SUPERVALUATION-STYLE TRUTH REVISITED

PABLO DOPICO & CARLO NICOLAI & JOHANNES STERN

ABSTRACT. Supervaluational fixed-point semantics for truth cannot be axiomatized because of its recursion-theoretic complexity. Johannes Stern (*Supervaluation-Style Truth Without Supervaluations*, Journal of Philosophical Logic, 2018) proposed a new strategy (supervaluational-style truth) to capture the essential aspects of the supervaluational evaluation schema whilst limiting its recursion-theoretic complexity, hence resulting in ($\mathbb{N}$-categorical) axiomatizations. Unfortunately, as we show in the paper, this strategy was not fully realized in Stern's original work: in fact, we provide counterexamples to some of Stern's key claims. However, we also vindicate Stern's project by providing different semantic incarnations of the idea and corresponding $\mathbb{N}$-categorical axiomatizations. The results provide a deeper picture of the relationships between standard supervaluationism and supervaluational-style truth.

1. INTRODUCTION

In an ideal world, semantic and proof-theoretic methods go hand in hand: after all, the first-order completeness theorem tells us that r.e. first-order theories can be characterized one way or the other. Yet, in practice the methods come apart. When working semantically one often focuses on intended models rather than arbitrary models of the language. If we then consider the theories of these intended models, we typically end up with a non r.e. set of sentences, that is, the set of sentences will not be axiomatizable. A prominent example of this mismatch between semantics and proof-theory arises in theories of truth. If one attempts to characterize the truth predicate semantically, one typically defines the (an) interpretation and the resulting truth theory is thus non-axiomatizable.

Nonetheless, axiomatic theories have been proposed as axiomatizations of certain semantic truth theories; for instance, the axiomatic truth theory KF from [Fef91] has been proposed as an axiomatization of the Strong Kleene version of Kripke's theory of truth [Kri75]. The question then arises: according to which criteria can KF be considered as an axiomatization of Kripke's theory of truth? And, more generally, when can an axiomatic truth theory be considered to axiomatize a non r.e. semantic truth theory? This question was addressed by [FHKS15], who propose a number of criteria to this effect. Arguably, amongst these criteria, $\mathbb{N}$-categoricity stands out as particularly compelling: an axiomatic theory Σ is an $\mathbb{N}$-categorical axiomatization of a semantic theory $\text{Th}_{\mathfrak{M}}$, where $\mathfrak{M}$ is a class of interpretations of the truth predicate defined over the standard model $\mathbb{N}$, iff

$$(1) \qquad (\mathcal{N}, S) \vDash \Sigma \text{ iff } S \in \mathfrak{M}.$$

Indeed, the theory KF is an $\mathbb{N}$-categorical axiomatization of Kripke's theory of truth, that is, a set $S \subseteq \omega$ is an interpretation of the truth predicate of KF over the standard model iff S is a fixed point of the Strong Kleene jump SK:

$$(\mathcal{N}, S) \vDash \text{KF iff SK}(S) = S.$$

Whilst it might have its limitations, the criterion of $\mathbb{N}$-categoricity is surely part of answering the question of whether a given axiomatic truth theory counts as an axiomatization of a given semantic truth theory.[1]

It may then come as a surprise that there are semantic truth theories that are not amenable to an $\mathbb{N}$-categorical axiomatization, i.e., to which the criterion of $\mathbb{N}$-categoricity is inapplicable. Indeed both the revision theory of truth (see, e.g., [GB93]) but also Kripke's theory of truth in its supervaluational variant are examples of such theories.[2] The reason is that the definition of the respective class of (intended) models is too complicated in the recursion-theoretic sense: satisfaction (the left side of the biconditional (1)) can be defined by a Δ^1_1-formula of the language of second-order arithmetic. By contrast, the definition of the respective class of models (the right side of the biconditional) requires a Π^1_1-formula in the case of the supervaluation scheme – and even more complicated formulae in the case of the revision theory or Field's theory. This observation entails that well-known theories intended to axiomatize supervaluational truth, notably the theory VF from [Can90], are not $\mathbb{N}$-categorical axiomatizations of the supervaluational version of Kripke's truth theory.

The lack of an $\mathbb{N}$-categorical axiomatization of the supervaluational version of Kripke's truth theory was one of the key aspects that motivated [Ste18]. In essence, he proposed the following:

> THE SSK-THESIS: there is a fixed-point semantics that captures the essential aspects of the supervaluational schema whilst limiting its recursion-theoretic complexity, hence resulting in an $\mathbb{N}$-categorical axiomatization.

Indeed, [Ste18] puts forward such a semantics, *supervaluation-style truth*, based on a satisfaction scheme that he calls SSK – for 'supervaluational Strong Kleene'. Moreover, Stern proposes an axiomatic truth theory, which he labels IT, and which he alleges to be an $\mathbb{N}$-categorical axiomatization of Kripke's truth theory based on the scheme SSK.

Of course, the SSK THESIS is somewhat vague, since we haven't specified what the 'essential aspects' of the supervaluational schema are. This can be clarified as follows. The starting point of Stern's analysis was the observation that the supervaluation schemata

$$(\mathbb{N}, S) \vDash_s \varphi \text{ iff } \forall S' \supseteq S(\Phi(S') \Rightarrow (\mathbb{N}, S') \vDash \varphi),$$

where Φ is an $\mathcal{L}_{\mathrm{Tr}}$-definable condition with parameters, can be understood as collecting classical consequences of the set of sentences Strong Kleene grounded in S over the standard model:

$$\exists G(S \vDash_{\mathrm{sk}} G \,\&\, \forall S' \supseteq S(\Phi(S') \,\&\, (\mathbb{N}, S') \vDash G \Rightarrow (\mathbb{N}, S') \vDash \varphi).[3]$$

The correspondence between the two characterization holds as long as Φ forces a precisification S' to be consistent with the starting hypothesis S, that is, as long as $\Phi(S')$ entails that $\{\varphi \mid \neg\varphi \in S\} \cap S' = \varnothing$.[4] In particular, the supervaluation schemata $\mathrm{VB}, \mathrm{VC}, \mathrm{MC}$ discussed below can be viewed as collecting classical *second-order* consequences of the set of Strong Kleene truths over the hypothesis S.

Stern observed that, instead of taking classical consequences of G relative to a specific class of truth models over the standard model, one can simply consider the classical first-order consequences of the

---

[1] $\mathbb{N}$-categoricity is mainly cogent in the context of classical theories of truth where sentences φ are in general not equivalent to $\mathrm{Tr}\ulcorner\varphi\urcorner$. When such an equivalence is in place – e.g. in the context of axiomatizations of Kripke's fixed-point semantics in a Strong Kleene setting –, $\mathbb{N}$-categoricity may be trivially obtained.

[2] The impossibility to $\mathbb{N}$-categorically axiomatize the revision theory of truth then permeates other truth theories based on revision rules, most notably Field's [Fie08].

[3] G is a set of sentences and by $(\mathbb{N}, S) \vDash_e G$ we intend to convey that $(\mathbb{N}, S) \vDash_e \psi$ for all $\psi \in G$.

[4] The correspondence will thus not hold for the trivial admissibility condition Φ with $\Phi(S') :\leftrightarrow S' = S'$ and thus fails for the scheme sv discussed below.

Strong Kleene-grounded sentences. This constitutes the SSK satisfaction scheme, and gives rise to what Stern called *supervaluation-style semantics*. This approach would lead to a significant reduction of the recursion-theoretic complexity and would even give rise to an arithmetically definable jump (or so he argued). Specifically, [Ste18] claimed that the fixed points of the SSK-scheme are the fixed points of a first-order arithmetical operator Θ (cf. Definition 1), and that the SSK schema has an $\mathbb{N}$-categorical axiomatization in the form of the theory IT. In addition, under these assumptions, he showed that the SSK-scheme has the same minimal fixed point as the supervaluation scheme.

This reconstruction gives us a precise way of understanding the SSK THESIS: the preservation of 'essential aspects' of the supervaluational scheme amounts to recovering all first-order penumbral truths of classical logic and arithmetic, something that one would hope to retain when moving from second-order to first-order formulations of the supervaluational schema.

Unfortunately, as we show in this paper, the SSK THESIS hasn't been successfully realized by [Ste18] since some of Stern's reasoning was flawed. Specifically:

- the fixed points of the operator associated with the SSK-schema and Θ constitute two distinct classes of fixed points;
- the fixed points of the SSK-schema are in general *not* closed under the ω-rule;
- as a consequence, the minimal fixed point of the SSK-schema and the supervaluation scheme differ;
- the theory IT is not $\mathbb{N}$-categorical with respect to SSK-fixed points.

However, not all is lost and, as we show, the SSK-THESIS can be vindicated, albeit in a less uniform way than argued in [Ste18]. In fact, since the fixed points of the SSK schema and Θ come apart, we end up with different semantic constructions – each of them recovering penumbral truths of classical logic and arithmetic – that enjoy $\mathbb{N}$-categorical axiomatizations in virtue of their lower recursion-theoretic complexity.[5] The choice forced upon us is essentially as follows. We can stick to the well-motivated SSK schema but lose the desirable closure under ω-logic. As a consequence, the universal quantifier will not commute with the truth predicate in $\mathbb{N}$-categorical axiomatizations of the SSK schema. Alternatively, we can adopt a more 'standard' arithmetical jump with familiar closure conditions. However, the arithmetical jump will, in contrast to the SSK schema and its associated operator, "merely" collect consequences of a given truth set/hypothesis rather than submitting the hypothesis to semantic scrutiny. This undermines the philosophical and semantic motivation underlying supervaluational-style semantics and leads to some undesirable formal features (e.g. failure of what will be called 'classical soundness').

In the paper we will study the classes of models given by fixed points of SSK, Θ, and Θ^* – a variant of Θ equipped with additional closure properties. We will also provide $\mathbb{N}$-categorical axiomatizations for each of them – as we have seen, a highly desirable feature. After describing some preliminaries in Section 2 and Section 3, we isolate some gaps and problems in the previous attempt to realize the SSK THESIS by [Ste18] in Section 4. These include SSK's failure to be closed under ω-logic. Section 5 provides two $\mathbb{N}$-categorical axiomatizations of SSK, that we will dub PK and PK$^+$. We then move to the study of Θ and its variation Θ^*: Section 6 presents $\mathbb{N}$-categorical axiomatizations for (the fixed points of) each operator, and establishes lower-bounds for their proof-theoretic strength. Section 7 draws connections between supervaluation-style fixed-point semantics and the standard supervaluational approaches: although the minimal fixed points of VB (a class of supervaluational fixed points that imposes a weak consistency condition), Θ and Θ^* coincide, this is not true for all fixed points. In fact, every VB fixed point is a

---

[5]We recall that, because of the higher recursion-theoretic complexity of supervaluational fixed-point semantics, $\mathbb{N}$-categorical axiomatizations will not be available for the latter. In some sense it is impossible to equip supervaluational fixed-point semantics with a fitting proof theory. This shows, for example, in the fact that the models of VF are not just the supervaluational fixed points but, in addition, fixed points of Θ or even the set of stable truth of the Herzberger sequence over the empty set [Can96, §96]. It thus seems difficult to associate VF with supervaluational fixed points in some strong sense.

Θ^*, and hence a Θ fixed point, but the converse provably fails. We conclude by assessing the results presented in the paper along some key dimensions.

2. Preliminaries and notation

2.1. **Language, Theories, Coding.** Our notation follows, for the most part, the conventions in [Hal14]. We will be working with languages whose logical symbols are $\neg, \vee, \wedge, \forall, \exists$. We also write $\varphi \to \psi$ as an abbreviation for $\neg\varphi \vee \psi$. As a default practice, our base language $\mathcal{L}_{\mathbb{N}}$ is a definitional extension of the language of Peano Arithmetic (PA) with finitely many function symbols for primitive recursive functions. We assume a standard formalization of the syntax of first-order languages as it is for instance carried out in [HP17]. The expression $\bar{n}$ stands for the numeral of the number n (although we omit the bar for specific numbers). We write $\#\varphi$ for the code or Gödel number of φ, and $\ulcorner\varphi\urcorner$ for the numeral of that code. $\mathcal{L}_{\mathbb{N}}$ is assumed to contain a finite set of function symbols that will stand for certain primitive recursive operations. In particular, we want to include function symbols for some primitive recursive syntactic operations on expressions of the language under consideration; for that, we use the traditional subdot notation. For example, $\dot{\neg}$ is the function symbol for the primitive recursive operation that, when inputted the code of a formula, yields the code of its negation. The same applies to $\dot{\vee}, \dot{\forall}, \dot{\wedge}, \dot{\exists}$. We assume a function symbol for the substitution function, and write $x(t/v)$ for the result of formally substituting v with t in x; $\ulcorner\varphi(\dot{x})\urcorner$ abbreviates $\ulcorner\varphi(v)\urcorner(\text{num}(x)/\ulcorner v\urcorner)$, for x a term variable (and provided φ only has one free variable) and $\text{num}(x)$ the function symbol corresponding to the operation of sending a number to its numeral. Furthermore, we avail ourselves to the notation t° for the result of applying to a term t the evaluation function (which outputs the value of the inputted term). Note that $\cdot^\circ$ is an abbreviation for a formula, and not a term of the language. The development of syntax just mentioned can be comfortably carried out in Peano arithmetic (PA), which will be our background theory of syntax.

For the formulation of the theories of truth, we will (mostly) be concerned with the extension of $\mathcal{L}_{\mathbb{N}}$ with a (unary) truth predicate Tr. PAT is then the theory formulated in the language $\mathcal{L}_{\text{Tr}}$ consisting of the axioms of PA with induction extended to the whole language. We will then write $\text{Var}_{\mathcal{L}_{\text{Tr}}}(x)$ for the formula representing the set of (codes of) variables, $\text{CT}_{\mathcal{L}_{\text{Tr}}}(x)$ for the formula representing the set of (codes of) closed terms of $\mathcal{L}_{\text{Tr}}$, $\text{For}_{\mathcal{L}_{\text{Tr}}}(x)$ ($\text{For}_{\mathcal{L}_{\text{Tr}}}(x, y_0, ..., y_n)$), for the formula representing the set of all formulae (formulae with $y_0, ... y_n$ free) of $\mathcal{L}_{\text{Tr}}$, and $\text{Sent}_{\mathcal{L}_{\text{Tr}}}(x)$, for the formula representing the set of all sentences of $\mathcal{L}_{\text{Tr}}$. We also use $\text{Var}_{\mathcal{L}_{\text{Tr}}}$, $\text{CT}_{\mathcal{L}_{\text{Tr}}}$, $\text{For}_{\mathcal{L}_{\text{Tr}}}$, and $\text{Sent}_{\mathcal{L}_{\text{Tr}}}$ to stand for the sets corresponding to these formulae. We will occasionally omit reference to $\mathcal{L}_{\text{Tr}}$ when this is clear from the context. We often quantify over variables s, t to refer to codes of closed terms, so that $\forall t\, \varphi(t)$ is short for $\forall x(\text{CT}_{\mathcal{L}_{\text{Tr}}}(x) \to \varphi(x))$.

We assume an appropriate ordinal notation system OT up to Γ_0, and the ordinal less-than relation $<$ on Γ_0 (for more detailes, see, e.g., [Poh09, Ch.3]). For simplicity, we identify each ordinal number with its notation. We use 0 as the ordinal number, the ordinal sum $\alpha + \beta$, and the Veblen function $\varphi_\alpha\beta$. Given a language $\mathcal{L}$, an ordinal number $\alpha < \Gamma_0$, and an $\mathcal{L}$-formula A, transfinite induction for A up to α is defined as the formula:

$$\text{TI}(\alpha, A) := \forall\beta(\forall\gamma < \beta A(\gamma) \to A(\beta)) \to \forall\beta < \alpha A(\beta).$$

In line with the latter, the schema $\text{TI}_{\mathcal{L}}(< \alpha)$ is defined to be the set $\{\text{TI}(\beta, A) \mid \beta < \alpha \,\&\, A \in \mathcal{L}\}$. Finally, if S is a theory, its proof-theoretic ordinal (denoted $|S|$) is the ordinal α such that $\text{PA} + \text{TI}_{\mathcal{L}_{\mathbb{N}}}(< \alpha)$ and S prove the same $\mathcal{L}_{\mathbb{N}}$-statements (and verifiably so within PA). If T, S are theories, we write $|T| \geqq |S|$ to indicate that the proof-theoretic ordinal of T is equal or greater than the proof-theoretic ordinal of S. Similarly, $|T| \equiv |S|$ is defined as $|T| \geqq |S|$ and $|S| \geqq |T|$.

2.2. **Supervaluations.** In his seminal [Kri75], Kripke proposes to determine the extension of the truth predicate via a fixed-point construction. Famously, this construction is carried out in Strong Kleene (SK) logic. However, SK fails to capture many *penumbral* truths, that is, truths that hold merely in virtue of the logical relations occurring between sentences of a language. An example of such a sentence is $\lambda \vee \neg\lambda$, for λ the Liar sentence. As a remedy, Kripke in his paper also suggests to run the construction with some form of supervaluational logic. This will be our departure point. For $X \subseteq \mathrm{Sent}_{\mathcal{L}_{\mathrm{Tr}}}$, we define

$$\mathrm{CON}(X) := \neg\exists\varphi(\{\#\varphi, \#\neg\varphi\} \subseteq X);$$
$$\mathrm{MXC}(X) := \neg\exists\psi(\{\#\psi, \#\neg\psi\} \subseteq X \wedge (\forall\varphi(\#\varphi \notin X \rightarrow \#\neg\varphi \in X))).$$

$\mathrm{MXC}(X)$ expresses that X is a maximally consistent set of $\mathcal{L}_{\mathrm{Tr}}$-sentences.

For a satisfaction relation e, we will write $X \vDash_e A$ to abbreviate $(\mathbb{N}, X, X^-) \vDash_e A$, where $X \subseteq \mathrm{Sent}_{\mathcal{L}_{\mathrm{Tr}}}$ is the relevant interpretation of the truth predicate and X^- is defined as $\{\#\varphi \mid \#\neg\varphi \in X\} \cup \{n \in \omega \mid \mathbb{N} \vDash \neg\mathrm{Sent}_{\mathcal{L}_{\mathrm{Tr}}}(\bar{n})\}$. Similarly, for the classical satisfaction relation ($\vDash$), $X \vDash A$ just stands for $(\mathbb{N}, X) \vDash A$. In the literature, one often distinguishes the following supervaluational satisfaction relations:

(2) $$X \vDash_{\mathrm{sv}} \varphi \text{ iff } \forall X'(X' \supseteq X \Rightarrow X' \vDash \varphi)$$

(3) $$X \vDash_{\mathrm{vb}} \varphi \text{ iff } \forall X'(X' \supseteq X \ \& \ X' \cap X^- = \varnothing \Rightarrow X' \vDash \varphi)$$

(4) $$X \vDash_{\mathrm{vc}} \varphi \text{ iff } \forall X'(X' \supseteq X \ \& \ \mathrm{CON}(X') \Rightarrow X' \vDash \varphi)$$

(5) $$X \vDash_{\mathrm{mc}} \varphi \text{ iff } \forall X'(X' \supseteq X \ \& \ \mathrm{MCX}(X') \Rightarrow X' \vDash \varphi)$$

To avoid triviality, and following the truth-theoretic literature on supervaluationism, we are mainly interested in consistent set of sentences. For $\mathrm{CON}(X)$ and $j \in \{\mathrm{sv}, \mathrm{vb}, \mathrm{vc}, \mathrm{mc}\}$, we let:

(6) $$J(X) := \{\#\varphi \mid X \vDash_j \varphi\}$$

$J(\cdot)$, for $J \in \{\mathrm{SV}, \mathrm{VB}, \mathrm{VC}, \mathrm{MC}\}$, is the *J-supervaluational jump*. It can be checked that, for arbitrary $X \subseteq \mathrm{Sent}_{\mathcal{L}_{\mathrm{Tr}}}$, the jump is monotone. It then follows by the theory of positive inductive definitions (see e.g. [Mos74, Ch.1], [McG91, Ch.5]) that $J(\cdot)$ has a fixed point, and in particular a minimal one obtained by iterating the operator over the empty set.

It is worth noting that, to establish the existence of non-trivial fixed points, we need a further property to guarantee that at every stage of the inductive definition we can find an admissible precisification. Otherwise triviality ensues. One property that guarantees the existence of non-trivial fixed points is to require the admissibility condition to be compact on the set of consistent interpretations: let $X := \{S \mid \mathrm{CON}(S)\}$, $\Phi^S := \{S' \mid S \subseteq S' \ \& \ \Phi(S')\}$ and for $Y \subseteq X$ set $\Phi(Y) := \{\Phi^S \mid S \in Y\}$. Then Φ is *compact on* X iff for all $Y \subseteq X$:

$$\text{if } \Phi^{S_1} \cap \ldots \cap \Phi^{S_n} \neq \varnothing \text{ for all } S_1, \ldots S_n \in Y \text{ and all } n \in \omega, \text{ then } \bigcap \Phi(Y) \neq \varnothing.$$

The admissibility condition associated with the schemes $\mathrm{SV}, \mathrm{VB}, \mathrm{VC}, \mathrm{MC}$ are compact in that sense.

3. Supervaluation-style truth

While Kripke's fixed-point semantics over supervaluational logics succeeds in recovering all penumbral truths, the theory is not free from objections. As argued in [Ste18] supervaluational truth fails to be compositional and, because of its recursion-theoretic complexity, is not $\mathbb{N}$-categorically axiomatizable. Accordingly, Stern develops a new theory of truth, which he calls *supervaluational-style truth*, aimed at addressing some of these shortcomings.

As we have seen, to address such worries Stern notes that supervaluation *satisfaction* involves recognizable *Strong-Kleene* and *classical* components. Stern studies a monotone evaluation schema isolating

these two elements of supervaluational satisfaction:

(7) $$X \vDash_{\text{ssk}} \varphi \text{ iff } \exists \psi (X \vDash_{\text{sk}} \psi \text{ and } \text{PAT} \vdash \psi \rightarrow \varphi)$$

In the definition of $\vDash_{\text{ssk}}$, $\vDash_{\text{sk}}$ is the familiar definition of satisfaction in Strong Kleene logic for $\mathcal{L}_{\text{Tr}}$. For $X \subseteq \omega$ consistent, the *Strong-Kleene supervaluational jump* is then:

(8) $$\text{SSK}(X) := \{ \#\varphi \mid X \vDash_{\text{ssk}} \varphi \}.$$

The jump is monotone, and therefore will have fixed points, including a minimal one.

To study the Strong-Kleene supervaluational jump, Stern also introduces a related jump based on a positive inductive definition in $\mathcal{L}_{\mathbb{N}}$ (with parameters).

Definition 1. *Let $\xi(x, X)$ abbreviate the familiar formula of $\mathcal{L}_{\mathbb{N}}$ employed to characterize Strong Kleene truth:*

$$
\begin{aligned}
& x \in \text{True}_0 \\
& \vee \, \exists y (x = \ulcorner \neg \urcorner y \wedge y \in X) \\
& \vee \, \exists y, z (x = (y \dot{\vee} z) \wedge (y \in X \vee z \in X)) \\
& \vee \, \exists y, z (x = \dot{\neg}(y \dot{\vee} z) \wedge (\dot{\neg} y \in X \wedge \dot{\neg} z \in X)) \\
& \vee \, \exists y, z (x = (y \dot{\wedge} z) \wedge (y \in X \wedge z \in X)) \\
& \vee \, \exists y, z (x = \dot{\neg}(y \dot{\wedge} z) \wedge (\dot{\neg} y \in X \vee \dot{\neg} z \in X)) \\
& \vee \, \exists y, v (x = (\dot{\forall} v y) \wedge \forall z (y(z/v) \in X)) \\
& \vee \, \exists y, v (x = \dot{\neg}(\dot{\forall} v y) \wedge \exists z (\dot{\neg} y(z/v) \in X)) \\
& \vee \, \exists y, v (x = (\dot{\exists} v y) \wedge \exists z (y(z/v) \in X)) \\
& \vee \, \exists y, v (x = \dot{\neg}(\dot{\exists} v y) \wedge \forall z (\dot{\neg} y(z/v) \in X)) \\
& \vee \, \exists t (x = \dot{\text{T}} \text{r}(t) \wedge t^\circ \in X) \\
& \vee \, \exists t (x = \dot{\neg} \dot{\text{T}} \text{r}(t) \wedge (\dot{\neg} t^\circ) \in X \vee \neg \text{Sent}(t^\circ))
\end{aligned}
$$

In the definition, True_0 *is the set of all codes of true arithmetic literals. Then define*

$$\theta(x, X) :\leftrightarrow \exists y (\xi(y, X) \wedge \text{Pr}_{\text{PAT}}(y \dot{\rightarrow} x)).$$

Finally, the operator $\Theta : \mathcal{P}(\omega) \rightarrow \mathcal{P}(\omega)$ *is defined as*

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

Stern provides arguments to show that (i) the fixed points of SSK and Θ coincide and, making crucial use of (i), that (ii) there is an $\mathbb{N}$-categorical (in the sense of [FHKS15]) axiomatization of those fixed points. To achieve (ii), Stern introduces the following theory.

Definition 2. *The theory* IT *extends* PAT *with the following axioms:*

IT1 $\forall s, t (\text{Tr}(s \dot{=} t) \leftrightarrow s^\circ = t^\circ) \wedge \forall s, t (\text{Tr}(s \dot{\neq} t) \leftrightarrow s^\circ \neq t^\circ)$

IT2 $\forall x, y (\text{Sent}(x \dot{\wedge} y) \rightarrow (\text{Tr}(x) \wedge \text{Tr}(y) \rightarrow \text{Tr}(x \dot{\wedge} y)))$

IT3 $\forall x, y (\text{Sent}(x \dot{\vee} y) \rightarrow (\text{Tr}(x) \vee \text{Tr}(y) \vee \exists z (\xi(z, \text{Tr}) \wedge \text{Pr}_{\text{PAT}}(z \dot{\rightarrow} x \dot{\vee} y)) \leftrightarrow \text{Tr}(x \dot{\vee} y)))$

IT4 $\forall x, v (\text{Sent}(\dot{\forall} v x) \rightarrow (\forall t \text{Tr}(x(t/v)) \rightarrow \text{Tr}(\dot{\forall} v x)))$

IT5 $\forall x, v (\text{Sent}(\dot{\exists} v x) \rightarrow (\exists t \text{Tr}(x(t/v)) \vee \exists w (\xi(w, \text{Tr}) \wedge \text{Pr}_{\text{PAT}}(w \dot{\rightarrow} \dot{\exists} v x)) \leftrightarrow \text{Tr}(\dot{\exists} v x)))$

IT6 $\forall x (\text{Tr}(x) \leftrightarrow \text{Tr}(\dot{\text{T}} \text{r} x))$

IT7 $\forall x (\text{Tr}(\dot{\neg} x) \vee \neg \text{Sent}(x) \leftrightarrow \text{Tr}(\dot{\neg} \dot{\text{T}} \text{r} x))$

IT8 $\forall x, y (\text{Tr}(x) \wedge \text{Pr}_{\text{PAT}}(x \dot{\rightarrow} y) \rightarrow \text{Tr}(y))$

IT9 $\forall x (\text{Tr}(\dot{\neg} x) \rightarrow \neg \text{Tr}(x))$

IT10 $\mathrm{Tr}\ulcorner \forall x(\mathrm{Tr}(x) \to \mathrm{Sent}(x))\urcorner$

Tr-Out $\forall t_1, ..., t_n(\mathrm{Tr}\ulcorner \varphi(t_1, ..., t_n)\urcorner \to \varphi(t_1, ..., t_n)$ *for every formula* $\varphi(t_1, ..., t_n)$ *of* $\mathcal{L}_{\mathrm{Tr}}$.

As it turns out the argument in support of (i) was flawed and, as a consequence, neither (i) nor (ii) can be achieved.

4. Supervaluational-style truth: some problems

Stern tried to establish that the fixed points of SSK and Θ coincide, i.e., that $S = \Theta(S) \Leftrightarrow S = \mathrm{SSK}(S)$. However, the proof relied on the mistaken assumption that $\Theta(S) \subseteq \mathrm{SSK}(S)$ for any consistent set of formulas S (see [Ste18, Lemma 3.4]). This fails for the following reason. We say that a set $X \subseteq \omega$ is sound with respect to some operator $\mathcal{J} : \mathcal{P}(\omega) \longrightarrow \mathcal{P}(\omega)$ when $X \subseteq \mathcal{J}(X)$. Then, the above failure is to be located in the fact that *any* set of formulas is sound with respect to Θ, and this does not happen for SSK—not even when the set of formulas is consistent.

Proposition 3. *Let S be a any set of codes of formulae of $\mathcal{L}_{\mathrm{Tr}}$. Then, $S \subseteq \Theta(S)$.*

Proof. If $\#\varphi \in S$, then $\mathbb{N} \vDash \xi(\ulcorner \varphi \wedge \varphi \urcorner, S)$ and, since $\varphi \wedge \varphi \to \varphi$ is provable in PAT, $\#\varphi \in \Theta(S)$. $\quad\square$

Proposition 3 immediately entails that iterations of Θ over any $X \subseteq \omega$ reach a fixed point. These fixed points may turn out to be rather ill-behaved – even if we focus on the non-trivial, consistent fixed points:

Proposition 4. *There are formulae φ, ψ of $\mathcal{L}_{\mathrm{Tr}}$ and S a consistent set of codes of sentences such that:*
  1. *$S = \Theta(S)$ and $\#\mathrm{Tr}\ulcorner \varphi \urcorner \in S$ but $\#\varphi \notin S$.*
  2. *$S = \Theta(S)$ and $\#\neg\mathrm{Tr}\ulcorner \varphi \urcorner \in S$ but $\#\neg\varphi \notin S$.*
  3. *$S = \Theta(S)$ and $\#\varphi \vee \psi \in S$ but $\#\varphi \notin S$ and $\#\psi \notin S$ and there is no χ such $S \vDash_{\mathrm{sk}} \chi$ and $\mathrm{PAT} \vdash \chi \to \varphi \vee \psi$.*
  4. *$S = \Theta(S)$ and $\#\exists v\varphi \in S$ but $\#\varphi(\bar{n}) \notin S$ for all $n \in \omega$ and and there is no χ such $S \vDash_{\mathrm{sk}} \chi$ and $\mathrm{PAT} \vdash \chi \to \exists v\varphi$.*

Proof. For 1: consider the fixed point that is obtained by iterating Θ over the set $X = \{\#\mathrm{Tr}\ulcorner \neg\tau \urcorner\}$, for τ a truth-teller. For 2: consider the fixed point that is obtained by iterating Θ over the set $X = \{\#\neg\mathrm{Tr}\ulcorner \neg\tau \urcorner\}$. For 3: consider the fixed point S that is obtained by iterating Θ over the set $X = \{\#\tau_0 \vee \tau_1\}$; clearly, $\#\tau_0 \notin S$ or $\#\tau_1 \notin S$. Now suppose there is a χ such that $S \vDash_{\mathrm{sk}} \chi$ and $\mathrm{PAT} \vdash \chi \to \tau_1 \vee \tau_2$. We obtain $S \vDash \chi$[6] and $S \vDash \chi \to \tau_0 \vee \tau_1$ and thus $S \vDash \tau_0 \vee \tau_1$. By the propeties of τ_0 and τ_1 this in turn implies that either $\#\tau_0 \in S$ or $\#\tau_1 \in S$. Contradiction. For 4: consider the fixed point that is obtained by iterating Θ over the set $X = \{\#\exists x(\tau \wedge x \geq 0)\}$, and apply the reasoning for (3) above. $\quad\square$

The proposition entails the key property of classical soundness (cf. Def. 8) does not obtain for all Θ fixed-points. Classical soundness will play a prominent role in what follows.

Corollary 5. *There are set of codes of sentences S such that $S \subseteq \Theta(S)$ and $\#\varphi \in S$ but $S \nvDash \varphi$. In particular, there are fixed points $S = \Theta(S)$ such that $\#\varphi \in S$ but $S \nvDash \varphi$.*

Proposition 3 fails for SSK. By taking, for instance, $S = \{\#\tau_0 \vee \tau_1\}$, where τ_0 and τ_1 are truth-tellers, one can immediately see that $S \nsubseteq \mathrm{SSK}(S)$. However, although none of the problems above apply to SSK fixed points, the latter are not very attractive either, as they are not closed under ω-logic, a basic feature of SK fixed points.

---

[6]We note that for consistent sets of sentences S we have that

$$\text{if } S \vDash_{\mathrm{sk}} \varphi, \text{ then } S \vDash \varphi,$$

for all sentences φ.

Proposition 6. *There are fixed points X of SSK and a formula $\varphi(x)$ of $\mathcal{L}_{\mathrm{Tr}}$ such that $\#\varphi(\bar{n}) \in X$ for any $n \in \omega$, but $\#\forall y\, \varphi \notin X$.*

Proof. By the diagonal lemma, define a parametrized truth-teller:

$$(9) \qquad\qquad\qquad \tau(x) :\leftrightarrow \mathrm{Tr}(\ulcorner \tau(x) \urcorner (\mathrm{num}(x)/\ulcorner x \urcorner)).$$

Assuming the ordinal notation system introduced in 2.1, let

$$\mathrm{TI}(\tau, \alpha) :\leftrightarrow \mathrm{Prog}(\tau) \to (\forall x < \alpha)\, \tau(x),$$

Now consider I_{ssk}. Since, it is well-known that $\mathrm{PAT} \vdash \mathrm{TI}(\tau, f(n))$, with $f(0) = 0$, $f(n+1) = \omega^{f(n)}$, $\varepsilon_0 := \sup\{f(n)\,|\,n \in \omega\}$, $\mathrm{TI}(\tau, f(n)) \in I_{\mathrm{ssk}}$.

Since $\mathrm{PAT} \nvdash \mathrm{TI}(\tau, \varepsilon_0)$, it can be shown by induction on the construction of I_{ssk} that $\mathrm{TI}(\tau, \varepsilon_0) \notin I_{\mathrm{ssk}}$. However, assuming $\mathrm{Prog}(\tau)$, an application of the ω-rule would enable one to go from $\forall x(x < f(\bar{n}) \to \tau(x)))$ for each n to $\forall y, x(x < f(y) \to \tau(x))$, which would entail the consequent of $\mathrm{TI}(\tau, \varepsilon_0)$, hence a contradiction.

$\square$

Remark 1. Nothing rests on the choice of the truth-teller in the previous proposition. For instance, we could have used a parametrized liar sentence or other 'ungrounded' sentences. It's worth noting that the 'SK-fragment' of the construction is still closed under the ω-rule, in the sense that, if each of the $A(\bar{n})$ are SK-satisfied in the relevant parameter, so is $\forall x A$.

The issues above also imply that Stern's theory IT is not an $\mathbb{N}$-categorical axiomatization of the fixed points of either SSK or Θ. In particular:

Proposition 7. *There are sets of codes of sentences S such that:*
1. *$S = \Theta(S)$ and $S \nVdash \forall t_1, ..., t_n(\mathrm{Tr}\ulcorner \varphi(t_1, ..., t_n) \urcorner \to \varphi(t_1, ..., t_n))$ for some formula $\varphi(t_1, ..., t_n)$ of $\mathcal{L}_{\mathrm{Tr}}$.*
2. *$S = \Theta(S)$ and $S \nVdash \forall t(\mathrm{Tr}(\neg \underaccent{\cdot}{\mathrm{T}}rt) \to \mathrm{Tr}(\neg t) \vee \neg \mathrm{Sent}(t))$*
3. *$S = \mathrm{SSK}(S)$ and $S \nVdash \forall v, x(\mathrm{Sent}(\forall vx) \to (\forall t \mathrm{Tr}(x\,(t/v)) \to \mathrm{Tr}(\forall vx)))$*
4. *$S = \mathrm{SSK}(S)$ and $S \nVdash \forall x, y(\mathrm{Sent}(x\vee y) \to (\mathrm{Tr}(x) \vee \mathrm{Tr}(y) \vee \exists z(\xi(z, \mathrm{Tr}) \wedge \mathrm{Pr}_{\mathrm{PAT}}(z \underaccent{\cdot}{\to} x\vee y)) \to \mathrm{Tr}(x\vee y)))$*
5. *$S = \mathrm{SSK}(S)$ and $S \nVdash \forall x, y(\mathrm{Sent}(\exists vx) \to (\exists t\mathrm{Tr}(x\,(t/v)) \vee \exists w(\xi(w, \mathrm{Tr}) \wedge \mathrm{Pr}_{\mathrm{PAT}}(w \underaccent{\cdot}{\to} \exists vx)) \to \mathrm{Tr}(\exists vx)))$*

Hence, no theory featuring Tr-Out *or* IT7 *(right-to-left) can be* $\mathbb{N}$-*categorical with respect to fixed points of* SSK, *and no theory featuring* IT4, IT3, *or* IT5 *can be* $\mathbb{N}$-*categorical with respect to fixed points of* Θ.

Proof. (1) follows from Proposition 4.1. (2) follows from Proposition 4.2. (3) follows from Proposition 6. (4) follows by letting z be the code for the formula $\mathrm{TI}(\tau, \varepsilon_0)$ as given in Proposition 6, and $x\vee y$ be the code for $\mathrm{TI}(\tau, \varepsilon_0) \vee \mathrm{TI}(\tau, \varepsilon_0)$: by inspecting the proof of that proposition, it is easy to see that $(\mathbb{N}, S) \vDash \xi(z, \mathrm{Tr})$. Finally, (5) follows, just like (4), by letting w be the code of $\mathrm{TI}(\tau, \varepsilon_0)$ and $\exists vx$ be the code for $\exists x_0(\mathrm{TI}(\tau, \varepsilon_0) \wedge x_0 > 0)$. $\square$

Although, by 1 and 2 of Proposition 7, Θ cannot be categorically axiomatized by IT, some fixed points of Θ are still models for IT. In particular, all fixed points that are classically sound, Tr-downwards closed, and $\neg \mathrm{Tr}$-downwards closed, will be models of IT (cf. Proposition 27). Since these notions will play a prominent role in what follows, we provide precise definitions for them.

Definition 8. *Let $S \subseteq \omega$, then*
1. *S is classically sound iff $(\mathbb{N}, S) \vDash \varphi$ for all $\#\varphi \in S$;*
2. *S is Tr-downwards closed iff $\#\mathrm{Tr}\ulcorner \varphi \urcorner \in S$ implies $\#\varphi \in S$;*

(Ax. 1) $\quad \left|\frac{\alpha}{\rho}\right.\ \Gamma, \varphi$, for φ a literal of $\mathcal{L}_{\mathbb{N}}$;

(Ax. 2) $\quad \left|\frac{\alpha}{\rho}\right.\ \Gamma, \mathrm{Tr}(t), \neg\mathrm{Tr}(s)$, with $\#s^{\mathbb{N}} = \#t^{\mathbb{N}}$

$$\frac{\left|\frac{\alpha_0}{\rho}\right.\ \Gamma, \varphi_i, \varphi_0 \vee \varphi_1 \ \text{ for } i \in \{0,1\}}{\left|\frac{\alpha}{\rho}\right.\ \Gamma, \varphi_0 \vee \varphi_1} \ (\vee) \qquad \frac{\left|\frac{\alpha_0}{\rho}\right.\ \Gamma, \varphi_0, \varphi_0 \wedge A_1 \quad \left|\frac{\alpha_0}{\rho}\right.\ \Gamma, \varphi_1, \varphi_0 \wedge \varphi_1}{\left|\frac{\alpha}{\rho}\right.\ \Gamma, \varphi_0 \wedge \varphi_1} \ (\wedge)$$

$$\frac{\left|\frac{\alpha_0}{\rho}\right.\ \Gamma, \exists x \varphi, \varphi(\overline{n}) \ \text{ for some } n \in \omega}{\left|\frac{\alpha}{\rho}\right.\ \Gamma, \exists x \varphi} \ (\exists) \qquad \frac{\left|\frac{\alpha_i}{\rho}\right.\ \Gamma, \forall x \varphi, \varphi(\overline{n}) \ \text{ for all } n \in \omega \text{ and some } \alpha_i < \alpha}{\left|\frac{\alpha}{\rho}\right.\ \Gamma, \forall x \varphi} \ (\omega)$$

$$\frac{\left|\frac{\alpha_0}{\rho_0}\right.\ \varphi}{\left|\frac{\alpha}{\rho}\right.\ \Gamma, \mathrm{Tr}\ulcorner\varphi\urcorner} \ \text{Tr-intro} \qquad \frac{\left|\frac{\alpha_0}{\rho_0}\right.\ \neg\varphi}{\left|\frac{\alpha}{\rho}\right.\ \Gamma, \neg\mathrm{Tr}\ulcorner\varphi\urcorner} \ \neg\text{Tr-intro}$$

Table 1. SV_∞

(3) S is $\neg$Tr-downwards closed iff $\#\neg\mathrm{Tr}\ulcorner\varphi\urcorner \in S$ when $\#\neg\varphi \in S$.

Proposition 4 showed that some Θ fixed points are not Tr-downwards closed, and hence not classically sound, and some are not $\neg$Tr-downwards closed. This being said, some Θ fixed points meet all the conditions: they are classically sound, Tr- and $\neg$Tr-downwards closed. One such fixed point is the minimal fixed point.

Proposition 9. $\mathrm{I}_\Theta \vDash \mathrm{IT}$.

Proof. We define the Tait language $\mathcal{L}_{\mathrm{Tr}}^{\mathrm{Tait}}$ as follows: given a formula A of $\mathcal{L}_{\mathrm{Tr}}$, we take the negation-normal form of A, i.e., negations are only allowed on literals, and double negations are eliminated from literals. Hence, any formula of $\mathcal{L}_{\mathrm{Tr}}^{\mathrm{Tait}}$ is the result of taking literals or negated literals and applying the connectives $\wedge, \vee, \forall, \exists$. We can safely assume that $\mathcal{L}_{\mathrm{Tr}}$ and $\mathcal{L}_{\mathrm{Tr}}^{\mathrm{Tait}}$ are extensionally equivalent, modulo the definition of negation.

Now take the system SV_∞, the infinitary (one-sided) sequent calculus in the language $\mathcal{L}_{\mathrm{Tr}}^{\mathrm{Tait}}$ given by the relation $\left|\frac{\alpha}{\rho}\right.\ \Gamma$, whose axioms and rules are displayed in Table 1. We write $\mathrm{SV}_\infty \vdash \varphi$ iff there are $\alpha, \rho \in \mathrm{On}$ s.t. $\left|\frac{\alpha}{\rho}\right.\ \varphi$. Cantini showed that the following two rules are admissible, given the invertibility properties of the system:

$$\frac{\mathrm{SV}_\infty \vdash \mathrm{Tr}\ulcorner\varphi\urcorner}{\mathrm{SV}_\infty \vdash \varphi} \ \text{Tr-Elim} \qquad\qquad \frac{\mathrm{SV}_\infty \vdash \neg\mathrm{Tr}\ulcorner A\urcorner}{\mathrm{SV}_\infty \vdash \neg A} \ \neg\text{Tr-Elim}$$

One can then show that, for any formula φ, $\mathrm{SV}_\infty \vdash \varphi \Leftrightarrow \#\varphi \in \mathrm{I}_\Theta$. The left-to-right direction is proved in [Ste18, Lemma 4.4]. The right-to-left direction is proven by induction on the stages of the construction of I_Θ, with a subinduction on the positive complexity of the φ. Finally, by induction on the length of the derivation in SV_∞, one can show that $\mathrm{SV}_\infty \vdash \varphi$ implies $(\mathbb{N}, \{\psi \mid \mathrm{SV}_\infty \vdash \psi\}) \vDash \varphi$. This suffices to prove that $(\mathbb{N}, \mathrm{I}_\Theta)$ is a model for axioms IT6, IT7 and Tr-Out. Verifying the rest is immediate. $\qquad\square$

5. The $\mathbb{N}$-categoricity of SSK

One major difficulty in achieving an $\mathbb{N}$-categorical axiomatization for SSK-fixed points is the inexpressibility, in first-order logic, of the notion of (Strong Kleene) satisfaction built in its definition. The satisfaction relation is hyperarithmetical, and hence inexpressible in $\mathcal{L}_{\mathrm{Tr}}$. This does not exclude, however, that there may be ways of capturing enough of Strong-Kleene satisfaction to obtain an extensionally correct characterization of SSK-fixed points. In this section we provide one such axiomatization.

By the diagonal lemma, PAT proves that there is a formula $\pi(x)$ such that:

$$\pi(x) \leftrightarrow \mathrm{True}_0(x) \vee$$
$$\exists t(x = \mathrm{T\dot{r}}t \wedge \mathrm{Tr}t^\circ) \vee$$
$$\exists t(x = \neg\mathrm{T\dot{r}}t \wedge (\mathrm{Tr}\neg\dot{\,}t^\circ \vee \neg\mathrm{Sent}_{\mathcal{L}_{\mathrm{Tr}}}(t^\circ)) \vee$$
$$\exists y(x = \neg\dot{\,}\neg\dot{\,}y \wedge \mathrm{Tr}^\ulcorner \pi(\dot{y})^\urcorner) \vee$$
$$\exists y, z(x = (y\dot{\wedge}z) \wedge \mathrm{Tr}^\ulcorner \pi(\dot{y})^\urcorner \wedge \mathrm{Tr}^\ulcorner \pi(\dot{z})^\urcorner)$$
$$\exists y, z(x = \neg\dot{\,}(y\dot{\wedge}z) \wedge \mathrm{Tr}^\ulcorner \pi(\neg\dot{y})^\urcorner \wedge \mathrm{Tr}^\ulcorner \pi(\neg\dot{z})^\urcorner)$$
$$\exists y(x = \forall\dot{v}y \wedge \forall t\, \mathrm{Tr}^\ulcorner \pi(\dot{y}(t/v))^\urcorner) \vee$$
$$\exists y(x = \neg\dot{\,}\forall\dot{v}y \wedge \forall t\, \mathrm{Tr}^\ulcorner \pi(\neg\dot{y}(t/v))^\urcorner).$$

Thus, the formula $\pi(x)$ performs a standard iteration for truth ascriptions, but exploits the diagonal lemma to yield Strong Kleene compositionality in the case of logically complex formulae – in much the same way that the well-known definition of KF over the theory PUTB does (see, e.g., [Hal14, Ch.19]).

The following result shows that, in a certain class of models, $\pi(x)$ captures Strong-Kleene satisfaction:

Lemma 10. *Any $S \subseteq \mathrm{Sent}_{\mathcal{L}_{\mathrm{Tr}}}$ with $(\mathbb{N}, S) \vDash \forall x(\mathrm{Tr}^\ulcorner \pi(\dot{x})^\urcorner \leftrightarrow \pi(x))$ is such that*

$$S \vDash \pi(^\ulcorner \varphi^\urcorner) \text{ iff } (\mathbb{N}, S, S') \vDash_{\mathrm{sk}} \varphi$$

Proof. One needs to induct on the positive complexity of φ. Suppose that φ is an arithmetic literal. Then

$$S \vDash \pi(^\ulcorner \varphi^\urcorner) \text{ iff } \#\varphi \in \mathrm{True}_0 \text{ iff } \mathbb{N} \vDash \varphi \text{ iff } (\mathbb{N}, S, S^-) \vDash_{\mathrm{sk}} \varphi.$$

The case for φ of the form $\mathrm{Tr}t$ is also straightforward:

$$S \vDash \pi(^\ulcorner \mathrm{Tr}t^\urcorner) \text{ iff } t^\circ \in S \text{ iff } (\mathbb{N}, S, S^-) \vDash_{\mathrm{sk}} \mathrm{Tr}t.$$

When φ is of the form $\neg\mathrm{Tr}t$, the argument is analogous. If φ is a logically complex formula, the claim follows almost immediately by IH. Take the case of φ being $\neg\neg\psi$:

$$S \vDash \pi(^\ulcorner \neg\neg\psi^\urcorner) \Leftrightarrow S \vDash \mathrm{Tr}^\ulcorner \pi(^\ulcorner \psi^\urcorner)^\urcorner \Leftrightarrow S \vDash \pi(^\ulcorner \psi^\urcorner) \Leftrightarrow (\mathbb{N}, S, S^-) \vDash_{\mathrm{sk}} \psi \Leftrightarrow (\mathbb{N}, S, S^-) \vDash_{\mathrm{sk}} \neg\neg\psi$$

where the third biconditional follows evidently by IH. $\square$

For future use, we recall that, for Tr-positive formulae, Strong-Kleene satisfaction and classical satisfaction coincide.

Lemma 11. *For φ Tr-positive: $(\mathbb{N}, S, S^-) \vDash_{\mathrm{sk}} \varphi$ iff $S \vDash \varphi$.*

Proof. We induct on the positive complexity of φ again. If φ is an arithmetic literal or negated literal, the claim is clear. If φ is of the form $\mathrm{Tr}t$, then $(\mathbb{N}, S, S^-) \vDash_{\mathrm{sk}} \mathrm{Tr}t$ iff $t^\circ \in S$ iff $S \vDash \mathrm{Tr}t$. The case in which φ is of the form $\neg\mathrm{Tr}t$ does not arise, since φ would not be Tr-positive. The inductive case follows immediately from the IH, since the SK satisfaction clauses for all connectives (except negation) and classical satisfaction coincide. $\square$

Definition 12. *Let* PK *be the theory*

(TBπ) $\forall x(\mathrm{Tr}^\ulcorner \pi(\dot{x})^\urcorner \leftrightarrow \pi(x))$

(PK-Tr) $\mathrm{Tr}(x) \leftrightarrow \exists y(\pi(y) \wedge \mathrm{Pr}_{\mathrm{PAT}}(y\dot{\to}x)).$

As the following lemma shows, the theory PK has similar compositional properties as the theory IT with the exception that the truth predicate does not commute with the universal quantifier.

Lemma 13. *The theory* PK *proves the following sentences:*

(1) $\forall x(\mathrm{Tr}(\underline{\mathrm{T}}\mathrm{r}(x)) \leftrightarrow \mathrm{Tr}(x))$;

(2) $\forall x(\mathrm{Tr}(\dot{\neg}\underline{\mathrm{T}}\mathrm{r}(x)) \leftrightarrow \mathrm{Tr}(\dot{\neg}x) \vee \neg\mathrm{Sent}(x))$;

(3) $\forall x(\mathrm{Sent}(x) \rightarrow (\mathrm{Tr}(\dot{\neg}\dot{\neg}x) \leftrightarrow \mathrm{Tr}(x)))$;

(4) $\forall x, y(\mathrm{Sent}(x\dot{\wedge}y) \rightarrow (\mathrm{Tr}(x\dot{\wedge}y) \leftrightarrow \mathrm{Tr}(x) \wedge \mathrm{Tr}(y)))$;

(5) $\forall x, y(\mathrm{Sent}(x\dot{\vee}y) \rightarrow (\mathrm{Tr}(x) \vee \mathrm{Tr}(y) \vee \exists z(\pi(z) \wedge \mathrm{Pr}_{\mathrm{PAT}}(z\dot{\rightarrow}x\dot{\vee}y)) \leftrightarrow \mathrm{Tr}(x\dot{\vee}y)))$;

(6) $\forall x, v(\mathrm{Sent}(\dot{\forall}vx) \rightarrow (\mathrm{Tr}(\dot{\forall}vx) \rightarrow \forall t\mathrm{Tr}(x(t/v))))$;

(7) $\forall x, y(\mathrm{Sent}(\dot{\exists}vx) \rightarrow (\exists t\mathrm{Tr}(x\,(t/v)) \vee \exists w(\pi(w) \wedge \mathrm{Pr}_{\mathrm{PAT}}(w\dot{\rightarrow}\dot{\exists}vx)) \leftrightarrow \mathrm{Tr}(\dot{\exists}vx)))$.

Now, we establish the $\mathbb{N}$-categoricity results for SSK fixed points:

Theorem 14. *For* $S \subseteq \mathrm{Sent}_{\mathcal{L}_{\mathrm{Tr}}}$, *and* $\mathrm{CON}(S)$, $\mathrm{SSK}(S) = S$ *iff* $(\mathbb{N}, S) \models \mathrm{PK}$.

We first prove (almost) one direction as a lemma:

Lemma 15. *Let* $S \subseteq \mathrm{Sent}_{\mathcal{L}_{\mathrm{Tr}}}$ *be such that* $\mathrm{CON}(S)$ *and* $S = \mathrm{SSK}(S)$. *Then,* $S \models \mathrm{TB}\pi$.

Proof. We check that $S \models \mathrm{Tr}\ulcorner \pi(\overline{n})\urcorner \leftrightarrow \pi(\overline{n})$ for all $n \in \omega$. The left-to-right direction follows for any n directly from the fact that SSK satisfies classical soundness, that is, one can immediately verify that $S \models_{\mathrm{ssk}} \pi(\overline{n}) \Rightarrow S \models \pi(\overline{n})$ (see also [Ste18, Lemma 3.7]). For the right-to-left direction, note that, by lemma 11, $S \models \pi(\overline{n})$ implies $S \models_{\mathrm{sk}} \pi(\overline{n})$, and so $S \models_{\mathrm{ssk}} \pi(\overline{n})$; so, since $S = \mathrm{SSK}(S)$, $\#\pi(\overline{n}) \in S$, and $S \models \mathrm{Tr}\ulcorner \pi(\overline{n})\urcorner$. $\qquad\square$

Proof of Theorem 14. For the left-to-right direction: lemma 15 shows that $S \models \mathrm{TB}\pi$. Now, we want to prove the axiom (PK-Tr). Assume $S \models \mathrm{Tr}\ulcorner \varphi \urcorner$. Then, since $S = \mathrm{SSK}(S)$, there is some ψ such that $S \models_{\mathrm{sk}} \psi$ & $\mathrm{PAT} \vdash \psi \rightarrow \varphi$. By lemma 10, and since $(\mathbb{N}, S)$ is a model of $\mathrm{TB}\pi$, $S \models_{\mathrm{sk}} \psi$ implies $S \models \pi(\ulcorner \psi \urcorner)$. Further, since $(\mathbb{N}, S)$ is a model of PAT, $S \models \exists y(\pi(y) \wedge \mathrm{Pr}_{\mathrm{PAT}}(y\dot{\rightarrow}\ulcorner \varphi \urcorner))$. The reverse reasoning yields the opposite direction of the axiom.

For the right-to-left direction: we reason as follows:

$$
\begin{aligned}
\#\varphi \in S &\Leftrightarrow S \models \mathrm{Tr}\ulcorner \varphi \urcorner \\
&\Leftrightarrow S \models \exists \ulcorner \psi \urcorner(\pi(\ulcorner \psi \urcorner \wedge \mathrm{Pr}_{\mathrm{PAT}}(\ulcorner \psi \urcorner \rightarrow \varphi \urcorner)) && \text{by (PK-Tr)} \\
&\Leftrightarrow S \models \pi(\ulcorner \chi \urcorner) \text{ and } S \models \mathrm{Pr}_{\mathrm{PAT}}(\ulcorner \chi \rightarrow \varphi \urcorner) && \text{for some } \chi \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{Tr}}} \\
&\Leftrightarrow S \models_{\mathrm{sk}} \chi \text{ and } \mathrm{PAT} \vdash \chi \rightarrow \varphi && \text{by lemma 10} \\
&\Leftrightarrow \#\varphi \in \mathrm{SSK}(S)
\end{aligned}
$$

$\qquad\square$

In a sense, PK amounts to the bare minimum required to establish its $\mathbb{N}$-categoricity with respect to SSK. We can extend the theory with principles that are independently justified and sound with respect to the intended semantics, and yet retain $\mathbb{N}$-categoricity. Principles that appear particularly natural are, on the one hand, a less artificial restriction of the naive Tr-schema allowing disquotation for all Tr-positive sentences; on the other, since the SSK-fixed points over $\mathbb{N}$ are classically sound, it is plausible to reflect this in the axiomatization by adding the schema Tr-Out to it.

Definition 16. *We call* PK^+ *the theory obtained by replacing, in* PK, *$\mathrm{TB}\pi$ with the schema*

(PUTB) $\qquad\qquad\qquad\qquad \mathrm{Tr}\ulcorner \varphi(\vec{t})\urcorner \leftrightarrow \varphi(\vec{t}^{\,\circ})$ *for φ Tr-positive*

and adding the schema Tr-Out *to it.*

PK^+ is also $\mathbb{N}$-categorical with respect to SSK. The argument above carries over with minor modifications: Lemma 15 still holds if one requires satisfiability of PUTB instead of just $TB\pi$. This is enough, in the proof of Theorem 13, to show that fixed points of SSK indeed satisfy PUTB.

Since PK^+ includes PUTB, it is proof-theoretically at least as strong as KF or ramified truth up to ε_0. IT (see Def. 2) is clearly an upper bound for PK^+, but it's evidently not sharp as the former has the same proof-theoretic strength of the theory of inductive definitions ID_1 [Ste18, Cor. 6.7]. It is an open question whether PK^+ has precisely the strength of PUTB. In addition, a full proof-theoretic analysis of PK is lacking.

6. The $\mathbb{N}$-categorical axiomatization of Θ, Θ^*, and variations

6.1. $\mathbb{N}$-categorical axioms for Θ.
By Proposition 7, IT isn't $\mathbb{N}$-categorical with respect to the Θ fixed points, but are there other axiomatic theories which are $\mathbb{N}$-categorical with respect to those fixed points? In this subsection, we answer that question in the affirmative.

Definition 17 (IT^-). *The theory* IT^- *extends* PAT *with the following axioms:*

IT1 $\forall s, t(\mathrm{Tr}(s \dot{=} t) \leftrightarrow s^\circ = t^\circ) \wedge \forall s, t(\mathrm{Tr}(s \dot{\neq} t) \leftrightarrow s^\circ \neq t^\circ)$

IT2 $\forall x, y(\mathrm{Sent}(x \dot{\wedge} y) \rightarrow (\mathrm{Tr}(x) \wedge \mathrm{Tr}(y) \rightarrow \mathrm{Tr}(x \dot{\wedge} y)))$

IT3 $\forall x, y(\mathrm{Sent}(x \dot{\vee} y) \rightarrow (\mathrm{Tr}(x) \vee \mathrm{Tr}(y) \vee \exists z(\xi(z, \mathrm{Tr}) \wedge \mathrm{Pr}_{PAT}(z \dot{\rightarrow} x \dot{\vee} y)) \leftrightarrow \mathrm{Tr}(x \dot{\vee} y)))$

IT4 $\forall v, x(\mathrm{Sent}(\dot{\forall} v x) \rightarrow (\forall t \mathrm{Tr}(x\,(\dot{t}/v)) \rightarrow \mathrm{Tr}(\dot{\forall} v x)))$

IT5 $\forall x, y(\mathrm{Sent}(\dot{\exists} v x) \rightarrow (\exists t \mathrm{Tr}(x\,(t/v))) \vee \exists w(\xi(w, \mathrm{Tr}) \wedge \mathrm{Pr}_{PAT}(w \dot{\rightarrow} \dot{\exists} v x)) \leftrightarrow \mathrm{Tr}(\dot{\exists} v x)))$

IT^-6 $\forall x(\mathrm{Tr}(x) \rightarrow \mathrm{Tr}(\dot{\mathrm{Tr}} x))$

IT^-7 $\forall t(\mathrm{Tr}(\dot{\neg} t) \vee \neg \mathrm{Sent}(t) \rightarrow \mathrm{Tr}(\dot{\neg} \dot{\mathrm{Tr}} t))$

IT8 $\forall x, y(\mathrm{Tr}(x) \wedge \mathrm{Pr}_{PAT}(x \dot{\rightarrow} y) \rightarrow \mathrm{Tr}(y))$

IT9 $\forall x(\mathrm{Tr}(\dot{\neg} x) \rightarrow \neg \mathrm{Tr}(x))$

IT10 $\mathrm{Tr}\ulcorner \forall x(\mathrm{Tr}(x) \rightarrow \mathrm{Sent}(x))\urcorner$

Thus, the major differences between IT^- and IT are simply the exclusion of T-Out and of the right-to-left directions of axioms IT6 and IT7. Note that Axiom IT9 is provable from axioms IT1, IT2 and IT8.

We shall prove:

Theorem 18. IT^- *is* $\mathbb{N}$-*categorical with respect to fixed points of* Θ, *i.e.,* $S = \Theta(S) \Leftrightarrow S \vDash IT^-$.

Proof. For the left-to-right direction: for axioms IT1-IT^-7 the proof proceeds just like [Ste18, Lemma 5.6]. IT8 follows from the fact that fixed points of Θ are closed under PAT-provability. As remarked, IT9 follows from IT1, IT2 and IT8. IT10 also follows from the definition of the Θ jump.

For the right-to-left direction: we need to show that $S = \Theta(S)$, provided $S \vDash IT^-$. But since we know that Θ is sound for any choice of S, what remains to be shown is $\Theta(S) \subseteq S$. Assume $\#\varphi \in \Theta(S)$. Then,

$$\mathbb{N} \vDash \exists \ulcorner \psi \urcorner (\xi(\ulcorner \psi \urcorner, S) \wedge \mathrm{Pr}_{PAT}(\ulcorner \psi \rightarrow \varphi \urcorner)).$$

We first show $\xi(\ulcorner \psi \urcorner, S)$ implies $\#\psi \in S$, which follows easily by an induction on the positive complexity of φ. If φ is atomic or negated atomic, we make use of IT1; if it is of the form $\mathrm{Tr}t^\circ$ or $\neg \mathrm{Tr}t^\circ$, we make use of IT^-6 and IT^-7, respectively; if it is a conjunction, we make use of IT2; and so on. Once we have shown that $\xi(\ulcorner \psi \urcorner, S) \Rightarrow \#\psi \in S$, $\#\varphi \in S$ follows by using IT8. $\square$

6.2. Θ^* and $\mathbb{N}$-categorical axioms for it.
Fixed points of Θ are not very attractive though. The reason is that they are not necessarily *Kripke-sets* in the sense that $\#\varphi \in X \Leftrightarrow \#\mathrm{Tr}\ulcorner \varphi \urcorner \in X$ and $\#\neg \varphi \in X \Leftrightarrow \#\neg \mathrm{Tr}\ulcorner \varphi \urcorner \in X$. In particular, as we showed in Proposition 4, some fixed points will contain codes of sentences of the form $\mathrm{Tr}\ulcorner \varphi \urcorner$, while not containing the code of φ – and similarly for $\neg \mathrm{Tr}\ulcorner \varphi \urcorner$ and $\neg \varphi$.

This phenomenon is reflected in the absence of the right-to-left directions of IT^-6 and IT^-7 in IT^-. However, by tweaking the formula $\xi(x, X)$ slightly, we can obtain a class of fixed points that are ($\mathbb{N}$-categorically) axiomatized by a theory which restores the missing components of the original IT6 and IT7:

Definition 19. *Let $\xi^*(x, X)$ be the formula obtained by supplementing $\xi(x, X)$ from Definition 1 with additional Tr-downwards closure properties:*

$$x \in \text{True}_0$$
$$\vee \, \exists y, z(x = (y \dot{\vee} z) \wedge (y \in X \vee z \in X))$$
$$\vee \, \exists y, z(x = (y \dot{\wedge} z) \wedge (y \in X \wedge z \in X))$$
$$\vee \, \exists y, v(x = (\dot{\forall} vy) \wedge \forall t(y(t/v) \in X))$$
$$\vee \, \exists y, v(x = (\dot{\exists} vy) \wedge \exists t(y(t/v) \in X))$$
$$\vee \, \exists t(x = \dot{\text{Tr}}(t) \wedge t^\circ \in X)$$
$$\vee \, \exists t(x = \dot{\neg}\dot{\text{Tr}}(t) \wedge (\dot{\neg}t^\circ) \in X \vee \neg\text{Sent}(t^\circ))$$
$$\vee \, \exists t(x = t^\circ \wedge \text{Sent}(t^\circ) \wedge \dot{\text{Tr}}(t) \in X)$$
$$\vee \, \exists t(x = \dot{\neg}t^\circ \wedge \text{Sent}(\dot{\neg}t^\circ) \wedge \dot{\neg}\dot{\text{Tr}}(t) \in X)$$

We then define:

$$\Theta^*(S) := \{n \in \omega \mid \mathbb{N} \models \exists y(\xi^*(y, X) \wedge \text{Pr}_{\text{PAT}}(y \dot{\to} x))[n, S]\}$$

Note that, for any set S, S is sound with respect to Θ^* – just like with Θ. However, it follows immediately from the definition of Θ^* that Θ^* fixed points are Kripke sets:

Proposition 20. *Let $X = \Theta^*(X)$. Then, $\#\varphi \in X \Leftrightarrow \#\text{Tr}\ulcorner\varphi\urcorner \in X$ and $\#\neg\varphi \in X \Leftrightarrow \#\neg\text{Tr}\ulcorner\varphi\urcorner \in X$.*

It is now easy to see what theory characterizes Θ^* fixed points $\mathbb{N}$-categorically:

Theorem 21. *Let IT^* be IT minus the axiom schema of T-Out, and replacing $\xi(x, X)$ in axioms IT3 and IT6 for $\xi^*(x, X)$. Then, IT^* is $\mathbb{N}$-categorical with respect to fixed points of Θ^*, i.e., $S = \Theta^*(S) \Leftrightarrow S \models \text{IT}^*$.*

Θ^* does not impose any requirement on what the truth predicate looks like *internally*. In a way analogous to that of the different supervaluation schemes, however, we can also require that, internally, the truth predicate be consistent, or consistent and complete. This is done by defining new operators:

Definition 22. *Let $\text{con}(x)$ be the formula $\exists s, t(x = \ulcorner\neg(\text{Tr}\,\underline{s} \wedge \text{Tr}\dot{\neg}\underline{t})\urcorner \wedge s^\circ = t^\circ)$, and define:*

$$\Theta_c^*(S) := \{n \in \omega \mid \mathbb{N} \models \exists y((\xi^*(y, X) \vee \text{con}(y)) \wedge \text{Pr}_{\text{PAT}}(y \dot{\to} x))[n, S]\}$$

Definition 23. *Let $\text{com}(x)$ be the formula $\exists s, t(x = \ulcorner\neg\text{Tr}\,\underline{s} \leftrightarrow \text{Tr}\dot{\neg}\underline{t}\urcorner \wedge s^\circ = t^\circ)$, and define:*

$$\Theta_{\text{mc}}^*(S) := \{n \in \omega \mid \mathbb{N} \models \exists y((\xi^*(y, X) \vee \text{com}(y)) \wedge \text{Pr}_{\text{PAT}}(y \dot{\to} x))[n, S]\}$$

In light of our $\mathbb{N}$-categoricity result for Θ^*, the following is easily established:

Proposition 24. *Let IT_c^* be IT^* plus the axiom $\forall x(\text{Tr}\ulcorner\neg(\text{Tr}\dot{x} \wedge \text{Tr}\dot{\neg}\dot{x})\urcorner)$, and let IT_{mc}^* be IT^* plus the axiom $\forall x(\text{Tr}\ulcorner\neg\text{Tr}\dot{x} \leftrightarrow \text{Tr}\dot{\neg}\dot{x}\urcorner)$. Then, IT_c^* is $\mathbb{N}$-categorical with respect to fixed points of Θ_c^*, and IT_{mc}^* is $\mathbb{N}$-categorical with respect to fixed points of Θ_{mc}^*.*

6.3. **Relationships between fixed points of Θ and Θ^*.** We start by providing a general result about Θ and Θ^*, which reveals that their constructions preserve classical soundness and, in the case of Θ, also Tr- and ¬Tr-downwards closure:

Lemma 25. *The following hold:*

(1) *Let X be a classically sound set of sentences, and define:*

$$\Delta_0 := X$$
$$\Delta_{\alpha+1} := \Theta^*(\Delta_\alpha)$$
$$\Delta_\lambda := \bigcup_{\beta<\lambda} \Delta_\beta$$

Then, $\bigcup_{\alpha\in\text{On}} \Delta_\alpha$ – the least Θ^* fixed point containing X – is classically sound.

(2) *Let X be a classically sound, Tr- and ¬Tr-downwards closed set of sentences, and define:*

$$\Gamma_0 := X$$
$$\Gamma_{\alpha+1} := \Theta(\Gamma_\alpha)$$
$$\Gamma_\lambda := \bigcup_{\beta<\lambda} \Gamma_\beta$$

Then, $\bigcup_{\alpha\in\text{On}} \Gamma_\alpha$ is classically sound, and Tr- and ¬Tr-downwards closed.

Proof. For (1): we show that $\#\varphi \in \Delta_\alpha$ implies $\Delta_\alpha \vDash \varphi$. The base case follows from the assumption that X is classically sound. For the successor case: we can assume that $\#\varphi \in \Delta_\alpha$ implies $\Delta_\alpha \vDash \varphi$ and prove the claim for $\alpha + 1$. Suppose that $\#\varphi \in \Delta_{\alpha+1}$ by $\xi^*(x, X)$. We induct on the positive complexity of φ:

- If φ is in True$_0$, this is immediate.
- If φ is of the form Tr$\ulcorner\psi\urcorner$, and $\#\psi \in \Delta_\alpha$: since $\Delta_\alpha \subseteq \Delta_{\alpha+1}$, $\psi \in \Delta_{\alpha+1}$, hence $\Delta_{\alpha+1} \vDash \text{Tr}\ulcorner\psi\urcorner$.
- If φ is of the form ¬Tr$\ulcorner\psi\urcorner$, and $\#\neg\psi \in \Delta_\alpha$: then $\#\neg\psi \in \Delta_{\alpha+1}$ and, since $\Delta_{\alpha+1}$ is consistent, $\#\psi \notin \Delta_{\alpha+1}$, hence $\Delta_{\alpha+1} \vDash \neg\text{Tr}\ulcorner\psi\urcorner$.
- If φ is a logically complex formula, the claim follows easily by IH.

If $\#\varphi \in \Delta_{\alpha+1}$ because there is ψ such that $\xi^*(\ulcorner\psi\urcorner, \Delta_\alpha)$ holds and PAT $\vdash \psi \to \varphi$: since $\Delta_{\alpha+1} \vDash$ PAT, and we just showed how $\xi^*(\ulcorner\psi\urcorner, \Delta_\alpha)$ implies $\Delta_{\alpha+1} \vDash \psi$, the claim follows.

For (2): to prove that $\bigcup_{\alpha\in\text{On}} \Gamma_\alpha$ is classically sound, we only need to run again the proof for (1) again. To prove the Tr- and ¬Tr-downwards closure, we need to verify that $\#(\neg)\text{Tr}\ulcorner\varphi\urcorner \in \Gamma_\alpha \Rightarrow \#(\neg)\varphi \in \Gamma_\alpha$. But this is immediate for the base case from our assumption that X is Tr and ¬Tr-downwards closed, and follows by IH for the successor case. The argument just given implies that, for any α, $\xi^*(\ulcorner\varphi\urcorner, \Gamma_\alpha) \Leftrightarrow \xi(\ulcorner\varphi\urcorner, \Gamma_\alpha)$, whence we can conclude that $\bigcup_{\alpha\in\text{On}} \Gamma_\alpha$ is also the least fixed point of Θ^* containing X. □

In the constructions above, we say that X is the *starting set* of the construction.

Next, we show that fixed points of Θ^* are also fixed points of Θ. This is immediate by inspecting the definitions of Θ and Θ^*:

Lemma 26. *Let $S = \Theta^*(S)$. Then, $S = \Theta(S)$. Conversely, if $S = \Theta(S)$ and S is Tr- and ¬Tr-downwards closed, then $S = \Theta^*(S)$.*

By Proposition 4, there are fixed points of Θ that are not Tr- and ¬Tr-downwards closed. However, Tr and ¬Tr-downwards closure are still not sufficient to yield the ℕ-categoricity of IT. There could still be complex sentences, such as $\tau_0 \vee \tau_1$, within Θ^* fixed points, without their constituent sentences being there—which would violate Tr-Out. However, if classical soundness is added as a condition on the fixed points, we obtain a more restricted class of fixed points that IT can ℕ-categorically axiomatize.

Proposition 27. *Let $S \subseteq \mathrm{Sent}_{\mathcal{L}_{\mathrm{Tr}}}$. Then the following claims are equivalent:*

(1) $S = \Theta^*(S)$ & S is classically sound;
(2) $S = \Theta(S)$ & S is classically sound and Tr- and $\neg\mathrm{Tr}$-downwards closed;
(3) $S \vDash \mathrm{IT}$.
(4) $S \vDash \mathrm{IT}^* + \mathrm{Tr}\text{-Out}$

Proof. $1 \Leftrightarrow 4$: In Theorem 21, we proved that $S = \Theta^*(S)$ iff $S \vDash \mathrm{IT}^*$. But if S is classically sound, then $S \vDash \mathrm{Tr}\ulcorner\psi\urcorner \to \psi$ for all ψ s.t. $\psi \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{Tr}}}$. The reverse also holds: the fixed-point model $(\mathbb{N}, S)$ satisfies T-Out only if S is classically sound.

$2 \Leftrightarrow 3$: IT is IT^- plus Tr-Out and the right-to-left directions of axioms IT6 and IT7. By Theorem 18, $S = \Theta(S)$ iff $S \vDash \mathrm{IT}^-$. Tr-Out holds iff S is classically sound; and the right-to-left directions of IT6 and IT7 hold exactly when S is Tr- and $\neg\mathrm{Tr}$-downwards closed.

$1 \Leftrightarrow 2$: given by Lemma 26. $\qquad\square$

6.4. **Proof-Theory.** IT^* also enjoys considerable proof-theoretic strength. In particular, we can show that it is at least as proof-theoretically strong as the theory KF. The proof-theoretic strength will be obtained by defining ramified truth predicates, so we refer the reader to [Hal14, Ch.9] for a definition of the relevant languages and systems. Instead of using IT^*, however, we will provide a more general result by using IT^-; since the latter is a subtheory of IT^*, the claim will follow immediately.

Proposition 28. IT^- *can define the truth predicates of* $\mathrm{RT}_{<\varepsilon_0}$.

Proof. Consider the formula

$$\mathrm{D}(x) := (\mathrm{Tr}x \vee \mathrm{Tr}\dot{\neg}x) \wedge \forall y, z(x = (y\dot{\vee}z) \to \mathrm{Tr}y \vee \mathrm{Tr}z) \wedge \forall y, v(x = (\dot{\exists}vy) \to \exists t\mathrm{Tr}y(t/v))$$

It is immediate to verify that $\mathrm{D}(x)$ satisfies the following properties:

(1) $\mathrm{D}(x) \to \mathrm{D}(\dot{\neg}x)$
(2) $\mathrm{D}(x) \wedge \mathrm{D}(y) \to \mathrm{D}(x\dot{\vee}y) \wedge \mathrm{D}(x\dot{\to}y)$
(3) $\mathrm{Sent}_{\mathcal{L}_{\mathrm{Tr}}}(\dot{\forall}vy) \wedge \forall t(\mathrm{D}(y(t/v))) \to \mathrm{D}(\dot{\forall}vy)$
(4) $\mathrm{D}(x) \to (\mathrm{Tr}(\dot{\neg}x) \leftrightarrow \neg\mathrm{Tr}(x))$
(5) $\mathrm{D}(x\dot{\vee}y) \to (\mathrm{Tr}(x\dot{\vee}y) \leftrightarrow \mathrm{Tr}(x) \vee \mathrm{Tr}(y))$
(6) $\mathrm{D}(x\dot{\to}y) \to (\mathrm{Tr}(x\dot{\to}y) \leftrightarrow \mathrm{Tr}(x) \to \mathrm{Tr}(y))$
(7) $\mathrm{D}(x) \to (\mathrm{Tr}(\dot{\forall}vx) \leftrightarrow \forall y\mathrm{Tr}(x(y/v)))$
(8) $\mathrm{D}(x) \to (\mathrm{Tr}(x) \leftrightarrow \mathrm{Tr}(\dot{\mathrm{Tr}}x))$

We now use this formula to adapt a proof by Fujimoto [Fuj10, Lemma 36]. We start by defining two p.r. functions *à la* Fujimoto. First, for $\beta < \varepsilon_0$, let $h(x, \beta)$ be a p.r. function from $\omega \times \mathrm{On}$ into ω defined by

$$h(x, \beta) = \begin{cases} x & \text{if } x \in \mathrm{Sent}_{<\beta} \\ \ulcorner 0 = 1\urcorner & \text{otherwise} \end{cases}$$

Second, we define a function $k(x)$ from $\mathcal{L}_{<\varepsilon_0}$ into $\mathcal{L}_{\mathrm{Tr}}$ by

$$k(x) = \begin{cases} x, & \text{if } x \text{ is an atomic formula of } \mathcal{L}_{\mathbb{N}} \\ \ulcorner\mathrm{Tr}(k \circ h(t, \beta))\urcorner & \text{if } x = \mathrm{Tr}_\beta t \text{ and } t \text{ is a closed term} \\ \ulcorner\neg\mathrm{Tr}(k(\dot{y}))\urcorner & \text{if } x = \dot{\neg}y \\ \ulcorner\mathrm{Tr}(k(\dot{y})) \vee \mathrm{Tr}(k(\dot{z}))\urcorner & \text{if } x = y\dot{\vee}z \\ \ulcorner\mathrm{Tr}k(\dot{y}) \wedge \mathrm{Tr}(k(\dot{z}))\urcorner & \text{if } x = y\dot{\wedge}z \\ \ulcorner\forall z\mathrm{Tr}(k(\dot{y}(u/\ulcorner z\urcorner)))\urcorner(z/\ulcorner u\urcorner) & \text{if } x = \dot{\forall}zy \text{ and } z \text{ is a variable} \\ \ulcorner 0 = 1\urcorner & \text{otherwise} \end{cases}$$

Using the axioms of IT^- and properties 1-8, one can show:

$$IT^- \vdash \forall \gamma \leq \beta(\text{Sent}_{<\gamma}(x) \to D(kx))$$

Finally, for any ordinal $\alpha < \varepsilon_0$, define $\sigma_\alpha : \mathcal{L}_{<\alpha} \longrightarrow \mathcal{L}_{\text{Tr}}$ as:

$$\sigma_\alpha(\varphi) = \begin{cases} s = t & \text{if } \varphi = (s = t) \\ \text{Tr}(kx) & \text{if } \varphi = \text{Tr}_\beta(x), \beta < \alpha \\ \neg\sigma_\alpha(\psi) & \text{if } \varphi = \neg\psi \\ \sigma_\alpha(\psi) \vee \sigma_\alpha(\chi) & \text{if } \varphi = \psi \vee \chi \\ \sigma_\alpha(\psi) \wedge \sigma_\alpha(\chi) & \text{if } \varphi = \psi \wedge \chi \\ \exists v \sigma_\alpha(\psi) & \text{if } \varphi = \exists v \psi \\ \forall v \sigma_\alpha(\psi) & \text{if } \varphi = \forall v \psi \end{cases}$$

It is easy to verify that translating $\text{Tr}_\alpha(x)$ as $\sigma_\alpha(x)$, for $\alpha < \varepsilon_0$, yields a truth-definition of $RT_{<\varepsilon_0}$ in IT^-. $\qquad \square$

Corollary 29. $|IT^*| \geqq |IT^-| \geqq |KF| \equiv |RT_{<\varepsilon_0}|$

Just like in the case of PK, IT^*, IT^- are sub-theories of IT. However, a sharp proof-theoretic analysis of IT^*, IT^- is not yet available.

By contrast, the proof-theoretic analysis of $IT^* + \text{Tr-Out}$ follows the one for IT from [Ste18]: the lower bound is provided by the provability of Bar-Induction in (a fragment of) $IT^* + \text{Tr-Out}$, which is known to be sufficient to derive the arithmetical theorems of ID_1; for the upper bound argument, the strategy is to formalize in the theory KPU the stages of the construction of the minimal fixed point of Θ, and establish the equivalence between membership in the fixed point and provability in the infinitary system defined in Proposition 9. Since, provably in KPU, $I_\Theta = I_{\Theta^*}$, we can use such a provability predicate as an adequate interpretation for the truth predicate of $IT^* + \text{Tr-Out}$. That is:

Proposition 30. $|IT^* + \text{Tr-Out}| = |IT| = |ID_1|$.

7. Relations with standard supervaluationist fixed points

In this section we establish some connections between supervaluation-style fixed-point semantics and the standard supervaluational approaches. The aim is to draw a precise picture of the prospects and limitations of supervaluation-style semantics and its connection to standard supervaluational semantics.

Recall the schema VB introduced in Section 2.2. It is easy to see that the minimal fixed points of Θ, Θ^* and VB coincide:

Corollary 31. $I_\Theta = I_{\Theta^*} = I_{vb}$.

Proof. The proof of Proposition 9 shows that $\#\varphi \in I_\Theta \Leftrightarrow SV_\infty \vdash \varphi$. But Cantini [Can90, §4, esp. Thm. 4.7] proved that $\#\varphi \in I_{vb} \Leftrightarrow SV_\infty \vdash \varphi$. It then follows that $\#\varphi \in I_\Theta \Leftrightarrow \#\varphi \in I_{vb}$.

We now show that $\#\varphi \in I_\Theta \Leftrightarrow \#\varphi \in I_{\Theta^*}$. We know that VB fixed points are Tr- and ¬Tr-downwards closed (Def. 8); combining this with the fact that I_Θ is a fixed point of Θ, it follows that it is also a fixed point of Θ^*. But if I_Θ were not the minimal fixed point of Θ^*, there would be a set $S = \Theta^*(S)$ such that $S \subsetneq I_\Theta$. Also, by Lemma 26, $S = \Theta(S)$. But this contradicts the fact that I_Θ is the Θ-least fixed point. $\qquad \square$

By contrast, it can be shown that the minimal fixed point is a proper subset of the least VB fixed point—and, as a consequence, of the least fixed points of Θ and Θ^*:

Proposition 32. $I_{ssk} \subsetneq I_{vb}$.

Proof. That $I_{ssk} \subseteq I_{vb}$ was already proven in [Ste18, Lemma 3.6]: indeed, the implication

$$\varphi \in \text{SSK}(S) \Rightarrow \forall S' \supseteq S(\Phi_{vb}(S') \Rightarrow S' \vDash \varphi)$$

holds because, as we have seen, (i) satisfaction in Strong Kleene entails VB-satisfaction, and (ii) VB-truth is closed under PAT-consequence. That the inclusion is proper follows from Proposition 6, where we showed how I_{ssk} fails to be closed under ω-logic. By contrast, as is well-known, VB fixed points (including I_{vb}) are closed under ω-logic. $\qquad \square$

It is clear that some fixed points of Θ are not VB fixed points. For instance, some fixed points of Θ fail to be transparent for the truth predicate (Proposition 4). It is also clear that some Θ^* fixed points are not VB fixed points: all fixed-point models of VB satisfy T-Out, but not all Θ^* fixed points do.[7] What if we restrict our attention to Θ^* fixed-point models of T-Out, i.e., that are classically sound? In Proposition 27, we showed that those models are precisely the ones that satisfy IT. It then follows that the class of classically sound fixed points of Θ^* cannot coincide with the class of supervaluational, VB fixed points. Otherwise, VB would have an $\mathbb{N}$-categorical axiomatization (namely, IT), which we know isn't the case by the complexity considerations described in the introduction.

There is more we can say about fixed points of VB and of Θ^*. First, that the class of VB fixed points is a subclass of the Θ^* fixed points:

Lemma 33. *Let $S = \text{VB}(S)$. Then, $S = \Theta^*(S)$, and $S = \Theta(S)$, and S is classically sound.*

Proof. Assume $S = \text{VB}(S)$. We want to show $S = \Theta^*(S)$. $S \subseteq \Theta^*(S)$ follows immediately, since we know that every set of formulae is Θ^*- sound. For the other direction, we induct on the clauses of $\xi^*(x, X)$, and employ the VB fixed-point property of S in order to prove that $\xi^*(x, S) \Rightarrow x \in S$. It is easy to see why this holds: any fixed point of VB is closed under ω-logic (and so it is closed under the clauses for the connectives and quantifiers in ξ^*), and is closed under Tr-Intro, Tr-Elim, $\neg$Tr-Intro and $\neg$Tr-Elim, and so closed under the Tr-clauses. Furthermore, VB fixed points are also closed under PAT-provability, so $x \in \Theta^*(S) \Rightarrow x \in S$.

Having shown $S = \Theta^*(S)$, that $S = \Theta(S)$ just follows from Lemma 26. Also, that S is classically sound follows immediately, since S is a VB-fixed point. $\qquad \square$

In fact, we can show that any least fixed point generated by a VB-sound set corresponds to the least fixed point of Θ generated from that set. To prove that, we need two observations.

Lemma 34. *Let $S \subseteq \text{Sent}_{\mathcal{L}_{\text{Tr}}}$ and $S \subseteq \text{VB}(S)$. Then, $\Theta(S) \subseteq \text{VB}(S)$.*

Proof. The lemma can be proved by first checking

$$\xi(\ulcorner \varphi \urcorner, X) \Rightarrow \#\varphi \in \text{VB}(X)$$

whenever X is VB-sound. The closure of the VB operator under PAT-consequence yields the claim.
$\qquad \square$

Lemma 35. *Given a set $S \subseteq \text{Sent}_{\mathcal{L}_{\text{Tr}}}$, we write $\text{PAT} \cup \{\text{Tr}\ulcorner \psi \urcorner \mid \#\psi \in S\} \vdash^\omega \varphi$ iff φ is provable in ω-logic from PAT and each sentence of the form $\text{Tr}\ulcorner \psi \urcorner, \#\psi \in S$, as axioms. Now, let $X = \Theta(X)$. Then:*

$$\text{PAT} \cup \{\text{Tr}\ulcorner \psi \urcorner \mid \#\psi \in X\} \vdash^\omega \varphi \Rightarrow \#\varphi \in X$$

Proof. This is immediate by induction on the length of the (ω)-derivation of φ in $\text{PAT} \cup \{\text{Tr}\ulcorner \psi \urcorner \mid \#\psi \in S\}$. The axioms of PAT are given by the θ formula, and the axioms of the form $\text{Tr}\ulcorner \psi \urcorner, \#\psi \in X$, follow from the fixed-point property. Since Θ fixed points are closed under modus ponens and ω-logic, the claim follows. $\qquad \square$

---

[7] Again, just take the fixed point that obtains by applying Θ^* to $\{\#\tau_1 \vee \tau_2\}$.

Proposition 36. *Let $X \subseteq \text{VB}(X)$. Consider the following constructions:*

$$\Gamma_0^{\text{vb}} = X \qquad\qquad\qquad\qquad \Gamma_0^{\Theta} = X$$

$$\Gamma_{\alpha+1}^{\text{vb}} = \text{VB}(\Gamma_\alpha^{\text{vb}}) \qquad\qquad\qquad \Gamma_{\alpha+1}^{\Theta} = \Theta(\Gamma_\alpha^{\Theta})$$

$$\Gamma_\lambda^{\text{vb}} = \bigcup_{\beta < \lambda} \Gamma_\beta^{\text{vb}}, \text{ for } \lambda \text{ a limit} \qquad\qquad \Gamma_\lambda^{\Theta} = \bigcup_{\beta < \lambda} \Gamma_\beta^{\Theta}, \text{ for } \lambda \text{ a limit}$$

Then, $\bigcup_{\beta \in \text{On}} \Gamma_\beta^{\text{vb}} = \bigcup_{\beta \in \text{On}} \Gamma_\beta^{\Theta}$.

Proof. For clarity of notation, let $\bigcup_{\beta \in \text{On}} \Gamma_\beta^{\Theta} = S$. We have that $\Theta(S) = S$, and also $\text{VB}(S) \supseteq \Theta(S)$. We assume, for reductio, that $S \neq \bigcup_{\beta \in \text{On}} \Gamma_\beta^{\text{vb}}$. By Lemma 34, we know that $\Theta(X) \subseteq \text{VB}(X)$; and, in fact, for any $\alpha \in \text{On}$, $\Theta^\alpha(X) \subseteq \text{VB}(\Theta^\alpha(X))$ (where Θ^α stands for α-th iterations of Θ). Therefore, $S \subseteq \text{VB}(S)$. But then $S \neq \bigcup_{\beta \in \text{On}} \Gamma_\beta^{\text{vb}}$ implies $\text{VB}(S) \supsetneq S$. So there must be some sentence φ s.t. $\#\varphi \in \text{VB}(S)$ but $\#\varphi \notin \Theta(S) = S$. By Lemma 35, we obtain:

$$\text{PAT} \cup \{\text{Tr}\ulcorner\psi\urcorner \mid \#\psi \in S\} \nvdash^\omega \varphi$$

By the completeness theorem, there must be a standard model $(\mathbb{N}, S')$, with $S' \supseteq S$, s.t. $(\mathbb{N}, S') \nvDash \varphi$. But then $\#\varphi \notin \text{VB}(S)$, yielding a contradiction. $\qquad\square$

Let us now examine in greater depth the connection between the class of classically sound Θ^* fixed points and the VB fixed points. First, let us denote:

$$\mathcal{Z} := \{S \mid S = \Theta^*(S) \ \& \ S \text{ is classically sound}\}$$

We will consider this class of fixed points and refer to them in the usual terms. In particular, we say that some fixed point S is $\mathcal{Z}$-*maximal* iff there is no $S' \in \mathcal{Z}$ s.t. $S \subsetneq S'$. By adapting a result from [Ste18] to the present setting, we can show that the $\mathcal{Z}$-maximal fixed points and the VB maximal fixed points are not the same.

Definition 37. *Let $\Phi : \mathcal{P}(\omega) \to \mathcal{P}(\omega)$ be an operator. Recall that $X^- := \{\#\neg\varphi \mid \#\varphi \in X\} \cup \{n \in \omega \mid \mathbb{N} \vDash \neg\text{Sent}_{\mathcal{L}_{\text{Tr}}}(\bar{n})\}$ and $Y^- := \{\#\neg\varphi \mid \#\varphi \in Y\} \cup \{n \in \omega \mid \mathbb{N} \vDash \neg\text{Sent}_{\mathcal{L}_{\text{Tr}}}(\bar{n})\}$. We define:*

- *X, Y are* compatible *iff $X \cap Y^- = \emptyset$.*
- *X is Φ-intrinsic (in short: intrinsic) iff $X \subseteq \Phi(X)$ and $\text{CON}(X)$ and for all Y s.t. $Y \subseteq \Phi(Y)$, X and Y are compatible.*

The set $\{n \in \omega \mid \exists X(X \text{ is intrinsic} \ \& \ n \in X)\}$ is known as the maximal intrinsic fixed point *of Φ. A formula φ is said to be Φ-intrinsic iff its code is in the maximal intrinsic fixed point of Φ.*

Notice that for the familiar operators considered in this paper, the existence of the maximal intrinsic fixed point follows from the existence of the relevant minimal fixed points, as these minimal fixed points are themselves intrinsic.

Lemma 38. *The $\mathcal{Z}$-maximal intrinsic fixed point and the maximal intrinsic fixed point of VB differ. Therefore,*

$$\{S \mid S \in \mathcal{Z} \ \& \ S \text{ is maximal}\} \neq \{S \mid S = \text{VB}(S) \ \& \ S \text{ is maximal}\}.$$

Proof. The first claim follows by complexity considerations. The upper-bound for the $\mathcal{Z}$-maximal intrinsic fixed point $(\text{I}_{\mathcal{Z}}^+)$ can be obtained as follows:

(10) $\qquad n \in \text{I}_{\mathcal{Z}}^+ \Leftrightarrow \exists X(X = \Theta^*(X) \ \& \ X \text{ is classically sound} \ \& \ \forall m \in X(\dot{\neg}m \notin \text{V}_{\Theta^*}^+) \ \& \ n \in X)$

Here, $V_{\Theta^*}^+$ stands for the set of sentences true in some Θ^* fixed point. Since the formula $\xi^*(x, X)$ of the Θ^* jump is arithmetical, $X = \Theta^*(X)$ is a Δ_1^1 formula, as the next definition shows:

$$X = \Theta^*(X) \Leftrightarrow \forall n(n \in X \Leftrightarrow \exists y(\xi^*(y, X) \wedge \text{Pr}_{\text{PAT}}(y \dot{\to} n)))$$

It follows that $V_{\Theta^*}^+$ has complexity at most Σ_1^1:

$$n \in V_{\Theta^*}^+ \Leftrightarrow \exists X(X = \Theta^*(X) \wedge n \in X)$$

Finally, the property of classical soundness is at most Δ_1^1:

$$X \text{ is classically sound } \Leftrightarrow \forall \varphi(\#\varphi \in X \Rightarrow X \vDash \varphi)$$

where we rely on the well-known fact that classical satisfaction is Δ_1^1.

Following (10), membership in $I_{\mathcal{Z}}^+$ is at most Σ_1^1-in-a-$(\omega \setminus \Sigma_1^1)$-parameter, i.e., Σ_1^1-in-a-Π_1^1-parameter. By contrast, Burgess [Bur88] showed that the maximal intrinsic fixed point of the VB scheme (I_{vb}^+) is precisely Δ_2^1-in-a-Π_2^1-parameter, so $I_{\text{vb}}^+ \neq I_{\mathcal{Z}}^+$.

The 'therefore' part of the claim follows from the fact that the maximal intrinsic point of a scheme or operator is the intersection of all maximal fixed points of the operator—see e.g. [Vis89, Th.2.18.2]. $\quad\square$

In fact, we can provide specific examples of simple fixed points that are in $\mathcal{Z}$ but are not VB fixed points. The following is a variation of an example we have used repeatedly:

Lemma 39. *Consider the following construction:*

$$\begin{aligned}
\Gamma_0 &:= \{\#(\neg\tau_1 \vee \neg\tau_2)\} \\
\Gamma_{\alpha+1} &:= \Theta^*(\Gamma_\alpha) \\
\Gamma_\lambda &:= \bigcup_{\beta<\lambda} \Gamma_\beta
\end{aligned}$$

Then, $\bigcup_{\alpha\in\text{On}} \Gamma_\alpha \in \mathcal{Z}$ but $\bigcup_{\alpha\in\text{On}} \Gamma_\alpha \neq \text{VB}(\bigcup_{\alpha\in\text{On}} \Gamma_\alpha)$.

Proof. We first show that $\bigcup_{\alpha\in\text{On}} \Gamma_\alpha$ is classically sound. We prove that $\#\varphi \in \Gamma_\alpha$ implies $\Gamma_\alpha \vDash \varphi$ by transfinite induction on α.

For the base case: it is clear that $\Gamma_0 \vDash \neg\tau_1 \vee \neg\tau_2$, since $\neg\tau_i \equiv \neg\text{Tr}\ulcorner\tau_i\urcorner$ and $\#\tau_i \notin \Gamma_0$ (i one of 1, 2). For the successor case: we can assume that $\#\varphi \in \Gamma_\alpha$ implies $\Gamma_\alpha \vDash \varphi$ and prove the claim for $\alpha + 1$. If $\#\varphi \in \Gamma_{\alpha+1}$ by $\xi^*(x, X)$, the claim follows by an induction on the positive complexity of φ. It's worth noting that in the case in which φ is of the form $\text{Tr}\ulcorner\psi\urcorner$ we employ the increasing nature of Θ^*, whereas if φ is of the form $\neg\text{Tr}\ulcorner\psi\urcorner$ we employ the consistency of Γ_α for any α.

On the other hand, if $\#\varphi \in \Gamma_{\alpha+1}$ because there is some ψ such that $\xi^*(\ulcorner\psi\urcorner, \Gamma_\alpha)$ holds and $\psi \to \varphi$ is provable in PAT, the previous argument shows $\Gamma_{\alpha+1} \vDash \psi$. But since this is a model of PAT, $\Gamma_{\alpha+1} \vDash \varphi$.

Now, we show that $\bigcup_{\alpha\in\text{On}} \Gamma_\alpha \neq \text{VB}(\bigcup_{\alpha\in\text{On}} \Gamma_\alpha)$. It is clear that $\#(\neg\tau_1 \vee \neg\tau_2) \in \bigcup_{\alpha\in\text{On}} \Gamma_\alpha$ and $\#\neg\tau_1 \notin \bigcup_{\alpha\in\text{On}} \Gamma_\alpha$, $\#\neg\tau_2 \notin \bigcup_{\alpha\in\text{On}} \Gamma_\alpha$. But no consistent fixed point of VB can contain $\#(\neg\tau_1 \vee \neg\tau_2)$ and fail to contain $\#\neg\tau_1$ or $\#\neg\tau_2$. Since $\#\neg\tau_1 \notin \bigcup_{\alpha\in\text{On}} \Gamma_\alpha$ and $\#\neg\tau_2 \notin \bigcup_{\alpha\in\text{On}} \Gamma_\alpha$, there are extensions $Y \supseteq \bigcup_{\alpha\in\text{On}} \Gamma_\alpha$ such that both $\#\tau_1 \in Y$ and $\#\tau_2 \in Y$, whence $(\mathbb{N}, Y) \vDash \neg(\neg\tau_1 \vee \neg\tau_2)$. But then, $\#(\neg\tau_1 \vee \neg\tau_2) \notin \text{VB}(\bigcup_{\alpha\in\text{On}} \Gamma_\alpha)$, so $\bigcup_{\alpha\in\text{On}} \Gamma_\alpha \neq \text{VB}(\bigcup_{\alpha\in\text{On}} \Gamma_\alpha)$. $\quad\square$

Proposition 40. $\{S \mid S = \Theta^*(S) \ \& \ S \text{ is classically sound}\} \neq \{S \mid S = \text{VB}(S)\}$

Proof. Lemma 38 yields the result. Note that one can also use Lemma 39, reasoning as follows. By that lemma, either (i) there is a $\mathcal{Z}$-maximal fixed point X which is not a VB-maximal fixed point, or (ii) there is a VB-maximal fixed point Y which is not a $\mathcal{Z}$-maximal fixed point. If (i): then either X is not a fixed point of VB, in which case the claim follows; or there is some $X' \supseteq X$ which is a fixed point of VB, and

cannot be a fixed point of $\mathcal{Z}$, since X is $\mathcal{Z}$-maximal, whence the claim follows. An analogous reasoning applies if (ii) is the case. $\square$

The situation is very similar when we consider the operator Θ_c and the supervaluationist scheme VC. We omit the proofs of the following results, which are just reproductions of all proofs we have carried out for VB:

Proposition 41. *The following hold:*

    (1) *Let* $S = \Theta_c^*(S)$. *Then,* $S = \Theta_c(S)$.
    (2) $I_{\Theta_c} = I_{\Theta_c^*} = I_{vc}$.
    (3) $I_{ssk_c} \subsetneq I_{vc}$
    (4) $S = \Theta_c^*(S)$ *&* S *is classically sound* $\Leftrightarrow S \vDash IT_c$.
    (5) *Let* $\mathcal{Z}_c = \{S \mid \Theta_c^*(S)$ *&* S *is classically sound*$\}$. *Then,* $\mathcal{Z}_c \supseteq \{S \mid S = VC(S)\}$.
    (6) *The* $\mathcal{Z}_c$-*maximal intrinsic fixed point and the maximal intrinsic fixed point of* VC *differ.*
    (7) *Let* Y_c *be the* Θ_c^* *fixed point obtained by the construction of Lemma 39, with* Θ_c^* *in place of* Θ^*. *Then,* $Y \in \mathcal{Z}_c$ *but* $Y \neq VC(Y)$.

In the case of MC and Θ_{mc}^*, it must be noted that the scheme MC is not itself classically sound. Therefore, we can forget about that aspect when establishing their relations. Thus, the connection is even more direct than the case of Θ^* and Θ_c^*, although the proofs do not change much:

Proposition 42. *The following holds:*

    (1) *Let* $S = \Theta_{mc}^*(S)$. *Then,* $S = \Theta_{mc}(S)$.
    (2) $I_{\Theta_{mc}} = I_{\Theta_{mc}^*} = I_{mc}$.
    (3) $I_{ssk_{mc}} \subsetneq I_{mc}$
    (4) $S = \Theta_{mc}^*(S) \Leftrightarrow S \vDash IT_{mc}$.
    (5) *If* $S = MC(S)$, *then* $S = \Theta_{mc}^*(S)$.
    (6) *The* Θ_{mc}^*-*maximal intrinsic fixed point and the maximal intrinsic fixed point of* MC *differ.*

Yet something makes the case of Θ_{mc}^* particularly different: the example given in Lemma 39 does not apply here. The reason is simple: when one considers *maximally consistent* extensions of a set Y containing $\#(\neg\tau_1 \vee \neg\tau_2)$, the maximal consistency requirement implies that at least one of the disjuncts ($\#\neg\tau_1$ or $\#\neg\tau_2$) must be in the extension as well.

This being said, here are a few more things we can uncover about where MC and Θ_{mc}^* come apart. As we did before with Θ^* and Θ_c^*, we display a concrete counterexample of a fixed point of Θ_{mc}^* which is not a fixed point of MC.

Proposition 43. *Let* λ_1, λ_2 *be two Liar sentences, i.e.,* PAT $\vdash \lambda_i \leftrightarrow \neg Tr\ulcorner\lambda_i\urcorner$ *(for* $i \in \{1, 2\}$*). Consider the following construction.*

$$
\begin{aligned}
\Gamma_0 &:= \{\#(\lambda_1 \leftrightarrow \lambda_2)\} \\
\Gamma_{\alpha+1} &:= \Theta_{mc}^*(\Gamma_\alpha) \\
\Gamma_\lambda &:= \bigcup_{\beta<\lambda}\Gamma_\beta
\end{aligned}
$$

Then, $\bigcup_{\beta\in On}\Gamma_\beta \neq MC(\bigcup_{\beta\in On}\Gamma_\beta)$.

Proof. First, it must be noted that $\bigcup_{\beta\in On}\Gamma_\beta$ is indeed a consistent Θ_{mc}^* fixed point: we know that every consistent set of sentences (as is $\{\#(\lambda_1 \leftrightarrow \lambda_2)\}$) generates a consistent Θ_{mc}^* fixed point. Moreover, as before, $\{\#(\lambda_1 \leftrightarrow \lambda_2)\} \subseteq \bigcup_{\beta\in On}\Gamma_\beta$. Secondly, we check that no fixed point of MC can contain $\#(\lambda_1 \leftrightarrow \lambda_2)$.

The reasoning is as follows. It is easy to verify that no fixed point can contain either λ_1 or λ_2. But then, if X were a fixed point containing $\#(\lambda_1 \leftrightarrow \lambda_2)$, there would be, e.g., a (maximally-consistent) extension $X' \supseteq X$ for which $\#\lambda_1 \in X'$ and $\lambda_2 \notin X'$. Then, $(\mathbb{N}, X') \nVDash \lambda_1 \leftrightarrow \lambda$, so $\mathrm{MC}(X) \neq X$. □

Corollary 44. *There is a Π_1^1 fixed point of Θ_{mc}^* which is not a fixed point of* MC.

Proof. The fixed point constructed in the proof of Proposition 43 suffices. The (positive) parameter in the induction, $\{\#(\lambda_1 \leftrightarrow \lambda_2)\}$, is arithmetical. Hence, the fixed point can be described by a Π_1^1-formula. □

Remark 2. The construction:

$$\begin{aligned}
\Gamma_0 &:= \{\#(\lambda_1 \leftrightarrow \lambda_2)\} \\
\Gamma_{\alpha+1} &:= \Theta_{\mathrm{e}}^*(\Gamma_\alpha) \\
\Gamma_\lambda &:= \bigcup_{\beta < \lambda} \Gamma_\beta,
\end{aligned}$$

for Θ_{e}^* one of $\Theta^*, \Theta_{\mathrm{c}}^*$, also yields a fixed point of the corresponding Θ^* operator which isn't a fixed point of VB or VC, respectively. This observation draws on [Bur86], who already proved that no supervaluationist fixed point can contain the code of the sentence $\#(\lambda_1 \leftrightarrow \lambda_2)$.

8. Conclusion

In the introductory Section 1, we have claimed that the SSK THESIS can be vindicated, albeit in a less uniform manner than was envisaged by [Ste18]. In the light of the results just presented, we now explain in more detail what the vindication and the non-uniformity consists in, and sketch their philosophical import.

We can say that the SSK THESIS has been vindicated on the following grounds. First, all the operators we have considered – SSK, Θ, and Θ^* – give rise to a fixed-point semantics that accounts for all first-order penumbral truths of classical logic and arithmetic. That is, they preserve what in Section 1 we dubbed the *essential aspects* of supervaluational truth. Second, the fixed points in question have a considerably lower recursion-theoretic complexity, and are therefore amenable to $\mathbb{N}$-categorical axiomatizations, as we have shown. Moreover, the SSK-fixed points are true to the semantic motivation supporting the supervaluation schemes we discussed in Section 1: the SSK-scheme collects the first-order logical consequences of Strong Kleene truths. In contrast, the Θ- and Θ^*-fixed points merely collect consequences of a given hypothesis. Yet, the flipside is that attractive Θ (Θ^*)-fixed points are fixed points of supervaluational truth, as they actually collect consequences in ω-logic. SSK-fixed points are not closed under ω-logic and will thus typically fall short of supervaluational truth, that is, SSK-fixed points are usually not supervaluational fixed points.[8]

We see that each of these incarnations of supervaluational-style truth has its own strengths and weaknesses. We now analyze them in turn along key dimensions:

Principles of Truth. Although the fixed-point models of SSK, Θ, and Θ^* have all $\mathbb{N}$-categorical axiomatizations, this does not entail that all such axiom systems are equally natural.

The principles of truth for IT$^-$, IT* and IT* + Tr-Out preserve the desirable features of the the theory IT discussed in [Ste18]. In particular. they are arguably as compositional as supervaluational-style truth can be. In fact, following Stern's characterization of IT in [Ste18, p. 841], they have the advantage that, when full compositionality fails, we can locate where it does and the extent of this failure by means of a

---

[8]We do not know whether there are SSK-fixed points that are also supervaluational fixed points, and whether there is a nice class of such fixed points. We conjecture that this is not the case.

first-order formula – $\xi(x, X)$ for IT$^-$, $\xi^*(x, X)$ for IT* and IT* + Tr-Out. But even between these theories important differences arise. As we have seen, the lack of a right-to-left direction in axioms IT$^-$6 and IT$^-$7, which is remedied in IT* and IT* + Tr-Out, makes the latter more natural than IT$^-$: unlike IT$^-$, IT* and IT* + Tr-Out allow for full truth disquotation inside the truth predicate. In addition, IT* + Tr-Out stands out for its proof-theoretic strength, which matches the one of IT. They amount to the strongest, $\mathbb{N}$-categorical axiomatic theories of truth available.

The truth predicate of the IT-theories commutes with the universal quantifier, in contrast to the truth predicate of the theories of PK and PK$^+$, as the Θ-fixed points, unlike the SSK-fixed points, are closed under the ω-rule. While there are good reasons why the SSK-fixed points are not closed under ω-logic, it means that the theories PK and PK$^+$ cannot give a satisfactory semantic explanation for the truth of universally quantified sentences. Putting universally quantified sentences aside, PK and PK$^+$ preserve the desirable features of theories of supervaluation-style truth as highlighted by Lemma 13:

- they allow for full disquotation inside the truth predicate;
- they allow for double negation introduction/elimination inside the truth predicate
- they commute with conjunction;
- they locate the failure of compositionality by means of the formula $\theta(x)$ (for disjunctions, existential sentences and universal sentences).

Moreover, from the semantic perspective PK and PK$^+$ are more successful in limiting the failure of compositionality since $\mathrm{Tr}\ulcorner\neg\tau_1 \vee \neg\tau_2\urcorner$ will be true in the models of PK and PK$^+$ only if either $\mathrm{Tr}\ulcorner\neg\tau_1\urcorner$ or $\mathrm{Tr}\ulcorner\neg\tau_2\urcorner$ (cf. Lemma 39) is true in the model.

However, the theory PK assumes an, arguably, fairly *ad hoc* axiom, TBπ, which seems difficult to motivate. The axiom is required to guarantee that we can express $S \vDash_{\mathrm{sk}} \varphi$ over the standard model. Introducing TBπ enables us to define this notion by a positive inductive definition (relative to the standard model). Without this axiom we would not be able to show the $\mathbb{N}$-categoricity of PK and the theory would not behave in the intended way. PK$^+$ remedies this by allowing disquotation for all Tr-positive sentences. Arguably, this is a more natural restriction to the naive Tr-schema, as it is justified by the role played by negation (or equivalent logical tools) in the Liar paradox. In addition, PK$^+$ features the axiom schema of Tr-Out, which naturally follows from the classical soundness of SSK-fixed points.

Semantic Rationale. We have already mentioned that the SSK fixed-point semantics comes equipped with semantico-philosophical rationale. For one, the fixed-point semantics is associated with a semantic evaluation scheme, the SSK-scheme, which, arguably, captures directly the core idea of 'first-orderizing' the 'second-order' supervaluation schema. For another, but for related reasons, the SSK-fixed points are submitted to semantic scrutiny: every sentence in a given SSK-fixed point S must be true in the model $(\mathbb{N}, S)$ according to the SSK-scheme, that is, SSK-fixed points are semantically self-supporting. Since the SSK-scheme is classically sound this also implies that the members of S are classically true in the model $(\mathbb{N}, S)$.

Neither holds for the fixed-point semantics associated with the operations Θ or Θ^*: there is no motivated semantic evaluation scheme associated with these fixed-point semantics, nor are the fixed points subjected to any form of semantic scrutiny. The latter shows in the fact that Θ and Θ^* are sound for any given hypothesis $S \in \mathcal{P}(\mathrm{Sent})$, that is, starting from any set of sentences S we can inductively define a fixed point $I_S \supseteq S$ of Θ and Θ^* respectively. This means that the sentences in the starting hypothesis are accepted as true without scrutiny independently of their semantic status. Unsurprisingly then, not all fixed points of Θ and Θ^* are classically sound – and, even if they are classically sound, they may contain, e.g., disjunctions that are neither supported by the Strong Kleene scheme nor PAT-truths (see the example of Lemma 39 discussed above). So, generally speaking, the Θ and Θ^* fixed points lack semantic justification. Rather, these fixed points can be conceived of as the set of theorems derivable within the

proof system outlined in Proposition 9 with S as additional axioms. Ultimately, this means that, from a semantic perspective, the fixed points of Θ and Θ^* will be well supported and motivated if(f) the initial hypothesis is. However, motivating a particular hypothesis is external to the fixed-point semantics of Θ and Θ^*. Yet, if such a motivation for an initial hypothesis S is given, e.g., that it is supervaluationally self-supporting, then this motivation is indeed fully transferred to the fixed point I_S constructed over S: I_S will be a supervaluational fixed point. Indeed, as we showed in Proposition 36, all supervaluational fixed points can be constructed using Θ and Θ^* in such a way.

Two ways of first-orderizing. Summing up, the SSK- and Θ-operators can be considered as two different ways of 'first-orderizing' the supervaluation scheme. The SSK-operator and scheme provide us with a first-order formulation of the semantic idea underlying the supervaluation scheme. In contrast to the supervaluation scheme, it *only* collects the first-order penumbral truths and falls short of collecting second-order penumbral truths. This also explains why the SSK-fixed points are not closed under the ω-rule. The SSK-scheme embraces the idea of the supervaluation scheme but implements it in an independent way that leads to a distinct (and disjoint) class of supervaluation-style fixed points and a novel supervaluation-style theory.

In contrast, the various Θ-operators 'first-orderize' the supervaluational fixed-point construction, as these operators are defined via a first-order arithmetical formula. We can then use the first-order construction to extract axiomatic truth theories that are as close as possible to a proof-theoretic characterization of supervaluational truth (cf. Proposition 27). Yet, in contrast to the SSK-scheme, there is no semantics in the ordinary sense associated with these operators, and the fixed-point construction does not lend any semantic justification to the sentences in the fixed points. The fixed points of the various Θ-operators are somewhat 'hit-and-miss', depending on the choice of the initial hypothesis. The moral of our discussion then is that we can first-orderize supervaluational truth either by giving a first-order formulation of the underlying semantic idea or by first-orderizing the fixed-point construction. However, these are two different projects that, unfortunately, cannot be performed in an integrated and uniform manner.

9. Open problems

After our investigation, there remain some open problems that may be addressed in future research. Some of them concern the proof-theoretic strength of many of the theories that we have introduced here.

Open problem 1. *What is the proof-theoretic strength of* PK? *And of* PK$^+$?

By definition, we know that PK$^+$ has at least the proof-theoretic strength of PUTB, but it is still open whether this lower-bound is precise. By contrast, both the lower- and the upper-bound of PK are still undetermined.

Open problem 2. *What is the proof-theoretic strength of* IT$^-$? *And* IT*?

Proposition 28 established a lower-bound for both IT$^-$ and IT*, namely the strength of RT$_{<\varepsilon_0}$. As with PK$^+$, it remains unknown whether this bound is exact; while the strength of IT represents an upper-bound, it is clearly not sharp.

A further question is related to SSK-fixed points. We have shown how some said fixed points are not closed under ω-logic, e.g., the least SSK fixed point. But we lack a general proof to the effect that no SSK fixed point is closed under ω-logic:

Open problem 3. *Are there* SSK-*fixed points which are closed under* ω-*logic? And if so, is there a general procedure for obtaining these fixed points?*

If, unlike we conjecture, the answer to the first question is in the affirmative, it would still be open whether any such fixed point coincides with a supervaluational fixed point:

Open problem 4. *Is there a class of* SSK-*fixed points that are also supervaluational fixed points?*

REFERENCES

[Bur86] John P. Burgess. The truth is never simple. *The Journal of Symbolic Logic*, 51(3):663–681, 1986.

[Bur88] John P. Burgess. Addendum to the "truth is never simple". *Journal of Symbolic Logic*, 53(2):390–392, 1988.

[Can90] Andrea Cantini. A theory of formal truth arithmetically equivalent to ID_1. *Journal of Symbolic Logic*, 55(1):244–259, 1990.

[Can96] Andrea Cantini. *Logical Frameworks for Truth and Abstraction: An Axiomatic Study.* Elsevier Science B.V., New York, 1996.

[Fef91] Solomon Feferman. Reflecting on incompleteness. *Journal of Symbolic Logic*, 56(1):1–49, 1991.

[FHKS15] Martin Fischer, Volker Halbach, Jönne Kriener, and Johannes Stern. Axiomatizing semantic theories of truth? *The Review of Symbolic Logic*, 8(2):257–278, 2015.

[Fie08] Hartry H. Field. *Saving Truth from Paradox.* Oxford University Press, 2008.

[Fuj10] Kentaro Fujimoto. Relative truth definability of axiomatic truth theories. *The Bulletin of Symbolic Logic*, 16(3):305–344, 2010.

[GB93] Anil Gupta and Nuel D. Belnap. *The revision theory of truth.* MIT Press, 1993.

[Hal14] Volker Halbach. *Axiomatic theories of truth.* Cambridge University Press, 2014.

[HP17] Petr Hájek and Pavel Pudlák. *Metamathematics of first-order arithmetic*, volume 3. Cambridge University Press, 2017.

[Kri75] Saul Kripke. Outline of a theory of truth. *The Journal of Philosophy*, 72(19):690–716, 1975.

[McG91] Vann McGee. *Truth, Vagueness, and Paradox: an essay on the logic of truth.* Hackett Pub. Co., 1991.

[Mos74] Yiannis Nicholas Moschovakis. *Elementary Induction on Abstract Structures.* Mineola, N.Y.: Dover Publications, 1974.

[Poh09] Wolfram Pohlers. *Proof Theory. The first step into impredicativity.* Springer, 2009.

[Ste18] Johannes Stern. Supervaluation-style truth without supervaluations. *Journal of Philosophical Logic*, 47(5):817–850, 2018.

[Vis89] Albert Visser. Semantics and the liar paradox. In D. Gabbay, editor, *Handbook of Philosophical Logic vol. 4*, pages 617–706. Dodrecht, 1989.