mathcal{M}_S$, goes beyond the consequence relation of full second-order logic. The consequence relation of full-second order logic fixes the standard model, whereas the consequence relation used in the supervaluation scheme further restricts the class of relevant model to the class $\mathcal{M}_S$, i.e., it only considers models that improve on the interpretation of the truth predicate that is currently assumed. In other words, the second-order consequence relation fixes the admissible model(s) for the language without the truth predicate whereas the consequence relation employed in the supervaluation scheme goes one step further by telling us what the admissible models for the language with the truth predicate are. It is already questionable whether full second-order logic and thus second-order consequence remains within the bounds of logic. In fact, many people agree that second-order logic goes beyond the bounds of logic because of its ontological commitment and its expressive strength—indeed it is often thought to be mathematical rather than logical in character. As we have mentioned, the consequence relation at play in the supervaluation scheme goes beyond second-order consequence. It uses all information available to relate sentences, rather than just simple *logical structure*. In light of these remarks it seems reasonable to search for an alternative condition accounting for penumbral truths, that is, a condition that is based on a stricter understanding of the term *logical*. By this we do not mean to discredit

[14] See Halbach (2011) for a definition of the strong Kleene satisfaction relation.

the supervaluation scheme as unreasonable or unmotivated. Indeed the claim is *not* that the supervaluation scheme gets things wrong. Rather the claim is that we do not need to appeal to this strong a consequence relation and, consequently, the supervaluation scheme to account for penumbral truths. In other words it seems that truth preservation in $\mathcal{M}_S$ is not a necessary condition for accounting for penumbral truths. Of course, there may be good *other* reasons to opt for a strong consequence relation and the supervaluation scheme. However, as far as penumbral truths are concerned there is room for an alternative to the supervaluation scheme, that is, a simpler and more transparent scheme, which remains somewhat closer to the strong Kleene scheme.

On a stricter understanding of *logical relation*, the appropriate consequence relation seems to be the one of first-order logic and, for most examples used to motivate the supervaluation schemes, first-order consequence proves sufficient—indeed we are not aware of a motivating example for which second-order consequence or the even stronger $\mathcal{M}_S$-consequence relation is required. This suggests that if we are interested in penumbral truths only, accounting for penumbral truths via first-order consequence is the way forward. First-order consequence is, in many respects, much simpler than the alternative condition of the supervaluation scheme. Whereas second-order and supervaluational consequence requires universal quantification over sets of natural numbers, first-order consequence can be spelled out by appeal to existential quantification over natural numbers. In other words the latter relation is recursively enumerable. Moreover, our alternative definition of the supervaluation scheme provides us with a strategy for constructing an alternative evaluation scheme that is inspired by the supervaluation scheme but uses the first-order consequence relation to account for penumbral truth. The scheme will be called *supervaluation-style truth scheme* (sst) where $\models^1_{\mathcal{L}_T}$ denotes the first-order consequence relation in the language with the truth predicate:

$$S \models_{\mathsf{sst}} \phi :\Leftrightarrow \exists G \left(\forall \psi \in G (S \models_{\mathsf{sk}} \psi) \,\&\, G \models^1_{\mathcal{L}_T} \phi \right)$$

Like the supervaluation scheme the new scheme collects the sentences, which (i) are true according to the strong Kleene scheme under the interpretation of the truth predicate S and (ii) whatever follows from these sentences. However, in contrast to the supervaluation scheme this second condition is no longer considered to be truth preservation in the class $\mathcal{M}_S$ but simple first-order consequence, that is, truth preservation in all models of the language. Despite this difference between the supervaluation scheme and the supervaluation-style truth scheme, we take it that the evaluation schemes are very close in spirit, both collect some classical consequences of the sentences that are true according to the strong Kleene scheme. They only disagree on exactly what these consequences are, or rather, how these consequences are best collected.

It may be argued that moving to first-order consequence comes at a cost: we can no longer rule out certain interpretations as being unprincipled, that is, we can no longer consider *admissible* interpretations only. But this is only partly true. As long as the admissibility condition can be expressed via a first-order formula we can simply add this condition to the set of strong Kleene truths, i.e. the set G. For example, according to one particular supervaluation evaluation scheme we shall be discussing later we should only consider consistent interpretations of the truth predicate, that is, interpretations such that for no sentence ϕ, ϕ and $\neg\phi$ are in the interpretation of the truth predicate. Now, if, for every sentence ϕ, we add the sentence $\neg(T\ulcorner\phi\urcorner \wedge T\ulcorner\neg\phi\urcorner)$ to the set G this will have the effect that we only consider consistent interpretations of the truth predicate since only models with a consistent interpretation of the truth predicate will be possible models of the set G. Of course, the supervaluation-style truth scheme cannot emulate admissibility conditions that cannot be expressed in first-order languages. But allowing for admissibility conditions that cannot be expressed in first-order languages would undermine the very motivation of the supervaluation-style truth scheme and thus this limitation seems to be well motivated and unproblematic.

If we have chosen an interpretation of the truth predicate for which the Kripkean process based on the sst-scheme terminates in a fixed point, then this means that either a sentence is true because of the truth value of its subsentences or because it logically follows from the sentences in the interpretation of the truth predicate. The scheme thus works compositionally for all sentences except for some penumbral truths, namely those that are *ungrounded* in the Kripkean sense according to the strong Kleene scheme. For these ungrounded penumbral truths the scheme uses the underlying *logical structure* in form of the first-order consequence relation. As we have mentioned first-order consequence, as opposed to second-order consequence, is a rather simple and transparent condition and can be expressed within a first-order language. This means that we can actually state what is happening when evaluating a sentence with respect to its compositional structure has proven insufficient. At least from the first-order perspective, nothing similar is possible for the supervaluation scheme: if the evaluation of a sentence via its compositional structure has failed, then the evaluation process becomes opaque and mysterious. We simply do not know why a given sentence is true or false.

As we have just seen, even if the evaluation of a sentence via its compositional structure fails, the evaluation process of the supervaluation-style truth scheme remains, at least to a certain extent, transparent. We take this to suggest that the failure of compositionality in the supervaluation-style truth scheme is somewhat less drastic than in the supervaluation scheme. The gap between the strong Kleene scheme qua compositional scheme and the supervaluation-style truth scheme can be bridged via a simple first-order condition. This is not possible for the supervaluation scheme. As a matter of fact, it is precisely for this reason that Kripke's theory of truth based on the supervaluation-style truth scheme will prove more amenable to a proof-theoretic characterization than its supervaluational counterparts. However, before we turn to proof-theoretic questions we investigate the new scheme in detail and compare it to the strong Kleene scheme and, especially, the supervaluation scheme.

3. Evaluation Schemes, Kripke Jumps, and Fixed Points

In the introduction to this paper, we have already pointed out that Kripke's theory of truth can be formulated using various evaluation schemes and in this section we introduce the evaluation schemes that are crucial for our investigation: the relevant supervaluation schemes, the strong Kleene scheme and the supervaluation-style truth schemes.

Each evaluation scheme e gives rise to a jump operation $\mathcal{J}_e : P(\omega) \to P(\omega)$ with

$$\mathcal{J}_e(X) = \{\#\phi : X \models_e \phi\}$$

for $X \subseteq \omega$. A jump operation $\mathcal{J}_e$ is called a Kripke jump iff it is monotone, i.e. iff

$$X \subseteq Y \Rightarrow \mathcal{J}_e(X) \subseteq \mathcal{J}_e(Y).$$

The monotonicity of the Kripke jump guarantees that there will be fixed points, that is, there will be sets $S \subseteq \omega$ such that

$$\mathcal{J}_e(S) = S.$$

Among these fixed points, one particular fixed point stands out, namely, the minimal fixed point. The minimal fixed point can be obtained by iterative application of the jump operation to the empty set. In other words there will be an ordinal number α such that $\mathcal{J}_e^{\alpha}(\emptyset) = \mathcal{J}_e^{\alpha+1}(\emptyset)$, where by $\mathcal{J}_e^{\alpha}(S)$ we denote that the jump operator has been applied α-times starting from the set S.[15] Kripke's theory of truth can be understood as advocating either arbitrary fixed points as suitable interpretation of the truth predicate or the minimal fixed point only.[16] We will come back to this distinction in Section 4.

Before we turn to the different evaluation schemes we introduce some notation and terminology we shall be using. Throughout the paper we assume some reasonable coding scheme for the expressions of $\mathcal{L}_T$. For terminology we mostly follow Halbach (2011). We denote the Gödel number of an expression ζ by $\#\zeta$ and the numeral of $\#\zeta$ will be denoted by $\ulcorner\zeta\urcorner$. For a closed term t we write $t^{\mathbb{N}}$ to denote its interpretation in the standard model. We let the sets $\mathsf{Sent}_{\mathcal{L}_T}$ ("$\mathcal{L}_T$-sentences") and $\mathsf{Cterm}_{\mathcal{L}_T}$ ("$\mathcal{L}_T$-closed terms") represent themselves and drop the subscript when no confusion can arise. Quantification of the form $\forall s, t\, \phi$ is short for $\forall x, y(\mathsf{Cterm}(x) \wedge \mathsf{Cterm}(y) \rightarrow \phi(x, y))$. In most cases, if $\rhd$ is a syntactic operation we represent it by $\underset{\cdot}{\rhd}$. However, there are few exceptions to this rule: we represent the ternary substitution function by $x(s/t)$ where $x(s/t)$ is a name of the expression that results from replacing t by s in x. Also, we let $^{\circ}$ represent the function that takes codes of closed terms as arguments and provides their denotation as output. Finally, $\ulcorner\phi(\underset{\cdot}{t})\urcorner$ is short for $\ulcorner\phi(x)\urcorner(t/\ulcorner x\urcorner)$ where t is the code of a closed term, i.e., $\ulcorner\phi(x)\urcorner(t/\ulcorner x\urcorner)$ is the name of a formula in which the free variable has been replaced by the closed term with code t.

3.1 Supervaluation

So far we only introduced a generic supervaluation scheme in which we left the admissibility condition unspecified. From now on we consider two particular supervaluation schemes with specific admissibility conditions:

$$X \models_{\mathsf{vb}} \phi :\Leftrightarrow \forall Y(X \subseteq Y \,\&\, Y \cap X^{-} = \emptyset \Rightarrow Y \models \phi)$$
$$X \models_{\mathsf{vc}} \phi :\Leftrightarrow \forall Y(X \subseteq Y \,\&\, Y \in \mathsf{CONS} \Rightarrow Y \models \phi)$$

X^{-} is the antiextension of the truth predicate, which is defined as the set of negations of the sentences in X. CONS is the set of consistent sets of sentences.[17] The schemes vb and vc are the most prominent supervaluation schemes in connection to truth but there is a further popular scheme, which we will mostly neglect. This scheme is based on a stronger admissibility

[15] The monotonicity of $\mathcal{J}_e$ guarantees the existence of fixed points and the existence of a minimal fixed point: there are more ordinal numbers than sentences of the language so at some point we run out of sentences that we can add to the interpretation of the truth predicate—we have reached a fixed point. If we start from the empty set the fixed point we obtain must, by monotonicity, be the minimal fixed point.

[16] There is a further fixed point which is of particular interest, namely, the maximal intrinsic fixed point. The maximal intrinsic fixed point contains all sentences for which the negation of the sentence is in no fixed point of the jump operation at stake. So the construction of the fixed point does not involve any arbitrary semantic decisions.

[17] A set of sentences X is consistent if for no sentence ϕ: $\#\phi \in X$ and $\#\neg\phi \in X$.

condition which stipulates that Y must be a maximally consistent set, i.e. $Y \in \mathsf{MAXCONS}$.[18] We omit discussion of this latter supervaluation scheme because it is not classically sound, that is, it lacks the following important property:

Lemma 3.1 (Classical Soundness)**.** *Let* $X \in \mathsf{CONS}$*. Then for all* $\phi \in \mathcal{L}_\mathsf{T}$

(*i*) $\quad X \models_{\mathsf{vb}} \phi \Rightarrow X \models \phi;$

(*ii*) $\quad X \models_{\mathsf{vc}} \phi \Rightarrow X \models \phi.$

The supervaluation schemes give rise to the following supervaluation jump operations. Let $X \in \mathsf{CONS}$ and set

$$\mathsf{VB}(X) := \{\#\phi : X \models_{\mathsf{vb}} \phi\};$$
$$\mathsf{VC}(X) := \{\#\phi : X \models_{\mathsf{vc}} \phi\}.$$

Notice that the jump operation is only defined for consistent sets. Inconsistent sets would immediately lead to the trivialization of the jump operation.[19] The scheme vc considers fewer models than the scheme vb. As a consequence the jump operation VC collects more sentences than the jump operation VB, that is, for $S \in \mathsf{CONS}$, $\mathsf{VB}(S) \subseteq \mathsf{VC}(S)$. Indeed, if S is not only consistent but also a partial interpretation of the truth predicate, i.e. $S \cup S^- \neq \mathsf{Sent}_{\mathcal{L}_T}$, then $\mathsf{VB}(S) \subsetneq \mathsf{VC}(S)$ and, as a consequence, the two jump operations do not share any fixed points.[20] Such fixed points exist since it is easy to verify that both jump operations are monotone and that VB and VC are thus Kripke jumps.

3.2 Strong Kleene

We provide two alternative jumps, which deliver the same fixed points. We refrain from introducing the strong Kleene scheme in detail, since it is well known and may be found in several textbooks.[21] In this section we assume negation to be defined and correspondingly take $\neg T$ to be a primitive expression of the language—the falsity predicate. Later in this paper, once we turn to theories of truth, we refrain from this understanding of the negation symbol and in later sections the following definitions have to be supplemented by the obvious duals to the compositional clauses. The first jump operation is the strong Kleene jump SK

$$\mathsf{SK}(X) := \{\#\phi : X \models_{\mathsf{sk}} \phi\}$$

$\models_{\mathsf{sk}}$ is the usual strong Kleene satisfaction relation defined in our restricted language.[22] Again, it is well known and easy to verify that the strong Kleene jump is monotone and thus a Kripke jump in our sense.

[18] $\mathsf{MAXCONS}$ is the set of maximally consistent sets of sentences. A set of sentences X is maximally consistent if it is consistent and, in addition, it is maximal, that is for all sentences ϕ: $\#\phi \in X$ or $\#\neg\phi \in X$. The scheme was suggested by, e.g., Kripke (1975).

[19] Trivialization is immediate in the case of the VC jump. In case of the VB there is actually no trivialization but for no inconsistent set S we have $S \subseteq \mathsf{VB}(S)$ and thus all fixed points are reached starting from consistent sets.

[20] See Fischer et al. (2015) for a proof of this observation and further discussion of the different supervaluation schemes.

[21] See, for instance, McGee (1991) or Halbach (2011).

[22] See, e.g., Halbach (2011).

It is also well-established that the strong Kleene fixed points can be obtained via an arithmetically definable operator. Let $\xi(x, X)$ be the following formula:

(1) $x \in \mathsf{True}_0 \vee$

(2) $\exists y, z\, (x = (y \underset{\cdot}{\wedge} z) \wedge (y \in X \wedge z \in X)) \vee$

(3) $\exists y, z\, (x = (y \underset{\cdot}{\vee} z) \wedge (y \in X \vee z \in X)) \vee$

(4) $\exists y\, (x = \underset{\cdot}{\forall} vy \wedge \forall t(y(t/v) \in X)) \vee$

(5) $\exists y\, (x = \underset{\cdot}{\exists} vy \wedge \exists t(y(t/v) \in X)) \vee$

(6) $\exists t(x = \underset{\cdot}{T}\, t \wedge t^\circ \in X) \vee$

(7) $\exists t\, (x = \underset{\cdot}{\neg}\, \underset{\cdot}{T}\, t \wedge (\underset{\cdot}{\neg}\, t^\circ \in X \vee t^\circ \notin \mathsf{Sent}))$

$x \in \mathsf{True}_0$ in line 1 denotes that x is a true arithmetical literal. In our case this means that x is either the code of a true identity statement or the code of a true refutation of an identity statement. The operator defined by $\xi(x, X)$ is given by:

$$\Xi(S) := \{n \in \omega : \mathbb{N} \models \xi(x, X)[n, S]\}$$

By a folklore theorem we know that the two different jump operations have the same fixed points:[23]

Theorem 3.2. *For all $S \subseteq \omega$*

$$\mathsf{SK}(S) = S \Leftrightarrow \Xi(S) = S.$$

However, this theorem does not imply that the operators agree on all stages of the inductive definition. Rather the strong Kleene scheme gathers "more" sentences than Ξ, i.e. , for all $S \subseteq \omega$, $\Xi(S) \subseteq \mathsf{SK}(S)$.

The strong Kleene scheme is more restrictive than the supervaluational schemes, that is, it excludes more sentences. In particular, as we have discussed in the introduction of this paper not all logical truths of the language $\mathcal{L}_\mathsf{T}$ will be true under the the strong Kleene scheme. As a consequence, for consistent, partial interpretations of the truth predicate, the output of the strong Kleene jump is a proper subset of the outputs of the supervaluation jumps:

Lemma 3.3. *Let $S \in \mathsf{CONS}$ and $S \cup S^- \neq \mathsf{Sent}_{\mathcal{L}_T}$. Then*

$$\mathsf{SK}(S) \subsetneq \mathsf{VB}(S) \subsetneq \mathsf{VC}(S).$$

After presenting the well-know supervaluation schemes and the strong Kleene scheme we finally move to a precise formulation of our novel evaluation scheme—supervaluation-style truth.

3.3 Supervaluation-Style Truth

In Section 2 we introduced the main idea behind the supervaluation-style truth scheme. The idea was that a sentence is true according to the supervaluation-style truth scheme if it is a classical consequence of the sentences that are true according to the strong Kleene scheme:

$$S \models_{\mathsf{sst}} \phi :\Leftrightarrow \exists G \Big(\forall \psi \in G(S \models_{\mathsf{sk}} \psi)\, \&\, G \models^1_{\mathcal{L}_T} \phi\Big).$$

[23] See (Halbach, 2011, pp. 202-210) for a detailed exposition of this result.

We will modify this definition in two minor respects. First, since the first-order consequence relation is compact we can replace second-order quantification over sets of sentences by first-order quantification over individual sentences. Second, we are working in an arithmetical language and we want our notion of first-order consequence to include certain facts about arithmetic and, by the same token, certain facts about syntax theory. The idea is to assume some reasonable arithmetical theory in $\mathcal{L}_T$, say PAT (Peano Arithmetic in the extended language), as an additional premiss. This would yield the following definiens of the sst-scheme:

$$\exists\psi\big(S \models_{\mathsf{sk}} \psi \,\&\, \mathsf{PAT}, \psi \models^1_{\mathcal{L}_T} \phi\big).$$

Due to the completeness theorem for first-order logic we know that we can equivalently replace the first-order consequence relation by its syntactic counterpart and we shall do so in the official version of the scheme. The official version will come under the label supervaluational strong Kleene (ssk):

$$S \models_{\mathsf{ssk}} \phi :\Leftrightarrow \exists\psi\big(S \models_{\mathsf{sk}} \psi \,\&\, \mathsf{PAT} \vdash \psi \to \phi\big).$$

The scheme directly leads to a jump operation—the supervaluational strong Kleene jump (SSK), which, as in the case of the supervaluational jump operations, is defined for consistent sets of sentences only. For inconsistent sets of sentences trivialization would arise since the output would always be the set of all sentences. Let $S \in \mathsf{CONS}$ and set

$$\mathsf{SSK}(S) := \{\#\phi : S \models_{\mathsf{ssk}} \phi\}.$$

Now, as in the strong Kleene case but in stark contrast to the supervaluation case we can, using an arithmetical formula, define a second jump operator, which will have the same fixed points as the SSK-operator. We obtain the relevant arithmetical formula θ by replacing the strong Kleene satisfaction relation by the formula ξ we introduced in the previous section and the derivability condition by its arithmetization. Accordingly, let $\theta(x, X)$ be the formula

$$\exists y\,(\xi(y, X) \land \mathsf{Pr}_{\mathsf{PAT}}(y \dot{\to} x))\,.$$

As for the ssk-scheme, it might sometimes be helpful to recall that θ is equivalent to the following disjunctive condition:

$$\xi(x, X) \lor \exists y\,(\xi(y, X) \land \mathsf{Pr}_{\mathsf{PAT}}(y \dot{\to} x))\,.$$

The corresponding operator Θ is then defined as follows for $S \in \mathsf{CONS}$:

$$\Theta(S) := \{n \in \omega : \mathbb{N} \models \theta(x, X)[n, S]\}.$$

As we shall see, the SSK operator, which can be readily verified to be a Kripke jump, is closely related to the supervaluational VB-operator. But we can also define a supervaluation-style truth operator that relates to the alternative supervaluation scheme VC, which only considers consistent interpretations of the truth predicate as admissible precisifications. For the case of supervaluation-style truth this has the consequence that we need to add the assumption that the truth predicate is consistent to our sets of premises, i.e., we have to make sure that sentences of the form $\neg(T\ulcorner\phi\urcorner \land T\ulcorner\neg\phi\urcorner)$ are available premises when determining the relevant first-order consequences. We write $\mathsf{Con}(\phi)$ to denote the fact that ϕ is of the form

$\neg(Ts \wedge T \underset{\cdot}{\neg}\, t)$ where $t^{\mathbb{N}} = s^{\mathbb{N}}$. The scheme $\mathsf{ssk_c}$ (consistent supervaluational strong Kleene) is then defined as follows:

$$S \models_{\mathsf{ssk_c}} \phi \Leftrightarrow \exists\psi((S \models_{\mathsf{sk}} \psi \text{ or } \mathsf{Con}(\psi))\ \&\ \mathsf{PAT} \vdash \psi \rightarrow \phi)$$

and, unsurprisingly, the corresponding operator is given by

$$\mathsf{SSK_c}(S) := \{\#\phi : S \models_{\mathsf{ssk_c}} \phi\}$$

for $S \in \mathsf{CONS}$.

Parallel to the case of the basic supervaluation-style truth scheme we can define a matching operator Θ_c using an arithmetical formula. To this end let $\mathsf{con(x)}$ be the formula

$$\exists s, t\, (x = \ulcorner\neg(T\, \underset{\cdot}{s} \wedge T \underset{\cdot}{\neg}\, \underset{\cdot}{t}\,)\urcorner \wedge t^{\circ} = s^{\circ})$$

and let $\xi_c(x, X)$ be the formula

$$\xi(x, X) \vee \mathsf{con}(x).$$

Finally, let θ_c be the following X-positive arithmetic formula:

$$\exists y\, (\xi_c(y, X) \wedge \mathsf{Pr}_{\mathsf{PAT}}(y \underset{\cdot}{\rightarrow} x)).$$

The resulting operator, Θ_c is defined as expected. Let $S \in \mathsf{CONS}$ and set

$$\Theta_c(S) := \{n \in \omega : \mathbb{N} \models \theta_c(x, X)[n, S]\}.$$

Similarly, we can provide a supervaluation-style truth scheme that can be associated with the third prominent supervaluation scheme, which quantifies over maximally consistent sets. For the corresponding supervaluation style truth scheme we need to add a further condition to the definition of $\mathsf{ssk_c}$, namely the completeness condition. This can be done by writing $\mathsf{Com}(\phi)$ if ϕ is of the form $Tt \vee T \underset{\cdot}{\neg}\, s$ where either $t^{\mathbb{N}} = s^{\mathbb{N}}$ and by supplementing the definition of $\mathsf{ssk_c}$ accordingly. We label the resulting scheme $\mathsf{ssk_{mc}}$ and the resulting operator $\mathsf{SSK_{mc}}$, which alike the other two supervaluational strong Kleene jumps is a Kripke jump in our sense. Similarly, Θ_{mc} is obtained from Θ_c by supplementing ξ_c and thus θ_c by an appropriate clause $\mathsf{com}(x)$. However, like its supervaluational counterpart the scheme SSK_{mc} is not classically sound, that is, Lemma 3.7 (below) and its corollaries will fail for this scheme. This will make reasoning about the fixed points of the corresponding operator more difficult and for this reason, as our remarks in Section 3.1 suggest, we will ignore the scheme for the rest of the paper.[24]

As we have mentioned the SSK- and the Θ-operators have the same fixed points and we prove this fact in Theorem 3.5 below. But first we observe that the SSK-operators collects at least as many sentences as the Θ-operators.

Lemma 3.4. *Let $S \in \mathsf{CONS}$. Then $\Theta(S) \subseteq \mathsf{SSK}(S)$ and $\Theta_c(S) \subseteq \mathsf{SSK}_c(S)$.*

Theorem 3.5. *For all $S \in \mathsf{CONS}$*

(*i*) $\mathsf{SSK}(S) = S \Leftrightarrow \Theta(S) = S$

(*ii*) $\mathsf{SSK}_c(S) = S \Leftrightarrow \Theta_c(S) = S.$

[24] It is perhaps worth mentioning that certain observations in this paper also hold for the mc scheme. For example, the results of Section 4 will carry over to the minimal fixed point of this scheme.

The proof is easy but quite painful. We give an outline of the crucial arguments.

Proof sketch. (i) For the left-to-right direction we assume $\mathsf{SSK}(S) = S$ and show for all ϕ,

$$\#\phi \in \Theta(S) \Leftrightarrow \#\phi \in S.$$

If $\#\phi$ is in $\Theta(S)$ then there exists a sentence γ such that

$$\xi(\ulcorner\gamma\urcorner, \overline{S}) \,\&\, \mathsf{PAT} \vdash \gamma \rightarrow \phi.$$

where $\overline{S}$ denotes a predicate constant that has been added to the language of first-order arithmetic that is interpreted by S.[25] By an induction on the positive complexity of γ we can convince ourselves that $\xi(\ulcorner\gamma\urcorner, \overline{S})$ implies $\#\gamma \in \mathsf{SSK}(S)$, which allows us to infer $\#\phi \in \mathsf{SSK}(S)$, that is, by assumption, $\#\phi \in S$. For the converse direction we assume $\#\phi \in S$. This implies $\#\phi \in \mathsf{SSK}(S)$ and there must be a sentence ψ such that $S \models_{\mathsf{sk}} \psi$ and $\mathsf{PAT} \vdash \psi \rightarrow \phi$. Now, we assume $\#\phi \notin \Theta(S)$ and $\neg\xi(\ulcorner\psi\urcorner, S)$. By induction on the positive complexity of ψ using the fact that S is a SSK-fixed point we can show that this assumption leads to a contradiction. We conclude $\#\phi \in \Theta(S)$.

Similarly, for the right-to-left direction we assume $\Theta(S) = S$ and show for all ϕ,

$$\#\phi \in \mathsf{SSK}(S) \Leftrightarrow \#\phi \in S.$$

Now, if $\#\phi \in S$, then by assumption $\#\phi \in \Theta(S)$ and by Lemma 3.4 $\#\phi \in \mathsf{SSK}(S)$. For the converse direction we assume $\#\phi \in \mathsf{SSK}(S)$. Then there must be a sentence γ such that

$$S \models_{\mathsf{sk}} \gamma \,\&\, \mathsf{PAT} \vdash \gamma \rightarrow \phi.$$

By a secondary induction on the positive complexity of γ we can convince ourselves that $S \models_{\mathsf{sk}} \gamma$ implies $\#\gamma \in \Theta(S)$. Thus there is a sentence γ' such that $\xi(\ulcorner\gamma'\urcorner, \overline{S})$ and $\mathsf{PAT} \vdash \gamma' \rightarrow \gamma$. Hence, $\mathsf{PAT} \vdash \gamma' \rightarrow \phi$ and we may infer $\#\phi \in \Theta(S)$, that is, by assumption, $\#\phi \in S$.

The proof of (ii) follows the pattern of (i). For the left-to-right direction we assume $\mathsf{SSK}_c(S) = S$ and show for all ϕ that

$$\#\phi \in \Theta_c(S) \Leftrightarrow \#\phi \in S.$$

The argument is like for (i) with the exception that γ might be of the form $\neg(Ts \wedge T \dot{\neg} t)$ where $t^{\mathbb{N}} = s^{\mathbb{N}}$. In this case the left-to-right direction the claim also follows directly from the definition of Θ_c and the fact that $\mathsf{con}(\ulcorner\gamma\urcorner)$. The converse direction is as for (i) modulo the necessary modifications.

For the right-to-left direction assume $\Theta_c(S) = S$ and show for all ϕ that

$$\#\phi \in \mathsf{SSK_c}(S) \Leftrightarrow \#\phi \in S.$$

Modulo the necessary modifications the reasoning we used for (i) can also be applied to (ii). □

[25] We implicitly assume the following equivalence:

$$\mathbb{N} \models \xi(x, X)[\#\gamma, S] \Leftrightarrow S \models \xi(\ulcorner\gamma\urcorner, \overline{S}).$$

In light of Theorem 3.5 we use, from now on, the jump operation that appeals most for the respective task when proving results about supervaluation-style truth fixed points. However, we repeat our warning that the fact that Θ and SSK have the same fixed points does not imply that the two operations agree on all the stages of the inductive definition. This is generally not the case.

We collect a number of properties relating the different jump operations. Most noteworthy is the fact that supervaluation-style truth jump operations combine aspects of the strong Kleene jump and the supervaluational jump and are situated in between the two jumps:

Lemma 3.6. *Let $S \in \mathsf{CONS}$ and $S \cup S^- \neq \mathsf{Sent}_{\mathcal{L}_T}$. Then*

(i) $\mathsf{SSK}(S) \subsetneq \mathsf{SSK}_c(S)$*;*

(ii) $\mathsf{SK}(S) \subsetneq \mathsf{SSK}(S) \subseteq \mathsf{VB}(S)$*;*

(iii) $\mathsf{SK}(S) \subsetneq \mathsf{SSK}_c(S) \subseteq \mathsf{VC}(S)$*.*

Proof. We only prove item (i) and (ii) for sake of illustration. For (i) the inclusion is a straightforward consequence of the definition of $\mathsf{SSK}(S)$ and $\mathsf{SSK}_c(S)$. To see that the inclusion is strict we notice that by assumption for each S there must be a sentence ϕ such that $\#\phi \notin S$ and $\#\neg\phi \notin S$. But then $\#\neg(T\ulcorner\phi\urcorner \wedge T\ulcorner\neg\phi\urcorner) \notin \mathsf{SSK}(S)$, yet by the definition of SSK_c we have $\#\neg(T\ulcorner\phi\urcorner \wedge T\ulcorner\neg\phi\urcorner) \in \mathsf{SSK}_c(S)$.

(ii) The first inclusion is trivial since $\mathsf{PAT} \vdash \phi \to \phi$ for all $\phi \in \mathcal{L}_T$. For the second inclusion we reason as follows using Lemma 3.3 and the fact that PAT is true in all models of $\mathcal{L}_T$ over the standard model:

$$
\begin{aligned}
\#\phi \in \mathsf{SSK}(S) &\Leftrightarrow \exists\psi(S \models_{\mathsf{sk}} \psi \,\&\, \mathsf{PAT} \vdash \psi \to \phi)\\
&\Rightarrow \exists\psi(S \models_{\mathsf{vb}} \psi \,\&\, \mathsf{PAT} \vdash \psi \to \phi)\\
&\Rightarrow \exists\psi\big(\forall S'(S \subseteq S' \,\&\, S^- \cap S' \Rightarrow S' \models \psi) \,\&\, \mathsf{PAT} \vdash \psi \to \phi\big)\\
&\Rightarrow \exists\psi\forall S'(S \subseteq S' \,\&\, S^- \cap S' = \emptyset \Rightarrow S' \models \psi \wedge \psi \to \phi)\\
&\Rightarrow \forall S'(S \subseteq S' \,\&\, S^- \cap S' = \emptyset \Rightarrow S' \models \phi)\\
&\Rightarrow \#\phi \in \mathsf{VB}(S).
\end{aligned}
$$
□

In Lemma 3.6 the inclusion between the strong Kleene scheme and the supervaluation-style schemes is strict whereas this is not the case for the subset relation between the supervaluation-style schemes and the supervaluation schemes. Moreover, in general, the latter relation cannot be turned into a strict subset relation because, as we shall see in the next section, the SSK (SSK_c) operator and the VB (VC) operator have the same minimal fixed point. Before we prove this claim we continue collecting some facts about the supervaluation-style truth operators.

Lemma 3.7. *Let $S \in \mathsf{CONS}$. Then for all $\phi \in \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}$*

(*i*) $$\#\phi \in \mathsf{SSK}(S) \Rightarrow S \models \phi$$

(*ii*) $$\#\phi \in \mathsf{SSK}_c(S) \Rightarrow S \models \phi.$$

Proof. (i) By definition of $\mathsf{SSK}(S)$ we know that for all $\#\phi \in \mathsf{SSK}(S)$ there exists a ψ such that $S \models_{\mathsf{sk}} \psi \,\&\, \mathsf{PAT} \vdash \psi \to \phi$. But this together with $S \in \mathsf{CONS}$ and the fact that all models based on the natural number structure are PAT-models implies $S \models \psi \,\&\, S \models \psi \rightarrow \phi$, which yields $S \models \phi$. For (ii) we can apply essentially the same reasoning since for $S \in \mathsf{CONS}$ and all $\phi \in \mathcal{L}_{\mathsf{PAT}}$, $\mathsf{Con}(\phi)$ implies $S \models \phi$. □

Corollary 3.8. *Let* $S \in \mathsf{CONS}$*. Then* $\mathsf{SSK}(S) \in \mathsf{CONS}$ *and* $\mathsf{SSK}_c(S) \in \mathsf{CONS}$.

Corollary 3.9 (Classical Soundness)**.** *Let* $S \subseteq \mathsf{SSK}(S)$ *or* $S \subseteq \mathsf{SSK}_c(S)$*. Then for all* $\phi \in \mathcal{L}_{\mathsf{T}}$

(*i*) $\#\phi \in S \Rightarrow S \models \phi$

(*ii*) $S \models T\ulcorner\phi\urcorner \rightarrow \phi.$

Corollary 3.9 implies that supervaluation-style fixed points are models of themselves. A final observation that comes in useful throughout the paper is that the supervaluation style fixed points are closed under modus ponens.

Lemma 3.10. *Let* $\mathsf{SSK}(S) = S$*. If* $\#\phi \in S$ *and* $\#(\phi \rightarrow \psi) \in S$*, then* $\#\psi \in S$.

Proof. Assume $\#\phi \in S$ and $\#(\phi \rightarrow \psi) \in S$. Then by assumption there exist γ, γ' such that $S \models_{\mathsf{sk}} \gamma \,\&\, \mathsf{PAT} \vdash \gamma \rightarrow \phi$ and $S \models_{\mathsf{sk}} \gamma' \,\&\, \mathsf{PAT} \vdash \gamma' \rightarrow (\phi \rightarrow \psi)$. This implies $\mathsf{PAT} \vdash \gamma \wedge \gamma' \rightarrow \psi$ and by definition of the sk-scheme we also obtain $S \models_{\mathsf{sk}} \gamma \wedge \gamma'$. By definition we have $\#\psi \in \mathsf{SSK}(S)$. □

So far we have discussed the different evaluation schemes and their associated jump from a very general perspective. The next section will be devoted to comparing the minimal fixed points of the supervaluation- and the supervaluation-style jumps.

4. Grounded Supervaluation-Style Truth

We have already hinted at two different ways Kripke's theory of truth can be understood. According to the first we take the theory to advocate the minimal fixed point of the relevant jump operation to be the intended interpretation of the truth predicate. According to the second way of understanding the theory all fixed points qualify as suitable interpretations of the truth predicate. In this section we focus on the first understanding of Kripke's theory and compare Kripke's theory in its supervaluational version to the version based on the supervaluation-style truth scheme. It turns out that if we are only interested in the minimal fixed point of the jump operations, supervaluational truth and supervaluation-style truth coincide: the supervaluation jumps have the same minimal fixed-point as their respective supervaluation-style truth jumps.

Before we start establishing these results, let us quickly note that the main reason that the first understanding of Kripke's theory of truth is very interesting from a philosophical point of view is that it gives rise to the notion of grounded truth: the sentences in the minimal fixed point of the various jump operations do not depend on particular assumptions regarding the notion of truth and no sentences are declared true at the outset of the construction process. In other words the sentences in the minimal fixed point only depend on, i.e. are grounded in, non-semantic states of affairs.[26] The fact that the supervaluational jumps and the supervaluation-style truth jumps have the same fixed points therefore means that the

[26] For further discussion of the notion of groundedness we refer the reader to Kripke (1975), Herzberger (1970), Yablo (1982) and Leitgeb (2005).

different schemes give rise to the same notion of grounded truth. This is a nice outcome from the perspective of supervaluation-style truth since the ssk-scheme is clearly simpler and more transparent than the supervaluational scheme. As a consequence the construction process leading to the minimal fixed point will be more transparent, which squares better with the notion of grounded truth since this notion seems to require a traceable path from a grounded sentence to the basic fact grounding it. The notoriously opaque nature of the supervaluation scheme makes it much more difficult to provide such a path. Moreover, the very nature of the supervaluation scheme of looking at all admissible possible interpretations of the truth predicate does not fit well with the idea of groundedness in the first place. From this perspective the supervaluation-style truth jump operations may be seen as providing a more transparent and less complex characterization of grounded (minimal) supervaluational truth.

In alignment with our previous discussion, the ultimate goal of this section will be to establish that the aforementioned jumps have the same minimal fixed point.

Theorem 4.1. *For a given evaluation scheme e let* I_e *denote the minimal fixed point of the jump operation associated with the scheme e. Then*

(*i*) $$\mathsf{I}_{\mathsf{vb}} = \mathsf{I}_{\mathsf{ssk}}$$

(*ii*) $$\mathsf{I}_{\mathsf{vc}} = \mathsf{I}_{\mathsf{ssk}_{\mathsf{c}}}.$$

Using the subset relation the theorem may be split up in two directions. The right-to-left direction, that is, the fact that the minimal supervaluation-style truth fixed points are contained in the minimal supervaluation fixed point is a direct consequence of Lemma 3.6:

Lemma 4.2.

(*i*) $$\mathsf{I}_{\mathsf{ssk}} \subseteq \mathsf{I}_{\mathsf{vb}}$$

(*ii*) $$\mathsf{I}_{\mathsf{ssk}_{\mathsf{c}}} \subseteq \mathsf{I}_{\mathsf{vc}}.$$

The converse direction, however, requires some more work. The result is proven by appeal to Cantini's infinitary Tait-style calculus (Cantini, 1990), which we label SV^c_∞. Cantini shows that

$$\mathsf{SV}^c_\infty \vdash \phi \Leftrightarrow \#\phi \in \mathsf{I}_{\mathsf{vc}}.$$

Moreover, it is clear from his construction that we can do the same for the supervaluation scheme vb.[27] Let us call the calculus that has been modified to this effect SV_∞ (see Definition 4.5 below).

Lemma 4.3 (Cantini)**.** *For all* $\phi \in \mathcal{L}_{\mathsf{T}}$,

(*i*) $$\mathsf{SV}_\infty \vdash \phi \Leftrightarrow \#\phi \in \mathsf{I}_{\mathsf{vb}}$$

(*ii*) $$\mathsf{SV}^c_\infty \vdash \phi \Leftrightarrow \#\phi \in \mathsf{I}_{\mathsf{vc}}.$$

[27] Notice that the only purpose of the initial sequent (AX.C) of SV^c_∞ (cf. Definition 4.5) is to account for the fact that in the scheme vc we only consider consistent precisifications of a given interpretation of the truth predicate. As a consequence all sentences of the form $\neg T\ulcorner\phi\urcorner \vee \neg T\ulcorner\neg\phi\urcorner$ are true in each such admissible precisification and are thus members of every VC-fixed point. (AX.C) guarantees that we can derive all sentences of this form and their consequences. Sentences of this form are, however, not always true in all admissible vb-precisifications and thus not always members of the VB-fixed points. For this reason we need to drop the initial sequent (AX.C) from the infinitary Tait-style calculus in this case. However, once we have dropped (AX.C) we can copy Cantini's reasoning and establish Lemma 4.3.

In virtue of Cantini's theorem it suffices to prove the following lemma for showing that the minimal supervaluation fixed points are contained in the respective minimal supervaluation-style truth fixed points.

Lemma 4.4. *For all $\phi \in \mathcal{L}_{\mathsf{T}}$,*

(*i*) $$\mathsf{SV}_\infty \vdash \phi \Rightarrow \#\phi \in \mathsf{I}_\Theta$$

(*ii*) $$\mathsf{SV}^c_\infty \vdash \phi \Rightarrow \#\phi \in \mathsf{I}_{\Theta_c}.$$

Before we start our proof of Lemma 4.4 we introduce the rules and axioms of the infinitary calculus. As in Section 3.2, the negation symbol is not officially part of our language but defined in the usual way. $\neg T$ should hence be understood as a primitive falsity predicate and the sentences of the language are supposed to be in their negation normal form.

Definition 4.5 (The calculus SV_∞)**.** *We use Γ for finite sets of sentences of the languages and $A, B, A_1, A_2, \ldots$ for the sentences of the language. The calculus has three basic axioms:*

(*AX.1*) $\vdash \Gamma, A$; *if A is a true arithmetic literal;*

(*AX.2*) $\vdash \Gamma, \neg Tt, Ts$; *if $t^{\mathbb{N}} = s^{\mathbb{N}}$;*

(*Sent*) $\vdash \Gamma, \neg Tt$; *if $t^{\mathbb{N}} \notin$* Sent.

The calculus SV^c_∞ has, in addition to three basic axioms, one further axiom that reflects the consistency condition that is required for the supervaluation scheme vc*:*

(*AX.C*) $\vdash \Gamma, \neg Ts, \neg T \dot{\neg} t$; *if $t^{\mathbb{N}} = s^{\mathbb{N}}$;*

In addition to the axioms SV_∞ and SV^c_∞have the following rules:

$$(\wedge)\,\frac{\vdash \Gamma, A \wedge B, A \qquad \vdash \Gamma, A \wedge B, B}{\vdash \Gamma, A \wedge B} \qquad (\vee)\,\frac{\vdash \Gamma, A_1 \vee A_2, A_i}{\vdash \Gamma, A_1 \vee A_2}, i \in \{1,2\}$$

$$(\omega)\,\frac{\textit{for all } n \in \omega \ \vdash \Gamma, \forall x A, A(\overline{n})}{\vdash \Gamma, \forall x A} \qquad (\exists)\,\frac{\textit{for some } n \in \omega \ \vdash \Gamma, \exists A, A(\overline{n})}{\vdash \Gamma, \exists x A}$$

$$(T)\,\frac{\vdash A}{\vdash \Gamma, Tt}, t^{\mathbb{N}} = \#A \qquad (\neg T)\,\frac{\vdash \neg A}{\vdash \Gamma, \neg Tt}, t^{\mathbb{N}} = \#A$$

We need one more auxiliary lemma for proving Lemma 4.4.

Lemma 4.6. *Let Γ be a set of formulas. Then*

(*i*) $$\mathsf{SV}_\infty \vdash \Gamma \Rightarrow \#\left(\bigvee \Gamma\right) \in \mathsf{I}_\Theta$$

(*ii*) $$\mathsf{SV}^c_\infty \vdash \Gamma \Rightarrow \#\left(\bigvee \Gamma\right) \in \mathsf{I}_{\Theta_c}.$$[28]

Proof. In the proof we shall focus on item (i). The proof is by an induction on the height of the derivation in SV_∞. As induction hypothesis we may assume that for all $\alpha < \beta$

$$\mathsf{SV}_\infty \vdash^\alpha \Gamma \Rightarrow \#\left(\bigvee \Gamma\right) \in \mathsf{I}_\Theta$$

[28] By $\mathsf{SV}_\infty \vdash \Gamma$ ($\mathsf{SV}^c_\infty \vdash \Gamma$) we denote that $\vdash \Gamma$ is a derivable sequent in SV_∞ (SV^c_∞).

where $\mathsf{SV}_\infty \vdash^\alpha \Gamma$ denotes that Γ has been derived in SV_∞ by a proof of height α. AX.1 follows from (1) of the definition of ξ , reflexivity of PAT-implication and classical logic, i.e., disjunction introduction. AX.2 may be obtained from (1) of the definition of ξ and, again, classical logic.

For the induction step we need to show that membership in I_Θ is closed under the rules of SV_∞. We deal with ($\vee$) and (ω) for sake of illustration. The remaining cases work in a similar way. For ($\vee$) we may assume $\mathsf{SV}_\infty \vdash^\beta \Gamma, \phi \vee \psi$ and $\mathsf{SV}_\infty \vdash^\alpha \Gamma, \phi \vee \psi, \chi$ for $\chi \in \{\phi, \psi\}$. By IH we obtain $\# \bigvee(\Gamma, \phi \vee \psi, \chi) \in \mathsf{I}_\Theta$. But $\mathsf{PAT} \vdash \bigvee(\Gamma, \phi \vee \psi, \chi) \rightarrow \bigvee(\Gamma, \phi \vee \psi)$ and thus $\# \bigvee(\Gamma, \phi \vee \psi, \chi) \rightarrow \bigvee(\Gamma, \phi \vee \psi) \in \mathsf{I}_\Theta$. By Lemma 3.10 we obtain $\# \bigvee(\Gamma, \phi \vee \psi) \in \mathsf{I}_\Theta$.

Let us now turn to (ω). We assume $\mathsf{SV}_\infty \vdash^\beta \Gamma, \forall x \phi$ and $\mathsf{SV}_\infty \vdash^\alpha \Gamma, \forall x \phi, \phi(\overline{n})$ for all $n \in \omega$. By IH we infer $\# \bigvee(\Gamma, \forall x \phi, \phi(\overline{n})) \in \mathsf{I}_\Theta$ for all $n \in \omega$. By (4) of the definition of ξ we obtain $\# \forall y \bigvee(\Gamma, \forall x \phi, \phi(y)) \in \mathsf{I}_\Theta$ where we choose y to be new to Γ and $\forall x \phi$. Then $\mathsf{PAT} \vdash \forall y \bigvee(\Gamma, \forall x \phi, \phi(y)) \rightarrow \bigvee(\Gamma, \forall x \phi)$ and, again, by Lemma 3.10 we conclude $\bigvee(\Gamma, \forall x \phi) \in \mathsf{I}_\Theta$.

Now, for item (ii) we have to deal, in addition to the basic axioms we discussed for (i), with Ax.C. This, however, can be done rather easily by appeal to the consistency condition in θ_c: by assumption $s = t$ and $\neg(Ts \wedge T \underset{\cdot}{\neg} t)$ is, by definition, $\neg Ts \vee \neg T \underset{\cdot}{\neg} t$. □

Since Lemma 4.6 has Lemma 4.4 as an immediate consequence, we can establish our main Theorem.

Proof of Theorem 4.1. Immediate consequence of Lemma 4.2 and Lemma 4.4. □

In virtue of Theorem 4.1 we know that supervaluation-style truth and the supervaluation schemes have the same minimal fixed points.[29] As a consequence, the different schemes give rise to the same notion of grounded truth.

Furthermore, as an immediate corollary of this fact we know the complexity of the minimal fixed point of the supervaluation-style truth jump operations.

Corollary 4.7. $\mathsf{I}_{\mathsf{ssk}}$ *and* $\mathsf{I}_{\mathsf{ssk}_\mathsf{c}}$ *are* Π^1_1*-complete sets of integers.*

Proof. By results of Kripke (1975) and Burgess (1986) we know that the minimal fixed point of VB and VC is Π^1_1-complete. The claim is thus immediate by Theorem 4.1. □

Now that we have seen that the supervaluation-style truth and the supervaluation scheme agree on the minimal fixed point, the immediate question arises whether the jump operations of the two types of schemes agree on all fixed points: does Kripke's supervaluation theory of truth also coincide with Kripke's theory based on the supervaluation-style truth scheme on the second understanding of his theory where we are interested in arbitrary fixed points? If the answer was yes, then this would show, as far as the notion of truth is concerned, that there is no difference between the supervaluation scheme and the supervaluation-style truth scheme. Moreover, the move from second-order to first-order consequence would not have any consequences when constructing suitable interpretations of the truth predicate.

This would be rather surprising, and indeed, the two schemes cannot agree on all fixed points. More specifically, there must be at least one supervaluation-style truth fixed point

[29] One might wonder how the minimal strong Kleene fixed point relates to the minimal supervaluation-style truth fixed point. The former is, obviously, a subset of the latter. Moreover, the minimal supervaluation-style fixed point is *not* just the minimal strong Kleene fixed-point closed under PAT. The sentence $T\ulcorner\lambda \vee \neg\lambda\urcorner$, where λ is a usual Liar sentence, is a member of the former but it is *not* a member of the minimal strong Kleene fixed point closed under PAT.

that is not a supervaluation fixed point: by results of Welch (2015) (see also Fischer et al. (2015)) we know that every supervaluation fixed point is Π^1_1-hard. But as a consequence of Theorem 5.1 below we know that there must be a supervaluation-style truth fixed point that is not Π^1_1-hard. Thanks to Philip Welch (personal communication) we can say a little bit more about the relation between supervaluation-style truth fixed points and supervaluation fixed points: there must be a maximal VB-fixed point that is not a SSK-fixed point. This follows from two observations: 1. Due to complexity considerations VB and SSK cannot have the same maximal intrinsic fixed point. The maximal intrinsic VB-fixed point is Δ^1_2-in-a-Π^1_2 parameter while the maximal intrinsic SKK-fixed point is at most Σ^1_1-in-a-Π^1_1 parameter.[30] 2. The maximal intrinsic fixed point is the intersection of all maximal fixed points of the respective scheme.[31] Now, by Lemma 3.6, if a maximal VB-fixed point is a SSK-fixed point, then it is a maximal SSK-fixed point. Hence, there must be a maximal VB-fixed point, which is not a SSK-fixed point for otherwise, by 2, the two schemes would have the same maximal intrinsic fixed point, which contradicts 1.[32] Moreover, all these remarks are independent of the particular choice of the supervaluation and the supervaluation-style truth scheme. In particular, everything would go through for the schemes vc and $\mathsf{ssk_c}$. Welch's observation has two immediate consequences. On the one hand, it shows that the set of supervaluational fixed points is not a subset of the set of supervaluation-style truth fixed points. Indeed putting this together with our previous observation the set of supervaluation fixed points and the set of suervaluation-style fixed points are incomparable with respect to the subset relation. On the other hand, Welch's observation shows that the evaluation schemes also diverge with respect to the the maximal intrinsic fixed point. The latter observation is interesting since the maximal intrinsic fixed point is the largest fixed point that can be obtained without making arbitrary decisions with respect to the truth value of certain sentences. As a possible interpretation of the truth predicate the maximal intrinsic fixed point is thus of particular interest and the fact that the schemes differ on the maximal intrinsic fixed point shows that despite similarities, the two schemes give rise to different versions of Kripke's theories of truth, at least, if we are not interested in grounded truth. In the next section, further to our last remark, we observe that no jump operation, which is definable by a first-order arithmetical formula, can have exactly the same fixed points as any of the supervaluational jumps we have discussed. This observation will kick off the discussion of axiomatic theories of supervaluation-style truth.

5. Axiomatic Theories of Supervaluation-Style Truth

The inductive definition of I_Θ can be used to extract an axiomatic theory of truth—as in the case of strong Kleene scheme and the axiomatic theory KF ("Kripke-Feferman").[33] The resulting axiomatic theory of truth will not uniquely determine I_Θ as the suitable interpretation of the truth predicate even if the interpretation of the arithmetical vocabulary is fixed at the outset. So, the theory will not be a theory of grounded supervaluation-style truth. In

[30] See Burgess (1988) for a proof of the former. The latter observation follows from the definition of the maximal intrinsic fixed point and the fact that the complexity of set of sentences true in some SSK fixed point is at most Σ^1_1, which follows immediately from the definition of the SSK scheme.

[31] See, for instance, (Visser, 1984, Theorem 2.18.2).

[32] Welch's observation is actually slightly stronger. It shows that there is a maximal VB-fixed point S such that $S \not\subseteq \mathsf{SSK}(S)$.

[33] See Halbach (2011) for a presentation and discussion of KF.

fact, Fischer et al. (2015) observe that, in general, there cannot be a recursively enumerable theory that uniquely characterizes a non-reducible Π^1_1-set of natural numbers relative to the standard model so not such axiomatic theory could be given. However, it is possible that the axiomatic theory, which is extracted from the clauses of the inductive definition, characterizes the lattice of fixed points generated by the SSK-jump relative to the standard model, that is, the interpretations of the truth predicate in the standard model must be fixed points of the SSK-jump. Such a theory is, in the terminology of Fischer et al. (2015), $\mathbb{N}$-categorical. This suggests that axiomatic theories of truth are an interesting tool for investigating Kripke's theory of truth on its arbitrary fixed-point reading.

The straightforward way to extract an $\mathbb{N}$-categorical theory from the operator Θ is to add to the axioms of PA formulated in the language $\mathcal{L}_\mathsf{T}$, that is PAT, the axiom

$$(\theta Ax) \qquad\qquad \forall x\,(Tx \leftrightarrow \theta(x, T))\,.$$[34]

We shall call this theory BIT.

Theorem 5.1. *Let* $\mathsf{S} \subseteq \omega$*. Then*

$$S \models \mathsf{BIT} \Leftrightarrow \Theta(S) = S.$$

Proof. $\Rightarrow$: Assume $S \models \mathsf{BIT}$ and assume for an arbitrary sentence $\phi \in \mathcal{L}_\mathsf{T}$ that $\#\phi \in S$:

$$\begin{aligned}
\#\phi \in S \Leftrightarrow\ & S \models T\ulcorner\phi\urcorner & \\
\Leftrightarrow\ & S \models \theta(\ulcorner\phi\urcorner, T) & (\theta Ax)\\
\Leftrightarrow\ & \mathbb{N} \models \theta(x, X)[\#\phi, S] & \text{Def.}\\
\Leftrightarrow\ & \#\phi \in \Theta(S) & \text{Def. } \Theta
\end{aligned}$$

This establishes the left-to-right direction.

$\Leftarrow$: Assume $S = \Theta(S)$ and suppose for arbitrary $n \in \omega$ that $S \models T\overline{n}$. Then $n \in S$ and thus by assumption $n \in \Theta(S)$. But the latter is equivalent to $\mathbb{N} \models \theta(x, X)[n, S]$ which by definition is equivalent to $S \models \theta(\overline{n}, T)$. This yields the desired result. □

Theorem 5.1 has the immediate consequence that supervaluation-style truth and supervaluational truth cannot agree on all fixed points. Indeed, as we have pointed out at the end of Section 4, Theorem 5.1 shows that there must be a supervaluation-style fixed point that is not a supervaluation fixed point since it implies that not all supervaluation-style fixed points can be Π^1_1-hard. The theorem establishes that the second-order property of being a supervaluation-style truth fixed point defines a Δ^1_1-set of sets of natural numbers. But no Δ^1_1-set of sets of natural numbers can have only Π^1_1-hard sets of natural numbers as its members. Rather such a second-order property would be a non-reducible Π^1_1-set of sets of natural numbers, which, unsurprisingly, is exactly the complexity of the supervaluation fixed-point property.[35] It is worth noting that in the proof of Theorem 5.1 we did not assume any particular property of θ and as a consequence we may infer that no jump operation that

[34] Here, and in the remainder of the paper $\theta(x, T)$ ($\theta_c(x, T)$) and $\xi(x, T)$ ($\xi_c(x, T)$) denote the formulas where all occurrences of the free second-order variable X have been replaced by the truth predicate.

[35] See Fischer et al. (2015) for further discussion and explanation.

is definable by a first-order arithmetical formula can have exactly the same fixed points as the supervaluational jump.

In the case of purely compositional jumps such as the jump Ξ in the strong Kleene case, using the strategy that led to the theory BIT not only produces $\mathbb{N}$-categorical but also attractive theories. In this case the theory obtained by adding

$$(\xi Ax) \qquad \forall x\,(Tx \leftrightarrow \xi(x,T))$$

to PAT has a neat alternative axiomatization. We can divide (ξAx) into a list of biconditionals that provide the truth conditions of a sentence of $\mathcal{L}_{\mathsf{T}}$ depending on the build-up of the sentence. For example, for disjunctions we can provide the following axiom:

$$\forall x,y\,(\mathsf{Sent}(x \mathbin{\underset{\cdot}{\vee}} y) \rightarrow (Tx \mathbin{\underset{\cdot}{\vee}} y \leftrightarrow Tx \vee Ty)).$$

The resulting theory is just the well-known theory KF.

5.1 The Theories of Inductive Truth

However, in the case of the supervaluation-style jumps we do not necessarily obtain nice theories in this way since we cannot break down (θAx) into a number of "interesting" biconditionals depending on the built-up or the compositional structure of a sentence. The reason for this limitation is, of course, that the supervaluation-style jump operation is closed under PAT-consequence and thus the jump is not compositional. Fortunately, the failure of compositionality is not as drastic as in the case of the supervaluational jumps. Rather, in the case of the supervaluation-style jumps by the means of a first-order formula we can point precisely where, and to which, extent compositionality—and thus compositional reasoning—fails. As a consequence, it turns out that, after all, we can provide a list of axioms that depending on the built up provides us with conditions for the truth of a sentence. In other words, we can provide an attractive theory of supervaluation-style truth, which is an $\mathbb{N}$-categorical axiomatization of the fixed points of the supervaluation-style jump operation. But we shall not be using the formula θ directly in constructing our theory and as a consequence establishing the $\mathbb{N}$-categoricity result will require some metatheoretic reasoning which goes beyond simply reading off the clauses of the inductive definition. A further point worth noting is that in this section, in contrast to the previous sections, we take $\neg$ to be a primitive logical constant of the language.

Definition 5.2 (Inductive Truth)**.** *Inductive Truth (*IT*) consists of all axioms of PA in the language $\mathcal{L}_T$ and the following axioms:*

(Ax1) $\forall s,t(Ts \mathbin{\underset{\cdot}{=}} t \leftrightarrow s^\circ = t^\circ)$

(Ax2) $\forall s,t(Ts \mathbin{\underset{\cdot}{\neq}} t \leftrightarrow s^\circ \neq t^\circ)$

(Ax3) $\forall x,y\big(\mathsf{Sent}(x \mathbin{\underset{\cdot}{\wedge}} y) \rightarrow (Tx \wedge Ty \rightarrow T(x \mathbin{\underset{\cdot}{\wedge}} y))\big)$

(Ax4) $\forall x,y\big(\mathsf{Sent}(x \mathbin{\underset{\cdot}{\vee}} y) \rightarrow$

$(Tx \vee Ty \vee \exists z\,(\xi(z,T) \wedge \mathsf{Pr}_{\mathsf{PAT}}(z \mathbin{\underset{\cdot}{\rightarrow}} x \mathbin{\underset{\cdot}{\vee}} y)) \leftrightarrow T(x \mathbin{\underset{\cdot}{\vee}} y))\big)$

(Ax5) $\forall v \forall x\big(\mathsf{Sent}(\underset{\cdot}{\forall}\, vx) \rightarrow (\forall t Tx(t/v) \rightarrow T \underset{\cdot}{\forall}\, vx)\big)$

(Ax6) $\forall v \forall x\big(\mathsf{Sent}(\underset{\cdot}{\exists}\, vx) \rightarrow (\exists t Tx(t/v) \vee \exists z\,(\xi(z,T) \wedge \mathsf{Pr}_{\mathsf{PAT}}(z \mathbin{\underset{\cdot}{\rightarrow}} \underset{\cdot}{\exists}\, vx)) \leftrightarrow T \underset{\cdot}{\exists}\, vx)\big)$

(Ax7) $\forall t(Tt^\circ \to T\,\underset{\cdot}{T}\,t)$

(Ax8) $\forall t(T \underset{\cdot}{\neg}\, t^\circ \vee \neg\mathsf{Sent}(t^\circ) \leftrightarrow T \underset{\cdot}{\neg}\, \underset{\cdot}{T}\, t)$

(Ax9) $\forall x, y(Tx \wedge \mathsf{Pr}_{\mathsf{PAT}}(x \underset{\cdot}{\to} y)) \to Ty)$

(Ax10) $\forall x(T \underset{\cdot}{\neg}\, x \to \neg Tx)$

(Ax11) $T\ulcorner\forall x(Tx \to \mathsf{Sent}(x))\urcorner$

(Ax12) $\forall t_1, \dots, t_n \big(T\ulcorner\phi(\underset{\cdot}{t_1}, \dots, \underset{\cdot}{t_n})\urcorner \to \phi(t_1^\circ, \dots, t_n^\circ)\big)$.

The reader acquainted with Cantini's (Cantini, 1990) theory VF will remark some striking similarities between VF and IT. We compare the two theories later in this paper, once we have investigated IT in some detail.

Lemma 5.3. *Let* Pos *denote the set of formulas in which the truth predicate occurs only in the scope of an even number of negation symbols. Then* IT *proves*

(i) $\forall x, y\big(\mathsf{Sent}(x \underset{\cdot}{\to} y) \to (Tx \wedge Tx \underset{\cdot}{\to} y \to Ty)\big)$

(ii) $\forall x, s, t\big(\mathsf{Sent}(\underset{\cdot}{\forall}\, vx) \to (s^\circ = t^\circ \to (Tx(s/v) \leftrightarrow Tx(t/v)))\big)$

(iii) $\forall x, y\big(\mathsf{Sent}(x \underset{\cdot}{\wedge} y) \to (Tx \wedge Ty \leftrightarrow T(x \underset{\cdot}{\wedge} y))\big)$

(iv) $\forall v \forall x\big(\mathsf{Sent}(\underset{\cdot}{\forall}\, vx) \to (\forall t Tx(t/v) \leftrightarrow T\,\underset{\cdot}{\forall}\, vx)\big)$

(v) $\forall t\,(T\,\underset{\cdot}{T}\,t \leftrightarrow Tt^\circ)$

(vi) $\forall x\,(Tx \to \mathsf{Sent}(x))$

(vii) $\forall t_1, \dots, t_n\big(T\ulcorner\phi(\underset{\cdot}{t_1}, \dots, \underset{\cdot}{t_n})\urcorner \leftrightarrow \phi(t_1^\circ, \dots, t_n^\circ)\big)$, *for* $\phi \in$ Pos

(viii) $\forall x\,(\xi(x, T) \to Tx)$

(ix) $\forall x, y\,(\xi(x, T) \wedge \mathsf{Pr}_{\mathsf{PAT}}(x \underset{\cdot}{\to} y) \to Ty)$.

Proof. (i) is proved by (Ax3) and (Ax9). For (ii) observe that by (Ax1) and (Ax9) we obtain $T(s \underset{\cdot}{=} t \underset{\cdot}{\to} (x(s/v) \leftrightarrow x(t/v)))$. The claim follows by (i) and (Ax1). (iii) by (Ax9) and (Ax3); (iv) by (Ax5), (Ax9) and (ii); (v) by (Ax7) and (Ax12); (vi) is proved by (Ax11) and (Ax12). The left-to-right direction of (vii) is by (Ax12), the converse direction is by an induction on the build-up of the sentences in Pos. (viii) is also proved by an induction on the positive complexity of ϕ; (ix) is immediate by (viii) and (Ax9). □

Lemma 5.3(vii) has the interesting consequence that we can interpret KF in IT.

Corollary 5.4 (Cantini/Halbach)**.** KF *is truth definable in* IT.[36]

Proof. See Halbach (2009, pp. 792/793). □

[36] The notion of truth definability was introduced by Fujimoto (2010). Roughly, it means that there is an unrelativized interpretation of KF in IT which keeps the arithmetical vocabulary fixed.

As in Cantini's theory VF (Cantini, 1990) we cannot turn IT into a compositional theory, that is, we cannot replace (Ax4) and (Ax6) by the following equivalences:

(Ax4+) $$\forall x, y\Big(\mathsf{Sent}(x \underset{\cdot}{\vee} y) \rightarrow (Tx \underset{\cdot}{\vee} y \leftrightarrow Tx \vee Ty)\Big)$$

(Ax6+) $$\forall v \forall x\Big(\mathsf{Sent}(\underset{\cdot}{\exists} vx) \rightarrow (\exists t Tx(t/v) \leftrightarrow T \underset{\cdot}{\exists} vx)\Big)$$

Each of the axioms so strengthened will cause the theory to become inconsistent. Actually, we can show that they cause inconsistency even if we dispense of (Ax12). To see this notice that each of the strengthened axioms implies $T\ulcorner\lambda\urcorner \vee T\ulcorner\neg\lambda\urcorner$ in IT (without (Ax12)). For (Ax4+) consider the instance $T\ulcorner\lambda \vee \neg\lambda\urcorner \leftrightarrow T\ulcorner\lambda\urcorner \vee T\ulcorner\neg\lambda\urcorner$. For (Ax6+) let $\psi(x)$ be the formula $(x = 0 \wedge \lambda) \vee (x = 1 \wedge \neg\lambda)$ and consider the corresponding instance of the converse direction of (Ax6): $T\ulcorner\exists x\psi(x)\urcorner \rightarrow \exists t T\ulcorner\psi(\underset{\cdot}{t})\urcorner$ where $\mathsf{PAT} \vdash \exists x\psi(x)$. But also by (Ax9) $\exists t T\ulcorner\psi(\underset{\cdot}{t})\urcorner$ implies $T\ulcorner\lambda\urcorner \vee T\ulcorner\neg\lambda\urcorner$.[37] Here λ is a sentence such that

(L) $$\mathsf{PAT} \vdash \lambda \leftrightarrow \neg T\ulcorner\lambda\urcorner.$$

We have

1. $T\ulcorner\lambda\urcorner \rightarrow T\ulcorner\neg T\ulcorner\lambda\urcorner\urcorner$ (L), (Ax9),Lemma 5.3 (i);
2. $T\ulcorner\lambda\urcorner \rightarrow \neg T\ulcorner T\ulcorner\lambda\urcorner\urcorner$ 1,(Ax10);
3. $T\ulcorner\lambda\urcorner \rightarrow T\ulcorner T\ulcorner\lambda\urcorner\urcorner$ (Ax7);
4. $T\ulcorner\lambda\urcorner \rightarrow \bot$ 2,3;
5. $T\ulcorner\neg\lambda\urcorner \rightarrow T\ulcorner\neg T\ulcorner\lambda\urcorner\urcorner$ (Ax8)
6. $T\ulcorner\neg\lambda\urcorner \rightarrow \neg T\ulcorner T\ulcorner\lambda\urcorner\urcorner$ 5,(Ax10);
7. $T\ulcorner\neg\lambda\urcorner \rightarrow T\ulcorner T\ulcorner\lambda\urcorner\urcorner$ (L), Lemma 5.3 (i);
8. $T\ulcorner\neg\lambda\urcorner \rightarrow \bot$ 6,7.
9. $T\ulcorner\lambda\urcorner \vee T\ulcorner\neg\lambda\urcorner \rightarrow \bot$ 4,8

Before we discuss the models of IT, we introduce the theory of Consistent Inductive Truth, which is intended to match the supervaluation-style jump operations SSK_c and Θ_c.

Definition 5.5 (Consistent Inductive Truth). *Consistent Inductive Truth* (IT_c) *consists of axioms (Ax1)-(Ax3), (Ax5) and (Ax7)-(Ax12) of* IT *together with*

(AxC) $$\forall t\, (T\ulcorner T \underset{\cdot}{\neg}\, \underset{\cdot}{t}\ \rightarrow \neg T\, \underset{\cdot}{t}\,\urcorner)\,;$$

(Ax4$_c$) $$\forall x, y\Big(\mathsf{Sent}(x \underset{\cdot}{\vee} y) \rightarrow$$
$$(Tx \vee Ty \vee \exists z\, (\xi_c(z, T) \wedge \mathsf{Pr}_{\mathsf{PAT}}(z \underset{\cdot}{\rightarrow} x \underset{\cdot}{\vee} y)) \leftrightarrow T(x \underset{\cdot}{\vee} y))\Big);$$

(Ax6$_c$) $$\forall v \forall x\Big(\mathsf{Sent}(\underset{\cdot}{\exists} vx) \rightarrow (\exists t Tx(t/v) \vee \exists z\, (\xi_c(z, T) \wedge \mathsf{Pr}_{\mathsf{PAT}}(z \underset{\cdot}{\rightarrow} \underset{\cdot}{\exists} vx)) \leftrightarrow T \underset{\cdot}{\exists} vx)\Big).$$

In Definition 5.5, the axiom (Ax10) and the left-to-right direction of (Ax8) are redundant, that is, they could be proved on the basis of the remaining axioms of IT_c.

[37] See Friedman and Sheard (1987, p. 15) for the argument leading to the inconsistency of (Ax6).

5.2 Inductive Truth and Supervaluation-style Fixed Points

As promised, we now turn to models of IT (IT_c) and show that IT (IT_c) characterizes the supervaluation-style truth fixed points relative to the standard model, that is, we show that IT (IT_c) is an $\mathbb{N}$-categorical axiomatization of the supervaluation-style fixed points. As a by-product we establish the consistency of IT (IT_c). Indeed, this follows directly from the next lemma which establishes that the axioms of IT (IT_c) are true in the classical fixed-point models of SSK (SSK_c).

Lemma 5.6. *Let $S \subseteq \omega$.*

(*i*) $$\mathsf{SSK}(S) = S \Rightarrow S \models \mathsf{IT}$$

(*ii*) $$\mathsf{SSK_c}(S) = S \Rightarrow S \models \mathsf{IT_c}.$$

Proof. (i) Ax1-7, Ax11 follow immediately from the definition of the jump Θ and the reflexivity of PAT-implication. Ax10 follows from the definition of the jump and Corollary 3.8. Ax12 follows directly from Corollary 3.9. The left-to-right direction of Ax8 follows again from the definition of the jump. For the converse direction we may assume that $\#\neg Tt^\circ \in \mathsf{SSK}(S)$. Then there must be a ψ such that $S \models_{\mathsf{sk}} \psi$ and $\mathsf{PAT} \vdash \psi \rightarrow \neg Tt^\circ$. Now, $\mathsf{PAT} \vdash \psi \rightarrow \neg Tt^\circ$ implies $\forall X \subseteq \omega(X \models \psi \Rightarrow X \not\models Tt^\circ)$. But this is guaranteed only, if t° is not the name of a sentence or if there exists a ϕ such that $t^\circ = \ulcorner\phi\urcorner$ and $X \models \psi$ implies $\#\neg\phi \in X$. Since for $S \in \mathsf{Cons}$, $S \models_{\mathsf{sk}} \psi$ implies $S \models \psi$, we may conclude to the desired.

(ii) We only discuss (AxC), as the remaining axioms are, modulo the necessary modifications, as in the proof for (i). We assume $\mathsf{SSK}_c(S) = S$ and need to show that for an arbitrary closed term t, $\#T \underset{\cdot}{\neg}\, t \rightarrow \neg Tt \in S$. By assumption S is closed under PAT and it suffices to show $\#\neg(T \underset{\cdot}{\neg}\, t \wedge Tt) \in S$. But since $\mathsf{Con}(\neg(T \underset{\cdot}{\neg}\, t \wedge Tt))$ we know that $\#\neg(T \underset{\cdot}{\neg}\, t \wedge Tt) \in \mathsf{SSK_c}(S)$ and thus $\#\neg(T \underset{\cdot}{\neg}\, t \wedge Tt) \in S$ by assumption. □

Corollary 5.7. IT *and* $\mathsf{IT_c}$ *are consistent.*

Next we show that the converse direction of Lemma 5.6 holds as well. This establishes that IT (IT_c) is an $\mathbb{N}$-categorical axiomatization of the SSK (SSK_c) fixed points.

Theorem 5.8 ($\mathbb{N}$-categoricity)**.** *Let $S \subseteq \omega$. Then*

(*i*) $$S \models \mathsf{IT} \Leftrightarrow \mathsf{SSK}(S) = S$$

(*ii*) $$S \models \mathsf{IT}_c \Leftrightarrow \mathsf{SSK}_c(S) = S.$$

Proof. The right-to-left direction is by Lemma 5.6. For the converse direction one shows by a routine induction on the positive complexity of ϕ that

(*i*) $$\#\phi \in S \Leftrightarrow \#\phi \in \Theta(S)$$

(*ii*) $$\#\phi \in S \Leftrightarrow \#\phi \in \Theta_c(S).$$

For both items we discuss the case $\phi \doteq \psi \vee \chi$ and leave the remaining cases to the reader:

(i) Before we start note that $S \models T\ulcorner\psi\urcorner \vee T\ulcorner\chi\urcorner$ implies $S \models \exists z(\xi(z,T) \wedge \mathsf{Pr}_{\mathsf{PAT}}(z \underset{\cdot}{\rightarrow} \ulcorner\psi \vee \chi\urcorner))$:

$$\begin{aligned} \#(\psi \vee \chi) \in S &\Leftrightarrow S \models T\ulcorner\psi \vee \chi\urcorner \\ &\Leftrightarrow S \models T\ulcorner\psi\urcorner \vee T\ulcorner\chi\urcorner \vee \exists z(\xi(z,T) \wedge \mathsf{Pr}_{\mathsf{PAT}}(z \underset{\cdot}{\rightarrow} \ulcorner\psi \vee \chi\urcorner)) \qquad (Ax4) \end{aligned}$$

$$\begin{aligned} &\Leftrightarrow S \models \theta(\ulcorner \psi \vee \chi \urcorner, T) && \text{Def.} \\ &\Leftrightarrow \mathbb{N} \models \theta(x, X)[\#(\psi \vee \chi), S]) && \\ &\Leftrightarrow \#(\psi \vee \chi) \in \Theta(S). && \text{Def.} \end{aligned}$$

(ii) The proof of (ii) is exactly parallel to the one of (i). It suffices to replace ξ, θ and Θ by, respectively, ξ_c, θ_c and Θ_c. □

Theorem 5.8 establishes that IT and $\mathsf{IT_c}$ are $\mathbb{N}$-categorical axiomatizations of Kripke's theory of truth based on the schemes SSK and $\mathsf{SSK_c}$ respectively, which substantiates our claim that supervaluation-style truth admits a neat proof-theoretic characterization.

5.3 Inductive Truth and Cantini's theory VF

For the proof of Theorem 5.8 to work it was crucial that we could provide a precise condition of when compositionality was allowed to fail. Moreover, it is important that this condition can be expressed by a first-order arithmetical formula (with one free second-order variable) for otherwise we could not express this condition appropriately within the truth theory. This is exactly the reason why we cannot provide an $\mathbb{N}$-categorical axiomatization of the Kripke's theory of truth based on a supervaluational scheme: for supervaluation schemes there will not be a first-order arithmetical condition that tells us when compositionality fails. As a consequence, the tie between the Cantini's theory VF and the scheme VC, which VF is thought to capture, cannot be as close as the relation between $\mathsf{IT_c}$ and $\mathsf{SSK_c}$.

Despite these differences the theories of inductive truth and Cantini's theory VF are very similar theories. In fact, VF is a proper subtheory of IT_c and differs from this theory only in the axioms $(Ax4_c)$ and $(Ax6_c)$, which are replaced by the following principles:

$$(\vee Ax) \qquad \forall x, y\Big(\mathsf{Sent}(x \underset{\cdot}{\vee} y) \rightarrow (Tx \vee Ty \rightarrow T(x \underset{\cdot}{\vee} y))\Big);$$

$$(\exists Ax) \qquad \forall v \forall x\Big(\mathsf{Sent}(\underset{\cdot}{\exists} vx) \rightarrow (\exists t Tx(t/v) \rightarrow T \underset{\cdot}{\exists} vx)\Big).$$

However, like IT_c the theory VF includes the axiom (AxC), which makes it incomparable to IT via the subtheory relation.

The crucial difference between the theories of Inductive Truth and VF consists thus (unsurprisingly) in the compositional axioms for $\vee$ and $\exists$. For all three theories full compositionality fails, but while VF remains silent with respect to the reason for this failure, the theories of inductive truth provide, as we have repeatedly urged, exact conditions under which compositional reasoning fails.

To see that VF is a proper subtheory of $\mathsf{IT_c}$ the $\mathbb{N}$-categoricity of the latter theory proves useful: it provides us with an argument that IT_c is not $\mathcal{L}_\mathsf{T}$-conservative over VF. In other words, IT_c is a proper extension of VF. By a result of Cantini (1996) we know that a set of stable sentences over the Herzberger sequence is a VF-model. But it cannot be an IT_c-model, as this would contradict Theorem 5.8, that is, the $\mathbb{N}$-categoricity of IT_c with respect to SSK_c fixed points.[38]

As we shall see in the next section the situation changes when we ask whether $\mathsf{IT_c}$ is $\mathcal{L}$-conservative, i.e. conservative in the arithmetical language, over VF. This leads to the

[38] No set S of stable sentences can be a Θ-fixed point since for each such S there will be Liar sentences λ_1 and λ_2 such that $\#(\lambda_1 \leftrightarrow \lambda_2) \in S$ but $\#(\lambda_1 \leftrightarrow \lambda_2)$ cannot be a member of any Θ-fixed point. See Burgess (1986) for further details.

question of the proof-theoretic strength of the theories of inductive truth. Due to the fact that IT$_c$ is a proper supertheory of VF we cannot rely on Cantini's (Cantini, 1990) results when determining the proof theoretic strength of the theories of inductive truth but have to check to what extent these results carry over.

6. Proof Theory

The theories IT and IT$_c$ are to date the strongest $\mathbb{N}$-categorical theories on the market. This follows from a result by Friedman and Sheard (1987) who show that a subtheory of IT is proof-theoretically equivalent to ID$_1$.

Theorem 6.1 (Friedman & Sheard)**.** *Let Σ be a theory extending* PAT *which proves (T-Out), (Ax1), (Ax2), (Ax9) and*

$$\forall x, v\Big(\mathsf{Sent}(\underset{\cdot}{\forall}\, vx) \rightarrow (\forall y Tx(y/v) \rightarrow T \underset{\cdot}{\forall}\, vx)\Big).$$

Then Σ proves all arithmetical theorems of ID$_1$.

This establishes the lower bound of the proof-theoretic strength of IT and IT$_c$. The upper bound is not that immediate. A starting point is to look at VF for which Cantini (1990) showed ID$_1$ to be the upper proof theoretic bound and to see whether Cantini's proof strategy works for IT$_c$ likewise. Cantini's strategy was to interpret VF in KPU formulated over number theory.

Theorem 6.2 (Cantini)**.** VF *can be interpreted in* KPU *formulated over number theory.*

But by a result of Jäger (Jäger, 1982) we know that KPU formulated over number theory is proof-theoretically equivalent to ID$_1$, which provides the intended result.

Theorem 6.3 (Jäger)**.** KPU *formulated over number theory and* ID$_1$ *have the same arithmetical theorems.*

Corollary 6.4 (Cantini)**.** VF *and* ID$_1$ *are proof-theoretically equivalent.*

Clearly if we can show that IT$_c$ is interpretable in KPU we could use the same reasoning to conclude that IT$_c$ is proof-theoretically equivalent to ID$_1$. However, since VF is a proper subtheory of IT$_c$ the interpretability of VF in KPU does not guarantee that IT$_c$ is interpretable in KPU. In this section we show how this fact can be established.

We will show that IT$_c$ is interpretable in KPU by, once more, hijacking Cantini's proof. As we have mentioned, Cantini (1990) showed that VF can be interpreted in KPU. For his result Cantini uses the fact that in KPU we can define the layers (stages) of an inductive definition and, as a consequence, we can define a predicate $\mathsf{Der}(\alpha, \rho, \ulcorner\Gamma\urcorner)$, which formalizes $\mathsf{SV}^c_\infty \vdash^\alpha_\rho \Gamma$, i.e., Γ is derivable SV_∞ with height α and truth rank ρ.[39] Cantini proposed to translate a formula Tt by $\mathsf{Der}(t)$, which stands for $\exists\alpha\exists\rho\mathsf{Der}(\alpha, \rho, t)$, while the arithmetical vocabulary is held fixed in the translation. He showed that under this translation VF can be interpreted in KPU. But since the interpretation holds with the arithmetical vocabulary fixed this establishes the upper bound of VF by the aforementioned result of Jäger (1982).

[39] The truth rank keeps track of the number of applications of the truth rules (T) and ($\neg T$).

For this proof to work it is important that we can prove the formalized versions of certain properties of SV^c_∞ in KPU and these properties will also be important for showing that IT_c can be interpreted in KPU.

We start the proof of this latter claim by observing that in KPU we can define a predicate constant $\mathsf{I}^\alpha_{\Theta_c}$, which formalizes the stages of the inductive definition of the minimal fixed point of Θ_c. Now, by translating a formula Tt by $t \in \mathsf{I}_{\Theta_c}$, i.e. $\exists\alpha(t \in \mathsf{I}^\alpha_{\Theta_c})$,—while again keeping the arithmetical vocabulary fixed—we can show that the translations of the axioms (Ax1-7), (Ax.9) and (Ax11) are provable in KPU. Now the remaining axioms of IT_c cannot be checked in a straightforward way under the chosen interpretation of the truth predicate because they cannot simply be read off from the clauses of the inductive definition. Fortunately these remaining axioms are axioms of VF and thus provable under the translation of the truth predicate as the formula Der. Moreover, by Lemma 4.1 we know that derivability in SV^c_∞ coincides with membership in I_{Θ_c} and this fact can be proven in KPU:

Lemma 6.5. *For all $\phi \in \mathcal{L}_\mathsf{T}$*

$$\mathsf{KPU} \vdash \mathsf{Der}(\ulcorner\phi\urcorner) \leftrightarrow \ulcorner\phi\urcorner \in \mathsf{I}_{\Theta_c}.$$

Proof sketch. The left-to-right direction of Lemma 6.5 is by a straightforward formalization of Lemma 4.6. For the converse direction we cannot rely on our previous proofs since it is not clear that they can be formalized in KPU. Fortunately it turns out that the converse direction of Lemma 4.1 can be proved in alternative way, that is, by an induction on the stages of the inductive definition and this proof can be formalized in KPU: as an induction hypothesis we assume that for all $\alpha < \beta$

$$\mathsf{KPU} \vdash \ulcorner\phi\urcorner \in \mathsf{I}^\alpha_{\Theta_c} \rightarrow \mathsf{Der}(\ulcorner\phi\urcorner)$$

and prove by a routine (secondary) induction on the complexity of ϕ that

$$\mathsf{KPU} \vdash \ulcorner\phi\urcorner \in \mathsf{I}^\beta_{\Theta_c} \rightarrow \mathsf{Der}(\ulcorner\phi\urcorner).$$

□

As a consequence of Lemma 6.5 and Cantini's interpretation of VF in KPU we know that the translations of the remaining axioms of IT_c are provable in KPU. Consequently, IT_c can be interpreted in KPU, which establishes the intended upper bound of the proof-theoretic strength of IT_c.

Lemma 6.6. *Let $\phi \in \mathcal{L}_\mathsf{T}$ and $*$ a translation function that commutes with all logical connectives and quantifiers such that $(s = t)^* = (s = t)$ and $(Tt)^* = t \in \mathsf{I}_{\Theta_c}$. Then, if $\mathsf{IT}_c \vdash \phi$, then $\mathsf{KPU} \vdash \phi^*$.*

As we have just mentioned, this establishes the upper proof theoretic bound of IT_c but by inspecting the proof it is immediate that we can run a parallel argument to show the theory IT can also be interpreted in KPU. This together with Friedman and Sheard's result, determines the proof-theoretic strength of IT and IT_c.

Corollary 6.7. IT *and IT_c have the same arithmetical consequences as ID_1.*

By Corollary 6.7 we arrive at the following picture:

$$\mathsf{VF} \equiv \mathsf{IT} \equiv \mathsf{IT}_c \equiv \mathsf{ID}_1 \equiv \Pi^1_1\text{-}\mathsf{CA}^-_0 \equiv \mathsf{KPU}.$$

The theories of inductive truth are thus not only neat proof-theoretic characterizations of Kripke's theory of truth based on supervaluation-style truth schemes but they are also, in general, the strongest $\mathbb{N}$-categorical theories of Kripke's theory of truth that have been proposed so far. With this observation we end our technical investigation of supervaluation-style truth and turn to a quick summary and evaluation of our findings.

7. Conclusion

We think that our work shows that the supervaluation-style truth scheme is an interesting evaluation scheme that leads to an attractive version of Kripke's theory of truth. Penumbral truths will be in the interpretation of the truth predicate under the supervaluation-style truth scheme but the scheme is much simpler and transparent than the supervaluation scheme. This shows in the fact that the evaluation condition, which guarantees that penumbral truths will always be true, can be expressed by a first-order formula. This is most decisively not possible in the supervaluation scheme. Due to its simplicity the supervaluation-style truth scheme remains somewhat closer to the strong Kleene scheme and also retains, at least in parts, the constructive touch of the latter. This should make the supervaluation-style truth scheme particularly interesting for the advocate of supervaluational truth who is interested in grounded truth. It is also this simplicity that enables a neat proof-theoretic characterization of Kripke's theory based on supervaluation-style truth in the form of the theories of inductive truth. We have shown these theories to be $\mathbb{N}$-categorical axiomatization of the supervaluation-style fixed points, while no $\mathbb{N}$-categorical axiomatization of the fixed points of the supervaluation scheme is possible. The theories of inductive truth are from a proof-theoretic perspective fairly strong. Indeed they are the strongest $\mathbb{N}$-categorical theories available to date and considerably stronger than the theory KF, which is an $\mathbb{N}$-categorical axiomatization of the strong Kleene fixed points. Moreover, in the theories of inductive truth, in contrast to KF, logical truths are true in the object-linguistic sense—they are in all supervaluation-style fixed points. However, in KF the truth predicate commutes with disjunction and the existential quantifier, which it does not in the theories of inductive truth. This is of course where the failure of compositionality of the supervaluation-style truth scheme shows. But it is also where we can see that the failure proves to be less drastic than in the supervaluation scheme: in the theories of inductive truth we can state the exact conditions under which a disjunction or an existential statement is true. Nothing of the like is possible in the case of theories like VF, which are inspired by the the supervaluation scheme. We take this to show that the supervaluation-style truth scheme really combines the best of both "worlds" the strong Kleene and the supervaluation world.

One might argue that in contrast to the supervaluation scheme or the strong Kleene scheme, the supervaluation-style truth scheme does not come with a semantic story motivating it. Rather it is *ad hoc* in flavor, so the argument would go, since it is exclusively motivated by the desire of adding penumbral truths to the strong Kleene truth sets. While there is certainly some truth in this, it seems worth pointing out that the supervaluation-style truth jumps can be neatly motivated by the story Kripke uses to motivate his theory of truth in the first place. Kripke described the construction process leading up to suitable interpretations of the truth predicate as of a process of an interlocutor acquiring greater and greater understanding of the language with the truth predicate starting from the same language without the truth predicate. At each stage of this process, so the story goes, the interlocutor comes to understand more and more truth ascriptions of the language. But in such a picture it seems only reasonable to suppose that the interlocutor pauses at each stage to reflect which further truth ascriptions she might infer using her deductive capacities. If we assume the interlocutor to reason in classical first-order logic, then we have a direct motivation for our supervaluation-style truth scheme.

This motivational story might also raise interest in alternative *disjunctive* schemes, that is, schemes that consist of the standard strong Kleene satisfaction relation combined with a suitable closure condition. In such schemes a sentence is true iff it is true according to the strong Kleene scheme *or* it follows from some strong Kleene truth. In the case of supervaluation-style truth the closure condition was taken to be first-order consequence but in principle one could use alternative, possibly non-classical consequence relations to this effect as long as the resulting jump operation will be monotone.[40] This should be of interest to philosophers and logicians who find the motivational story, or, more generally, the strong Kleene construction, appealing but do not agree with the idealizing assumption that takes the interlocutor to reason in classical logic. Moreover, if the consequence relation at stake is first-order arithmetically definable we can apply the strategy sketched at the beginning of Section 5 and obtain an axiomatic theory matching the resulting version of Kripke's theory of truth.[41]

While supervaluation-style truth might not have a neat story that provides independent motivation for the scheme as such, it seems that it is well-motivated within the framework of Kripke's theory of truth. Moreover, the motivation that supports supervaluation-style truth points to a wealth of possible *disjunctive* evaluation schemes that have not been sufficiently explored. Therefore, we hope that our investigation into supervaluation-style truth will serve a twofold purpose. Firstly, we hope to have established supervaluation-style truth as an interesting alternative to the more complicated supervaluation schemes. In particular, the scheme should be appealing to theorists who are interested in penumbral truths and the idea of groundedness but who dislike the complex and intransparent character of the supervaluation scheme.[42] Secondly, we hope to have provided a starting point for further exploration of the unknown territory of *disjunctive* evaluation schemes.

Acknowledgments This work was supported by the DFG-funded research project *Syntactical Treatments of Interacting Modalities*, the German BA, and by the European Commission through a Marie Sklodowska-Curie Individual Fellowship (TREPISTEME, Grant No. 703539). I wish to thank Catrin Campbell-Moore, Martin Fischer, Carlo Nicolai, Philip Welch and two anonymous referees for helpful comments on and discussion of my paper. Early versions of this paper were presented at Bristol, Bath, and Munich. I thank the audiences for their feedback.

[40] If the consequence relation is reflexive we can omit—like in the case of the skk scheme—the first disjunct, that is, the strong Kleene satisfaction relation in the formulation of the scheme.

[41] Such an axiomatic theory of truth should also be possible if the relevant consequence relation is Δ^1_1. However, if the consequence relation is of greater complexity, then this will only be possible if there is a simpler way to define the resulting truth sets.

[42] A further interesting question that we have not addressed in this paper is whether the supervaluation-style truth schemes can be fruitfully applied to the study of vagueness. This would be interesting since vagueness was one of the principle fields of applications of the supervaluation schemes.

References

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

J. P. Burgess. Addendum to "The Truth is Never Simple". *The Journal of Symbolic Logic*, 53(2): 390–392, 1988.

A. Cantini. A theory of formal truth arithmetically equivalent to ID_1. *The Journal of Symbolic Logic*, 55:244–259, 1990.

A. Cantini. *Logical Frameworks for Truth and Abstraction*. Elsevier Science Publisher, Florenz, 1996.

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

K. Fine. Vagueness, Truth and Logic. *Synthese*, 30:265–300, 1975.

M. Fischer, V. Halbach, J. Kriener, and J. Stern. Axiomatizing semantic theories of truth? *The Review of Symbolic Logic*, 8(2):257–278, 2015. doi: 10.1017/S1755020314000379.

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

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

V. Halbach. Reducing compositional to disquotational truth. *The Review of Symbolic Logic*, 2: 786–798, 2009.

V. Halbach. *Axiomatic Theories of Truth*. Cambridge University Press, 2011.

H. G. Herzberger. Paradoxes of grounding in semantics. *The Journal of Philosophy*, 67:145–167, 1970.

G. Jäger. Zur Beweistheorie der Kripke-Platek Mengenlehre. *Archiv für Mathematische Logik und Grundlagenforschung*, 22:121–139, 1982.

P. Kremer and M. Kremer. Some supervaluation-based consequence relations. *Journal of Philosophical Logic*, 32:225–244, 2003.

P. Kremer and A. Urquhart. Supervaluation fixed-point logics of truth. *Journal of Philosophical Logic*, 37:407–440, 2008.

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

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

R. L. Martin and P. W. Woodruff. On Representing "true-in-L" in L. *Philosophia. Philosophical Quarterly of Israel*, 5:213–217, 1975.

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

B. C. van Fraassen. Presupposition, implication, and self-reference. *The Journal of Philosophy*, 65(5):136–152, 1968.

B. C. Van Fraassen. Presuppositions, Supervaluations and Free Logic. In K. Lambert, editor, *The Logical Way of Doing Things*. Yale University Press, New Haven, 1969.

A. Visser. Semantics and the Liar Paradox. In D. Gabbay, editor, *Handbook of Philosophical Logic*, pages 617–706. Dordrecht, 1984.

P. Welch. The Complexity of the Dependence Operator. *Journal of Philosophical Logic*, 44(3): 337–340, 2015.

S. Yablo. Grounding, dependence, and paradox. *Journal of Philosophical Logic*, 11:117–137, 1982.