Journal of Philosophical Logic manuscript No.
(will be inserted by the editor)

Revision without revision sequences: Self-referential truth

Edoardo Rivello

Received: date / Accepted: date

Abstract The model of self-referential truth presented in this paper, named *Revision-theoretic supervaluation*, aims to incorporate the philosophical insights of Gupta and Belnap's *Revision Theory of Truth* into the formal framework of Kripkean fixed-point semantics. In Kripke-style theories the final set of *grounded* true sentences can be reached from below along a strictly increasing sequence of sets of grounded true sentences: in this sense, each stage of the construction can be viewed as an improvement on the previous ones. I want to do something similar replacing the Kripkean sets of grounded true sentences with revision-theoretic sets of *stable* true sentences. This can be done by defining a monotone operator through a variant of van Fraassen's supervaluation scheme which is simply based on ω-length iterations of the Tarskian operator. Clearly, all virtues of Kripke-style theories are preserved, and we can also prove that the resulting set of "grounded" true sentences shares some nice features with the sets of stable true sentences which are provided by the usual ways of formalising revision. What is expected is that a clearer philosophical content could be associated to this way of doing revision; hopefully, a content directly linked with the insights underlying finite revision processes.

Keywords Self-referential truth · Revision · Supervaluation

1 Introduction

The revision-theoretic method was introduced in the field of semantic theories of self-referential truth by Herzberger [15] and by Gupta [7], independently of each other. Since then, the method has been applied by many authors, in a variety of forms and having in mind different purposes and different subject matters. This story shows revision to be a versatile method of conceptual analysis, mostly in dealing with notions or definitions which present some aspect of

Department of Philosophy and Educational Sciences // University of Torino // Via S. Ottavio, 20 // 10124 Torino (Italy) // E-mail: rivello.edoardo@gmail.com

circularity; the notion of truth still remaining the most notable example. However, it is not so easy to recognise a common philosophical method of analysis behind all proposals which are labelled "revision-theoretic". Much more likely, this label seems to identify a common formalism, a shared mathematical tool, rather than a single conception of truth or circularity. The reason is that all formal versions of revision theories are based on the same mathematical notion of *revision sequence* and thus, on a formal level, they can be easily identified as variants of a same pattern.

Revision sequences are a particular kind of transfinite iterations of a unary operation, the *revision operator*, which is supposed to capture the revision-theoretic character of the phenomenon under examination: for instance, if we are interested in a self-applicable truth predicate, the Tarskian rules for evaluating truth in a model play the role of a revision operator in that, given a hypothesis about the extension of the truth predicate, by applying the rules we obtain a revised extension. The formalism of revision sequences involves some technicalities which have been made object of criticism, in the literature, in particular with respect to some aspects of the role these technicalities play in the application of the revision-theoretic approach to philosophical issues. Among others: (a) the need of arbitrary choices in order to prolong finite iterations of the revision operator into the transfinite; (b) the philosophical interpretation of the formalism; (c) the complexity of the full machinery of revision based on revision sequences.

This situation suggested to me the general project of "doing revision without revision sequences". The hypothesis underlying the project is the idea that a "process of revision" bears an intuitive content which, in principle, might be captured by a formalism other than the one based on revision sequences.

The solution I want to explore firstly identifies the common mathematical core of all proposed revision theories as the notion of *ω-revision*, namely an iteration of length ω of the revision operator resulting in the standard notion of *ω-stability*[1]. Assuming that (1) ω-revision is universally accepted as an adequate formalisation of the intuitive idea of an infinite revision process, and (2) ω-revision does not suffer the same drawbacks of transfinite revision sequences, I offer an alternative way of prolonging the revision process into the transfinite. In a nutshell, while each revision sequence restarts the process at every limit stage by applying some "bootstrapping policy" and only at the end of the process the elements which are stable in all revision sequences are collected together, I propose first to collect all elements which are stable after each ω-length iteration of the revision operator and secondly restart the process by applying a "supervaluational" move: this strategy leads to a kind of revision process which is monotonic and so, arguably, more tractable than revision sequences.

In the end, we find ourselves with two competing formalisms: (I) the notion of revision sequence, used in all variants of revision theories so far proposed in the literature, and (II) the newly introduced notion of *revision-theoretic*

[1] See Section 3 for a definition of ω-stability.

supervaluation. In which sense the latter can be understood as a way of "doing revision" (without revision sequences) and, hopefully, as a way of doing so "better"? Clearly, these are not precise mathematical questions, however, I believe that some formal results which will be proved in this paper may support the answers. We can distinguish, in these questions, one component which is largely independent from the specific philosophical topic the revision-theoretic approach is applied to, from a second component more affected by the intended application. The notion of revision sequence represents an adaptable mathematical tool which, when applied to the study of a specific philosophical topic, like self-referential truth, needs to be integrated with other features in order to provide a full revision-theoretic account of the problem. This fact leads to the variety of variants of revision which are proposed in the logico-philosophical literature. What makes all these theories "revision-theoretic", from a mathematical point of view, is the fact that all are based on the same mathematical notion of revision sequence. Analogously, revision-theoretic supervaluation intends to be a formal tool able to be adapted in different ways in order to handle different kinds of philosophical problems. Taken as a general and purely mathematical tool, the formalism proposed here is clearly "revision-theoretic" in the sense of being focused on the same basic notion the standard revision theory is based on, namely, the notion of ω-revision. And, as a general tool to be applied in addressing philosophical problems, the same formalism can also be judged "simpler" than revision sequences, at least by those authors that judge fixed-point semantics simpler than revision-theoretic semantics.

The second component of our questions, asks for a comparison of the two methods — standard revision and revision-theoretic supervaluation — in handling specific philosophical problems. In this case we have to consider: which role revision sequences are intended for; which kinds of problems they help to solve; which reasons are applied when one claims that some formal accounts based on revision sequences is superior to some other based on a different formalism. Questions like these ones can hardly be answered — and, in some cases, even be formulated — at this level of abstraction and generality. We have to turn our attention to specific applications of the revision-theoretic approach. The two main applications of revision sequences considered in Gupta and Belnap's book [11] — by far the main reference on revision — are: (i) the theory of (circular) definitions, and (ii) the theory of (self-referential) truth. Accordingly, I chose to test the revision-theoretic supervaluational approach on these two topics first. I dealt with the theory of definitions in [25], where (a) it is shown how to apply the revision-theoretic supervaluational approach in order to get a theory of definitions substantially equivalent to that provided in Gupta and Belnap's book, and (b) it is suggested a way of modifying the revision-theoretic supervaluational operator in order to overcome some alleged deficiencies of the standard revision theory of definitions.

The goal of the present paper is to complement its companion [25] by showing that, even in the case of self-referential truth, a revision-theoretic analysis without using revision sequences is possible. The result is a formal theory of truth which combines ω-revision with fixed-point semantics. As expected, these two components produce a theory which overcomes some of the technical issues of the standard approach to revision and, at the same time, largely preserves the truth-theoretic content of this latter.

I will provide in Section 2 a quick review of the standard revision-theoretic formalism based on revision sequences, as presented by Gupta and Belnap in their book. In Section 3, I present my own proposal, *revision-theoretic supervaluation*. In order to make the present paper self-contained, the revision-theoretic supervaluational approach is described from scratch, repeating to some extent the presentation already done in the companion paper [25]. The two papers are intended to cover different aspects of the revision-theoretic supervaluation which are best exemplified, respectively, either by the application of the method to the semantics of circular definitions or to the semantics of self-referential truth (over the standard model of arithmetic). More details on this latter application are given in Sections 4 and 5, where the outcomes of revision-theoretic supervaluation are contrasted, respectively, with the standard supervaluational and with the revision-theoretical approaches to self-referential truth. The mathematical claims are stated, in the main part of the text, referring to their intended application, namely, the theory of truth: however, their proofs are postponed in the appendices and, in some cases, these proofs refer to a more abstract reformulation of the claims, in order to emphasise the generality of the theorems and to make it easier to link these results to the corresponding ones left unproved in [25].

2 The revision theory of truth

The core concepts of revision are introduced by Gupta as follows:

> I propose [...] that we view the concept of truth as characterized by a *revision* procedure. [...] The revision rule associated with truth is very simple: it is essentially the rule that was formalized by Tarski in his definition of "truth-in-a-model".
>
> [...]
>
> Intuitively what is wanted is a way of summing up the improvements that are brought about by each successive application of [the revision rule]. That is, we want a way of going from the improvements that are collectively brought about by these applications. I suggest that to achieve this we rely on the *stability* property of improvements [7, pp. 37, 39].

Let me recall here how the above conception of truth is formalised in Gupta and Belnap book *The Revision Theory of Truth* [11]. For the sake of simplicity, we will stick to the standard arithmetical setting, which is one of the most studied in the literature on formal theories of truth[2].

[2] An important feature of the standard arithmetical setting is that the rule of revision associated with truth has no fixed points, because of the presence of the Liar sentence. For a discussion of cases in which one rule of revision can have one or more fixed points see [25].

We will denote the language of arithmetic by L. It is assumed that L is a first-order language suitable to formalise Peano Arithmetic. Formulæ and sentences of L will be denoted by variously decorated Greek letters like $\phi, \psi, \psi', \ldots$

The set of all nonnegative integers is denoted by ω and it will also be identified with the first infinite ordinal number. The standard model of arithmetic, denoted by $\mathbb{N}$, is given by ω equipped with all objects mentioned in L.

An additional unary predicate T, intended to represent "true", yields the language $\mathsf{L}_\mathsf{T} = \mathsf{L} \cup \{\mathsf{T}\}$. By identifying each sentence ϕ of L_T with its corresponding Gödel code in ω, subsets of ω are intended as possible interpretations of the truth predicate T and will be called *hypotheses*. We will write $\phi \in \mathbb{N}$ and $X \subseteq \mathbb{N}$ to refer, respectively, to the code of ϕ belonging to ω and to a set X of codes of formulæ as a subset of ω. Usually we will identify a hypothesis $X \subseteq \mathbb{N}$ with its corresponding *characteristic function* $h_X : \omega \to \{\mathbf{t}, \mathbf{f}\}$ defined by

$$h_X(\phi) = \mathbf{t} \Leftrightarrow \phi \in X,$$

underlining the distinction between X and h whenever it makes a difference. The *expansion* of $\mathbb{N}$ by a hypothesis X (or, h) will be denoted by $\mathbb{N}+X$ ($\mathbb{N}+h$, respectively).

The "revision rule associated with truth" is formalised by the *Tarskian operator* τ on hypotheses, defined by

$$\tau(X) = \{\phi \mid \mathbb{N} + X \models \phi\}.$$

Thus, if h is the characteristic function of X, then $\tau(h)$ is defined as the characteristic function of $\tau(X)$.

Finite iterations of the Tarskian operator are inductively defined[3] as follows:

- $\tau^0(h) = h$.
- $\tau^{n+1}(h) = \tau(\tau^n(h))$.

The "stability property" of successive applications of the revision rule is defined as follows. Given any sequence $S = \langle h_\alpha \mid \alpha \in \mathsf{lh}(S)\rangle$ of hypotheses, where $\mathsf{lh}(S)$ — the *length of* S — is either a limit ordinal or the class On of all ordinals, we define:

- $\mathsf{stab}^+(S) = \{\phi \in \mathbb{N} \mid \exists\alpha < \mathsf{lh}(S)\, \forall\beta\, (\alpha \le \beta < \mathsf{lh}(S) \implies h_\beta(\phi) = \mathbf{t})\}$.
- $\mathsf{stab}^-(S) = \{\phi \in \mathbb{N} \mid \exists\alpha < \mathsf{lh}(S)\, \forall\beta\, (\alpha \le \beta < \mathsf{lh}(S) \implies h_\beta(\phi) = \mathbf{f})\}$.

We will denote by $\mathsf{stab}(S)$ the function[4] defined by

- $\mathsf{Dom}(\mathsf{stab}(S)) = \mathsf{stab}^+(S) \cup \mathsf{stab}^-(S)$.
- $\mathsf{stab}(S)(\phi) = \mathbf{t} \Leftrightarrow \phi \in \mathsf{stab}^+(S)$, for every $\phi \in \mathsf{Dom}(\mathsf{stab}(S))$.

[3] In dealing with sequences, we will find more convenient to think at the hypotheses as characteristic functions, rather than subsets.

[4] $\mathsf{stab}(S)$ is a *partial characteristic function*, namely, a function with values in $\{\mathbf{t}, \mathbf{f}\}$ whose domain, $\mathsf{Dom}(\mathsf{stab}(S))$, is a (possibly proper) subset of $\mathbb{N}$.

For any sequence S and ordinal $\alpha < \mathsf{lh}(S)$, let $S \restriction \alpha$ denote the restriction of S to α, namely the sequence of length α we obtain from S by considering only the first α members of S. A *revision sequence* is an ordinal-length sequence $S = \langle h_\alpha \mid \alpha \in \mathsf{On} \rangle$ of hypotheses satisfying the following two requirements:

1. $h_{\alpha+1} = \tau(h_\alpha)$, for every $\alpha \in \mathsf{On}$ (namely, S is a "transfinite iteration" of the Tarskian operator).
2. $\mathsf{stab}(S \restriction \beta) \subseteq h_\beta$, for every β limit (the so-called "coherence condition").

A hypothesis h occurs cofinally many times in a revision sequence S if for every $\alpha < \mathsf{lh}(S)$ there exists $\beta \geq \alpha$ such that $h = h_\beta$. Given a class $\mathcal{C}$ of revision sequences, a hypothesis h is said to be *recurring* in $\mathcal{C}$ if and only if there exists one revision sequence S in $\mathcal{C}$ such that h occurs cofinally many times in S. We denote by $\mathcal{R}$ the set of all recurring hypotheses, namely the set of all hypotheses which occur cofinally many times in at least one revision sequence.

The notion of stability yields a tripartite classification of all sentences of $\mathsf{L_T}$ which we can represent by the partial characteristic function stab^*, defined as follows:

$$\mathsf{stab}^* = \bigcap \{\mathsf{stab}(S) \mid S \text{ is a revision sequence}\}.$$

An alternative classification is provided by Gupta and Belnap by using the notion of *near stability*. Mimicking the definition of the function $\mathsf{stab}(S)$ we define the function $\mathsf{stab}^\#(S)$ of near stability as follows:

- $\mathsf{stab}^{\#+}(S) =$
  $\{\phi \in \mathbb{N} \mid \exists \alpha < \mathsf{lh}(S)\, \forall \beta\, (\alpha \leq \beta < \mathsf{lh}(S) \implies \exists m\, \forall n \geq m\, (h_{\beta+n}(\phi) = \mathbf{t}))\}$.
- $\mathsf{stab}^{\#-}(S) =$
  $\{\phi \in \mathbb{N} \mid \exists \alpha < \mathsf{lh}(S)\, \forall \beta\, (\alpha \leq \beta < \mathsf{lh}(S) \implies \exists m\, \forall n \geq m\, (h_{\beta+n}(\phi) = \mathbf{f}))\}$.
- $\mathsf{Dom}(\mathsf{stab}^\#(S)) = \mathsf{stab}^{\#+}(S) \cup \mathsf{stab}^{\#-}(S)$.
- $\mathsf{stab}^\#(S)(\phi) = \mathbf{t} \Leftrightarrow \phi \in \mathsf{stab}^{\#+}(S)$, for every $\phi \in \mathsf{Dom}(\mathsf{stab}^\#(S))$.

The notion of near stability yields a classification $\mathsf{stab}^\#$, defined in the obvious way:

$$\mathsf{stab}^\# = \bigcap \{\mathsf{stab}^\#(S) \mid S \text{ is a revision sequence}\}$$

Finally, V^* and $\mathsf{V}^\#$ will denote the sets of *valid* sentences yielded by the classifications stab^* and $\mathsf{stab}^\#$, respectively:

- $\mathsf{V}^* = \{\phi \in \mathbb{N} \mid \mathsf{stab}^*(\phi) = \mathbf{t}\}$.
- $\mathsf{V}^\# = \{\phi \in \mathbb{N} \mid \mathsf{stab}^\#(\phi) = \mathbf{t})\}$.

Criticisms of revision In the literature on truth we can find several variants of the two revision-theoretic formalisms presented above. All these variants share the same central notion of *revision sequence* and differ in the class of revision sequences which they choose to consider[5]. Some criticisms moved

[5] Besides the original use of revision sequences made by Gupta [7] and Herzberger [15], see also Yablo [33], Gupta and Belnap [11], Chihara [4], and Welch [32].

against the revision theory of truth actually are directed towards the formalism of revision sequences rather than towards the core insight of the theory, which we can roughly identify in the idea of taking the Tarskian operator as a rule of revision. Gupta and Belnap themselves seem to agree with the different theoretical status of the revision operator vs the revision sequences[6]. In this section we will quickly review three kinds of criticism of this sort.

One first remark concerns the so-called *arbitrariness of the limit rule*, or, in other words, "what to do at limit stages". The notion of revision sequence only imposes limit stages to meet the coherence condition. Each revision sequence corresponds to infinitely many choices, one for each limit stage of the sequence. Gupta and Belnap suggest to take under consideration all possible choices, that amounts to define the recurring hypotheses with respect to the class of all revision sequences. But other strategies are possible and indeed have been explored. Therefore, the "limit rule" emerges as a "fourth parameter" [33, p. 91] which is not dictated by the other three conceptual components (the ground model, the initial hypotheses and the revision operator) of a revision process. This situation can be felt as a lack of necessity in the very definition of revision process, as prompted out in a doubtful form by Löwe:

> The fact that there are so many different systems of revision theory, all with slightly different requirements on the sequences or variations of the semantic predicate, each of them with some other set of advantages and disadvantages, is raising a concern: we are trying to model a phenomenon as central as truth; if revision theory is a fundamental tool to understanding it, shouldn't it provide answers that do not depend on such minor details? [20, p. 31][7].

Secondly, it is not entirely clear what a philosophical interpretation of the revision sequences should be[8]. One possibility, suggested by Gupta, is to read the process of revision as a process of *improvement*:

> When we learn the meaning of 'true' what we learn is a rule that enables us to improve on a proposed candidate for the extension of truth [7, p. 37].

However, transfinite revision sequences formalise the intuitive idea of a process of improvement in a very deceptive way[9]. Gupta himself warns:

> Let us note to avoid misunderstanding that sentences that are locally stable at a limit ordinal α may not be locally stable at higher limit ordinals. So we should understand the process at limit levels as summing up "seeming improvements". These "seeming improvements" may turn out to be illusory in light of later revisions [7, p. 41].

[6] See, for instance, Belnap [2, 105] and Gupta [10, 423].
[7] See also Halbach [12, p. 167, 168] and Meadows [24, Section 3.2.1, p. 6].
[8] See Shapiro [28] for an articulate analysis of this problem.
[9] See Belnap [2, p. 104], Halbach [12, p. 164], Meadows [24, Section 3.2.2, p. 6], among others.

Thirdly, there are concerns about the *complexity* of the mathematical machinery of revision sequences, as reported, for instance, by Horsten and Halbach[10]:

> One objection that is often raised to the revision theory of truth is that its notions of truth (stable truth, nearly stable truth) are too complicated: it is difficult to believe that our notion of truth is that complex [16, p. 379].

The complexity of the revision theory can be precisely computed in logical terms: the set of (the Gödel codes of) sentences which are declared stably true by the revision theories is a Π^1_2-complete set of natural numbers[11]. One consequence of this computational fact noticed by McGee, among others, is that it becomes difficult to understand the revision sequences as an idealised process of revision:

> ... an agent who had an oracle to inform her about the nonsemantic truths and who had the ability of flawlessly produce arbitrarily long first-order deductions wouldn't have the ability to recognize the stably true sentences of the truth theory for the language of arithmetic [...] Indeed, even an agent who could conduct infinitary proofs in ω-logic couldn't recognize the valid sentences, since the set of sentences derivable from the arithmetical truths in ω-logic is only Π^1_1 [23, p. 395][12].

Another consequence of the logical complexity of revision is that dealing with revision sequences can become problematic even from a purely mathematical standpoint. For instance, Löwe and Welch, investigating in particular the logical complexity of Gupta and Belnap's revision theory of truth, find that "it is easy to ask questions concerning such revision-theoretic truth sets that are independent of the axioms of ZFC" [21, p. 39]. Their conclusion is that:

> ... one lesson can be drawn immediately from the discussion above: a revision theoretical definition of truth, if it is to be applied over the natural numbers, raises questions unresolvable by the best of our current attempts to formalise set theory, and thus the foundations of mathematics [21, p. 40].

The three sorts of criticism directed towards the formalism of revision sequences, can be summarised as follows:

- The arbitrariness of choices at limit stages.
- The lack of a clear interpretation of the formalism as a process of improvement.
- The philosophical concerns connected with the high logical complexity of the notion of stable (or near stable) truth.

In the subsequent sections I will present a different formalism with the aim of formalising the intuitive idea of a revision process while avoiding the above-mentioned drawbacks of revision sequences.

[10] See also Sheard [29, p. 177], for a similar concern.
[11] Burgess [3], P. Kremer [17], Antonelli [1].
[12] See also Meadows [24, p. 6].

3 Revision-theoretic supervaluation: A sieve and a jump

There is no doubt that finite iterations $\tau^n(h)$ of the Tarskian operator τ capture the idea of revising one initial hypothesis h. Moreover, we surely have no reason to stop this process after a specific finite number of iterations. Thus we are led to consider the ω-length iteration of τ starting with a hypothesis h, also named the *trajectory* $\tau^\frown(h)$ *of* τ *starting with* h, namely:

$$\tau^\frown(h) = \langle \tau^n(h) \mid n \in \omega \rangle.$$

The notion of stability, previously defined for any sequence of limit length, clearly applies to the particular case of a sequence of length ω, so we can put[13]:

$$\tau^\omega(h) = \mathsf{stab}(\tau^\frown(h)).$$

There is the natural temptation of taking this notion $\tau^\omega(h)$ of ω-*stability*[14] as the correct formalisation of revision, so avoiding all kinds of criticism examined above. Indeed: (1) ω-stability does not introduce spurious elements other than the Tarskian operator and its iteration; (2) ω-length iterations clearly improve on the initial hypothesis as long as we accept the idea that a single application of the Tarskian operator does; (3) the mathematical definition of the set of all ω-stable elements only requires induction over the natural numbers. For instance, the attractiveness of the finite levels of revision is expressed by Halbach as follows:

> Concentrating on the finite levels of the revision process is worthwhile: one can thereby avoid all the difficult issues concerning limit levels and just capture the chief attractive feature of revision semantics, which is the revision process via the operator [τ] [12, p. 168].

Unfortunately, it is also well known that, in certain cases, countable revision only leads to unsatisfactory (say, counterintuitive) theories of truth. For instance, Visser [31, pp. 210–211] illustrates this fact by the following example. He considers the sequence of sentences:

ϕ_0 := "Snow is white"
ϕ_{n+1} := $\ulcorner$"ϕ_n" is true$\urcorner$
ϕ_ω := "$\forall n\, \phi_n$ is true"

and remarks that any countable process of revision evaluates ϕ_ω as false, but ϕ_ω is intuitively true.

The standard solution to problems of this kind is to introduce revision sequences to extend countable revision into the transfinite, but we have seen that revision sequences lead to a different kind of issues. So, we are facing a tension between two ways of doing revision: countable iterations are good

[13] This notation is borrowed from P. Kremer [18, p. 382].
[14] Note that for a sequence S of length ω the two notions of stability, $\mathsf{stab}(S)$, and near stability, $\mathsf{stab}^\#(S)$, coincide.

in formalising the intuitive idea of a revision process, but do not deliver a satisfactory theory of truth; while for revision sequences is the other way round.

Halbach [12, pp. 164–167] suggests an alternative way of seeing revision as a process of improvement. The picture is that of a "sieving procedure". Under this view, the purpose of revision is to refine the set of all possible hypotheses by repeated applications of the Tarskian operator, so leading to a set of "better" hypotheses.

In the standard presentations based on revision sequences, the set of better hypotheses is represented by the set $\mathcal{R}$ of all recurring hypotheses. The crucial role played by the recurring hypotheses in Gupta and Belnap revision theory is exemplified by the fact that both classifications of sentences, stab^* and $\mathsf{stab}^\#$, provided by the theory can be recovered from $\mathcal{R}$. Indeed, we can check that the following characterisations of stabilities in terms of recurring hypotheses hold[15]:

- $\mathsf{stab} = \bigcap\{h \mid h \in \mathcal{R}\}$.
- $\mathsf{stab}^\# = \bigcap\{\tau^\omega(h) \mid h \in \mathcal{R}\}$.

Halbach considers an interesting alternative sieving procedure: one consisting in *simultaneously* applying the Tarskian operator to all possible starting hypotheses, gradually excluding more and more candidate extensions for the truth predicate. The extensions which survive the process can be judged to be "better" than the others. Unfortunately, as Halbach shows, no hypothesis survives all finite stages of the sieving procedure which in the end results in an empty set of candidates. So the problems of "how to move from finite to transfinite revision?" or of "which is the 'right' formalisation of revision?" reappear again under a different form.

Combining the idea of a sieving procedure with that of ω-stability, we can provide a satisfactory solution to the problem of formalising revision beyond finite-length iterations. We define a "sieve" Δ_τ which filters a set of hypotheses $\mathcal{H}$ by evaluating each hypothesis $h \in \mathcal{H}$ against the outcome of performing ω-length revision starting with h. The formal definition of the operator is the following:

Definition 1 The *revision-theoretic supervaluational sieve* operator Δ_τ is the operator on sets of hypotheses defined by:

$$\Delta_\tau(\mathcal{H}) = \{h \mid \bigcap\{\tau^\omega(g) \mid g \in \mathcal{H}\} \subseteq h\}.$$

The choice of the adjectives "revision-theoretic" and "supervaluational", for the sieve operator[16] Δ_τ, will be motivated in the subsequent paragraphs[17].

[15] This fact follows from Theorem 5C.7 in Gupta and Belnap [11, p. 170].

[16] The definition of Δ_τ is formally identical to that of the operator Δ_δ given in the context of circular definitions in [25]. In the following we will often omit the subscript τ from Δ_τ since this latter is the only sieving operator we consider in this paper.

[17] See also a similar use of the adjective "supervaluational", in connection with revision theories, in Hansen [13, pp. 7–9].

A crucial point in the above definition is that the operator Δ on sets of hypotheses is *monotonic* (namely, $\mathcal{H} \subseteq \mathcal{K} \implies \Delta(\mathcal{H}) \subseteq \Delta(\mathcal{K})$)[18]. Thus, starting with the set of *all* hypotheses and iterating Δ we obtain a decreasing (under inclusion) sequence of sets of hypotheses and, at limit stages, we can take intersections in order to continue the iteration. Well known results on monotone operators on partially ordered sets[19] ensure that at some stage of the iteration we reach a fixed point of Δ (actually, the greatest one, denoted by $\mathsf{gfp}(\Delta)$). This fixed point can be chosen as the set of hypotheses to be assigned to the revision operator τ, as resulting from our "sieving procedure".

The other crucial fact about Δ is that, in contrast with the outcome of Halbach's sieving procedure, the set of hypotheses $\mathsf{gfp}(\Delta)$ is non-empty. This and other interesting mathematical properties of $\mathsf{gfp}(\Delta)$ are more easily proved by switching to a dual presentation of revision-theoretic supervaluation.

The idea, roughly speaking, is to define a sort of Kripkean "jump" operator by using ω-stability to jump from one partial interpretation to another. In Kripke's well-known construction[20], a partial interpretation for the truth predicate is represented by a pair (X^+, X^-) of disjoint subsets of $\mathbb{N}$, called the *extension* X^+ and the *antiextension* X^-. Here we find more convenient to identify each partial interpretation (X^+, X^-) with its corresponding *partial characteristic function* p defined, as expected, by

- $\mathsf{Dom}(p) = X^+ \cup X^-$.
- $p(\phi) = \mathbf{t} \Leftrightarrow \phi \in X^+$, for every $\phi \in \mathsf{Dom}(p)$.

Given a partial interpretation p of the truth predicate we consider, simultaneously, all total hypotheses h which extend p and, for each such h, we perform the corresponding ω-length iteration $\tau^{\frown}(h)$. Then we define a new partial hypothesis by collecting all sentences whose truth values are stable in all trajectories of τ. Formally, we define the jump operator[21]

$$\sigma^\omega_\tau(p) = \bigcap\{\tau^\omega(h) \mid p \subseteq h\}.$$

The jump operator σ^ω is monotonic[22], so we can consider the nonempty set of its fixed points, as in Kripke's theory. In particular, we denote by $\mathsf{lfp}(\sigma^\omega)$ the least fixed point of σ^ω and by V the set of sentences declared true by σ^ω:

$$\mathsf{V} = \{\phi \in \mathbb{N} \mid \mathsf{lfp}(\sigma^\omega)(\phi) = \mathbf{t}\}.$$

There is quite a natural connection between the partial interpretations investigated by the fixed-point semantics and the sets of hypotheses considered by the revision-theoretic semantics. Given a partial interpretation p of

[18] For, let $\mathcal{H} \subseteq \mathcal{K}$. Hence, $\{\tau^\omega(g) \mid g \in \mathcal{H}\} \subseteq \{\tau^\omega(g) \mid g \in \mathcal{K}\}$. By taking the intersections the inclusion reverses: $\bigcap\{\tau^\omega(g) \mid g \in \mathcal{K}\} \subseteq \bigcap\{\tau^\omega(g) \mid g \in \mathcal{H}\}$. Therefore, every hypothesis $h \in \Delta(\mathcal{H})$ also belongs to $\Delta(\mathcal{K})$.

[19] See, for instance, Fitting [6].

[20] Kripke [19].

[21] As in the case of the operator Δ_τ we will often omit the subscript τ from σ^ω_τ.

[22] Let $p \subseteq q$. Hence, $\{\tau^\omega(h) \mid q \subseteq h\} \subseteq \{\tau^\omega(h) \mid p \subseteq h\}$. By taking the intersections the inclusion reverses: $\bigcap\{\tau^\omega(h) \mid p \subseteq h\} \subseteq \bigcap\{\tau^\omega(h) \mid q \subseteq h\}$. Therefore, $\sigma^\omega(p) \subseteq \sigma^\omega(q)$.

the truth predicate, we can consider the set $J(p) = \{h \mid p \subseteq h\}$ of those hypotheses h which keep unchanged the truth-values assigned by p to the sentences belonging to its domain. Conversely, given a set of hypotheses $\mathcal{H}$, the set $K(\mathcal{H}) = \bigcap\{\tau^\omega(h) \mid h \in \mathcal{H}\}$ of all sentences which receive the same truth value when evaluated from a hypothesis $h \in \mathcal{H}$ forms a partial interpretation. In the revision theories described in Gupta and Belnap's book [11], we are applying K when we move from the set $\mathcal{R}$ of all recurring hypotheses to the partial interpretation $\mathsf{stab}^\# = \bigcap\{\tau^\omega(h) \mid h \in \mathcal{R}\}$ of sentences which are nearly stable in all revision sequences. Conversely, the map J is implicitly applied by the supervaluational version of the Kripkean fixed-point semantics when, at each successor stage of the approximation process, we apply the Tarskian operator to all hypotheses which extend the partial interpretation we have reached at that stage.

The two operators $\varDelta$ and σ^ω are built up from the same ingredients (the Tarskian operator, ω-stability and the two functions J and K) differently composed. Their duality patently shows in the following picture:

$$\emptyset \subseteq \left\{\begin{array}{c} h \\ g \\ \vdots \\ h' \\ g' \\ \vdots \end{array}\right\} = \mathcal{H} \mapsto \left\{\begin{array}{c} \tau^\omega(h) \\ \tau^\omega(g) \\ \vdots \\ \tau^\omega(h') \\ \tau^\omega(g') \\ \vdots \end{array}\right\} \bigcap = \sigma^\omega(\emptyset) \subseteq$$

$$\subseteq \left\{\begin{array}{c} g \\ \vdots \\ h' \\ \vdots \end{array}\right\} = \varDelta(\mathcal{H}) \mapsto \left\{\begin{array}{c} \tau^\omega(g) \\ \vdots \\ \tau^\omega(h') \\ \vdots \end{array}\right\} \bigcap = \sigma^\omega(\sigma^\omega(\emptyset)) \dots$$

Starting with the empty partial interpretation, $\emptyset$, we perform the approximation process by considering ω-length iterations from all hypotheses $h, g, h', \dots$ which extend $\emptyset$, and then by taking the intersection of the corresponding stability sets $\tau^\omega(h), \tau^\omega(g), \tau^\omega(h'), \dots$: this amounts to apply the jump operator σ^ω to the empty partial interpretation $\emptyset$. On the other hand, starting with the set $\mathcal{H}$ of all hypotheses (which obviously coincides with the set of all hypotheses extending $\emptyset$), we perform the sieving procedure by considering the partial interpretation $\sigma^\omega(\emptyset)$ obtained by ω-stabilities and intersection and then the set of all hypotheses $g, h', \dots$ which extend $\sigma^\omega(\emptyset)$: this amounts to apply the sieve operator $\varDelta$ to the set of all hypotheses $\mathcal{H}$. So the interleaved sequences of partial interpretations and of sets of hypotheses which underly the approximating and the sieving processes are substantially the same. This observation is enough to see why the set $\mathcal{H}_0 = \mathsf{gfp}(\varDelta)$ has to be nonempty: being a fixed point, $\mathcal{H}_0 = \varDelta(\mathcal{H}_0) = \{h \mid p_0 \subseteq h\}$, for some partial interpretation p_0; therefore, whichever p_0 is, $\mathcal{H}_0$ contains at least one element.

Actually, $p_0 = \mathsf{lfp}(\sigma^\omega)$, namely:

$$\mathsf{gfp}(\Delta) = \{h \mid \mathsf{lfp}(\sigma^\omega) \subseteq h\}.$$

More general, the two maps J and K form an order-preserving bijection between the sets of all fixed points of σ^ω and Δ, the former ordered by inclusion and the latter by reverse inclusion, as stated in the following

Theorem 1 *Let $J : p \mapsto \{h \mid p \subseteq h\}$ be the map which assigns to every partial hypothesis p the set $J(p)$ of all total hypotheses extending p; and let $K : \mathcal{H} \mapsto \bigcap\{\tau^\omega(h) \mid h \in \mathcal{H}\}$ be the map which assigns to every set of hypotheses $\mathcal{H}$ the intersection $K(\mathcal{H})$ of all sets of stabilities obtained performing ω-revision from a hypothesis in $\mathcal{H}$. Then J and K are order isomorphisms between the set of all fixed points of σ^ω (ordered by inclusion) and the set of all fixed points of Δ (ordered by reverse inclusion). In particular,*

$$\mathsf{lfp}(\sigma^\omega) = K(\mathsf{gfp}(\Delta)) = \bigcap\{\tau^\omega(h) \mid h \in \mathsf{gfp}(\Delta)\},$$

and

$$\mathsf{gfp}(\Delta) = J(\mathsf{lfp}(\sigma^\omega)) = \{h \mid \mathsf{lfp}(\sigma^\omega) \subseteq h\}.$$

Proof Theorem 1 immediately follows from its order-theoretic version proved in Appendix A.

In the light of Theorem 1, by the term *revision-theoretic supervaluation* we will refer to the common structure underlying both the operators Δ and σ^ω.

4 Revision-theoretic supervaluation and Kripke's fixed points

In this section we will focus on the σ^ω presentation of revision-theoretic supervaluation, which can be understood as a way of "doing revision in Kripkean clothes", so making easier a comparison with standard supervaluation[23].

The operator σ^ω can simply be viewed as a Kripkean jump based on a variant of van Fraassen's *supervaluation* scheme[24]. The original supervaluational Kripkean jump σ, in the present setting can be defined as follows:

$$\sigma(p) = \bigcap\{\tau(h) \mid p \subseteq h\}.$$

It becomes evident that σ^ω is obtained from σ just by replacing the single application $\tau(h)$ of the Tarskian operator by the notion $\tau^\omega(h)$ of ω-stability[25].

[23] The idea of combining supervaluation and revision is not entirely new, see Herzberger [15, p. 96]. Note, however, that with respect to Herzberger's suggestion the jump σ^ω combines the two approaches in the opposite way.

[24] First considered in the context of truth theories in Kripke [19, p. 711].

[25] The jump σ^ω is "supervaluational" in that it arises from the definition of σ just by replacing the function τ with the function τ^ω. The jump σ^ω, however, is not supervaluational in the sense of arising from a supervaluational evaluation scheme of the sort of those considered, for instance, in Fischer et al. [5, p. 269] or in Schindler [27, p. 12].

Conversely, σ can be viewed as a "one-step revision" since, instead of using denumerable iterated applications of τ, we apply the Tarskian operator only once[26].

The idea of preserving the philosophical theses of revision theory while adopting a Kripkean fixed-point formalism was already discussed in a debate on the revision theory between D. A. Martin, V. McGee and A. Gupta hosted in 1997 in a special number of *Philosophical Issues*[27]. The present proposal pushes this idea one step further, showing that it is possible not only to reproduce some of the outcomes of the revision theory in a Kripkean framework, but even that we can do this by directly incorporating the mathematical core of revision — i.e., ω-length iterations — in the fixed-point construction.

Which are the advantages — from a revision-theoretic standpoint — of adopting the jump σ^ω in the role of the supervaluational jump σ? The least we can say is that the arithmetical partial interpretation of the truth predicate provided by $\mathsf{lfp}(\sigma^\omega)$ avoids the criticisms made by Gupta in [7, pp. 33–36] against Kripke's fixed points. Let us briefly examine two of Gupta's objections[28].

The first one (the second in Gupta's paper) concerns the classical logical laws, which are not grounded in the least Kripkean fixed point, using either the Weak or the Strong Kleene scheme for partial logic. To overcome this problem it is enough, as remarked by Kripke himself, to replace Kleene's schemes with supervaluation. Moving from supervaluation to revision-theoretic supervaluation, this nice feature is preserved: It is easy to prove by induction that the least fixed-point $\mathsf{lfp}(\sigma^\omega)$ of revision-theoretic supervaluation validates all classical logical laws.

Gupta's second objection (the third in his paper) concerns variants of what is sometimes known in the literature as *Gupta's puzzle*. The simplest variant, in Visser's reformulation [31, p. 213] is the following:

> Call the following sentences respectively "A_1", "A_2", "B":
> – B is true.
> – B is false.
> – At most one of A_1, A_2 is true.
>
> Clearly, A_1, A_2 contradict each other, so B must be true, hence A_1 is true and A_2 is false.

Gupta observes that such pieces of reasoning (or some variants of it) are to be rejected if we accept Kripke's theory in its "least fixed point" form, even if we adopt supervaluation (in all its variants considered by Kripke) as our evaluation scheme or if we adopt an intrinsic fixed point of any scheme. By

[26] This observation motivates the choice of the notation "σ^ω" to denote the revision-theoretical jump.

[27] D. Martin [22, p. 410], McGee [23, p. 400], Gupta [10, pp. 430–434].

[28] Of Gupta's four criticisms, the first and the last one, which are concerned with models "in which there is no vicious self-reference", fall outside the scope of the present paper, which only deals with the standard model of arithmetic.

contrast, all four variants of Gupta's puzzle[29] receive the intuitively expected answer when evaluated according to revision-theoretic supervaluation[30].

From an extensional point of view the comparison between standard supervaluation and revision-theoretic supervaluation can be summarised as follows:

1. $\mathsf{lfp}(\sigma^\omega)$ is not included in the greatest intrinsic fixed point of σ.
2. For each σ-sound partial interpretation p (i.e., such that $p \subseteq \sigma(p)$) there exists the least fixed point of σ^ω above p.
3. $\mathsf{lfp}(\sigma^\omega)$ is exactly the least fixed point of σ^ω above $\mathsf{lfp}(\sigma)$.

For the first claim, the following variant of Gupta's puzzle provides an example of a sentence which belongs to the domain of the least fixed point of σ^ω, but not to the domain of the greatest intrinsic fixed point of σ. Call the following sentences "A_1", "A_2", and "B", respectively:

- "A_1 is true" is true.
- "A_1 is not true" is true.
- At most one of A_1, A_2 is true.

It is not difficult to check that the sentence B represents the sought example.

For the second and the third claim a proof is given in Appendix B, as Theorem 2.

Gupta [7, p. 37] attributes more value to the least fixed point (of any scheme) than to the other intrinsic ones (in particular, the greatest intrinsic one) due to the fact that the latter lack a stage-by-stage process to reach them. Under this respect, $\mathsf{lfp}(\sigma^\omega)$, as a least fixed point, seems to occupy a comfortable position relatively to the lattice of the intrinsic fixed points of supervaluation since, from the above-mentioned results, it follows that $\mathsf{lfp}(\sigma^\omega)$ properly extends the least fixed point of σ and is also compatible with the greatest intrinsic fixed point of σ.

5 Revision-theoretic supervaluation and Gupta-Belnap's revision

In the previous sections we have seen that the operator σ^ω can be viewed as a way of improving supervaluation in order to meet some revision-theoretic desiderata about truth.

We turn now to our original claim that revision-theoretic supervaluation can provide a better formalisation of revision than revision sequences. We can split our claim into two parts:

[29] Gupta, [7, pp. 35–37] and [10].

[30] It should be noticed that the same — and much more, in fact — happens if we use Hansen's supervaluation on trees. Hansen's proposal shares with revision-theoretic supervaluation the "idea of doing supervaluation that is not limited to one iteration of the truth predicate" [13, p. 73]. However, both the aims and the results of the two supervaluational variants are different: for one thing, Hansen's theory leads to one fixed point of the Strong Kleene jump, so preserving the compositionality of the truth predicate, while a constitutive feature of revision-theoretic truth — as reminded above — is to lack compositionality in favour of the preservation of all classical logical truths.

1. Revision-theoretic supervaluation formalises revision, and
2. Revision-theoretic supervaluation formalises revision better than any other formalisation based on revision sequences.

On the "better" part of the claim, we can observe that the sorts of criticism against revision sequences we have examined in Section 2, are often stated in the literature comparatively to the fixed-point theories[31]: it has been argued that (a) the revision-theoretic limit rule is less natural than that of a Kripkean-style approximation process; (b) the notion of "improvement" associated to the revision theories is less clear than that associated to the fixed-point semantics; (c) the logical complexity of the sets of validities V^* and $\mathsf{V}^\#$ is greater than the logical complexity of the sets of all sentences declared true by, say, the least fixed point of supervaluation (let us denote this latter by $\mathsf{V}^\sigma = \{\phi \in \mathbb{N} \mid \mathsf{lfp}(\sigma)(\phi) = \mathsf{t}\}$). Therefore, by adopting the supervaluational fixed-point formalism of the σ^ω presentation, these kinds of criticism dissolve. In particular, the logical complexity of the set V of all sentences declared true by the least fixed point of σ^ω has the same upper bound (Π^1_1) of V^σ, hence it can also be defined in purely algebraic terms without reference to transfinite ordinal numbers. We prove[32] that V is Π^1_1 in Appendix C, Theorem 3.

In the end, we can conclude that revision-theoretic supervaluation is a better formalism than revision sequences in that (a) there is no "fourth parameter" in its definition other than the ground model, the class of the initial hypotheses and the Tarskian operator, (b) it admits a clear interpretation in terms of a process of improvement and (c) does not increase the logical complexity of the set of validities beyond that of inductive definitions.

It remains to examine the first part of the claim: in which sense revision-theoretic supervaluation is a formalisation of revision at all? We can try answering this (unprecise) question in several ways.

First, we can evaluate revision sequences and revision-theoretic supervaluation just as competing mathematical methods to give "an account of those sentences that are paradoxical and those that are not, and for the latter sentences [...] an account of the conditions under which they are assertible and the conditions under which they are not assertible" Gupta [7, p. 4][33]. Since we are happy with the fact that the set V of all sentences made valid by the least fixed point of σ^ω is Π^1_1, we cannot expect V to coincide with anyone of the sets of revision-theoretically valid sentences on the market, since all these latter are at least Π^1_2-complete. A direct comparison with the two most studied sets of stabilities in Gupta and Belnap's book [11] — the sets stab^* of all sentences stable in all revision sequences, and the set $\mathsf{stab}^\#$ of all sentences nearly stable in all revision sequences — shows that (1) $\mathsf{lfp}(\sigma^\omega)$ is not included

[31] D. Martin [22, pp. 408, 416], Horsten and Halbach [16, p. 379].

[32] We owe to an anonymous referee the suggestion that the estimation of the upper bound of the logical complexity of V should be directly established by using the standard definition of the satisfaction relation and by proving by induction that, for all n, $k \in \tau^n(X)$ is Δ^1_1 in X. The proof given in Appendix C follows a more abstract reasoning which makes it immediately applicable also to the analogous result announced without proof in [25, n. 19].

[33] See also Visser [31, pp. 204–205].

in stab^*, and that (2) $\mathsf{lfp}(\sigma^\omega)$ is compatible with $\mathsf{stab}^\#$. The first claim follows from the fact that V validates more semantic principles than V^*: see below. The second claim is proved in Appendix D, Theorem 4. It follows that $\mathsf{lfp}(\sigma^\omega)$ is also compatible with stab^*, since stab^* is known to be included in $\mathsf{stab}^\#$.

In Gupta's words, one formal theory of truth "is to be tested by how well it captures our intuition about what is paradoxical and what is assertible in a given situation" [7, p. 6]. So, it seems reasonable to test the classification of sentences provided by $\mathsf{lfp}(\sigma^\omega)$ against the same intuitive examples which Gupta and Belnap use to test their classifications: the result is that for all examples treated in Gupta's and Belnap's book the answers provided by $\mathsf{stab}^\#$ and by $\mathsf{lfp}(\sigma^\omega)$ are the same.

Secondly, we can contrast revision sequences and revision-theoretic supervaluation as two alternative ways of formalising a same intuitive process of revision[34], as the one depicted in the quotation from Gupta at the beginning of Section 2. We have already argued in Section 3 that the presentation of $\mathsf{lfp}(\sigma^\omega)$ as the outcome of the sieving procedure based on the operator Δ actually captures the core insight of revision, consisting in (a) ω-stability and (b) a method to extend revision into the transfinite producing better and better hypotheses. The resulting set $\mathsf{gfp}(\Delta)$ of the hypotheses which survive the process replaces the set $\mathcal{R}$ of the recurring hypotheses in playing the role of the set of "best candidates" for the extension of the truth predicate. This substitution helps to eliminate another possible source of "arbitrariness" in the standard presentation of revision. We have seen that there are two competing ways of defining the set of all valid sentences in terms of $\mathcal{R}$. The first one — the set V^*, defined in terms of stability — has a more natural definition, as the set of all sentences which are declared true by all recurring hypotheses; while the second one — the set $\mathsf{V}^\#$, defined in terms of near stability — can be preferred as providing a nicer theory of truth. Taking $\mathsf{gfp}(\Delta)$ in the role of $\mathcal{R}$ this contrast disappears: both definitions of validity yield the same set V, namely the set of sentences which are declared true by the least fixed point of σ^ω. More explicitly, we have the following two equivalent[35] characterisations of the set $\mathsf{V} = \{\phi \in \mathbb{N} \mid \mathsf{lfp}(\sigma^\omega)(\phi) = \mathbf{t}\}$:

- $\mathsf{V} = \{\phi \in \mathbb{N} \mid \forall h \in \mathsf{gfp}(\Delta)\,(h(\phi) = \mathbf{t})\}$.
- $\mathsf{V} = \{\phi \in \mathbb{N} \mid \forall h \in \mathsf{gfp}(\Delta)\,(\tau^\omega(h)(\phi) = \mathbf{t})\}$.

Thirdly, we can contrast the sets of validities provided by standard revision theory (V^* or $\mathsf{V}^\#$) and by revision-theoretic supervaluation (V) as alternative theories of truth, namely as sets of sentences which validate one or another principle about truth.

To say the least, V shares with the other revision theoretic proposals the following "nice" properties [11, Cfr. Theorem 6C.1, p. 219]:

1. $\phi \in \mathsf{V} \Leftrightarrow \mathsf{T}\ulcorner \phi \urcorner \in \mathsf{V}$.

[34] Visser [31, p. 204].

[35] The equivalence immediately follows from the duality between the sieve Δ and the jump σ^ω established by Theorem 1.

2. V is *consistent*, namely there is no sentence ϕ such that both ϕ and $\neg\phi$ belong to V.
3. Neither the liar sentence λ nor its negation $\neg\lambda$ is in V.
4. V is closed under classical logical consequence.
5. All arithmetical truths belong to V.
6. Very long truth iterations hold in V: for instance, "$\forall n\, \mathsf{T}^{n} \ulcorner 0 = 0 \urcorner$" $\in \mathsf{V}$.

Actually, a more accurate inspection reveals that revision-theoretic supervaluation really looks as "near stability in Kripkean clothes": indeed, V, as well as $\mathsf{V}^{\#}$, validates all axioms of the Friedman-Sheard axiomatic theory of truth FS (as formulated in, for instance, Halbach [12, p. 161])[36].

6 Conclusion

Formal revision theories of truth based on revision sequences are supposed to formalise an intuitive notion of revision process. Several criticisms against revision theories are in fact directed to the formalism of revision sequences. This raises the demand for an alternative way of formalising revision.

Is revision-theoretic supervaluation the right answer? Of course, this is not a precise mathematical question and much more work has to be done in order to convert some Gupta's and Belnap's desiderata about revision in mathematical statements.

In this paper I have suggested revision-theoretic supervaluation as a possible "formalisation of revision in a Kripkean framework". To sustain this proposal I illustrated three possible lines of argumentation:

1. A strict connection between revision-theoretic supervaluation and Halbach's picture of the revision process as a "sieving procedure".
2. The intuitively correct answers given by revision-theoretic supervaluation to the examples taken by Gupta to support his criticism towards Kripkean-style theories of truth.
3. Some mathematical results which contrast the set of validities provided by revision-theoretic supervaluation with those provided either by standard supervaluation or by standard revision theories, showing a strong similarity between the revision-theoretic supervaluational and the nearly stable sets of validities.

I do not claim that the revision-theoretic proposal solves all issues concerning revision, yet it represents an interesting opportunity of changing our

[36] By McGee's ω-inconsistency theorem this fact comes with the ω-inconsistency of V as well as of $\mathsf{V}^{\#}$. This might be regarded as a weakness of V and $\mathsf{V}^{\#}$ with respect to V^{*} (for a discussion, see Gupta and Belnap [11, p. 227]). However, I stress one more time that I am not concerned myself here with the merits and flaws of revision-theoretic supervaluation as one revision theory of truth or as a theory of truth at all. What I am concerned with is to assess if revision-theoretic supervaluation can be a suitable formalisation of a revision theory of truth: whichever merits or flaws we ascribe to the latter will be inherited by the former.

perspective. The main question remains in the background: Is there a recognisable notion of revision independent from its instantiations in formal revision theories? At least, revision-theoretic supervaluation helps disentangling the informal idea of revision from the formalism of transfinite revision sequences.

The technique of revision sequences applies to any operator δ taken to replace the Tarskian operator τ in the role of the *revision* operator. The same holds for our newly introduced technique of revision-theoretic supervaluation. All mathematical claims proved in the appendices hold *verbatim* replacing τ by δ throughout the proofs[37]. This fact can be exploited in looking at possible applications of revision-theoretic supervaluation in domains other than self-referential truth. For instance, in [25], I have applied revision-theoretic supervaluation to the theory of circular definitions, taking δ to be the operator induced on the standard model of arithmetic by any first-order definition (circular or not) of a unary predicate, as in Gupta and Belnap's book [11, p. 30].

Taken together, the present paper and [25] provide two case-studies from which we can extract a new way of doing revision, namely revision-theoretic supervaluation. In both papers it is shown that the resulting theory (of definitions, in one case, and of truth, in the other) is quite similar to that yielded by the standard revision-theoretic approach in its "near stability" version. This fact strengthens our view of revision sequences and revision-theoretic supervaluation as two alternative ways of formalising a same informal process of revision. However, this fact does *not* force us to endorse the resulting theory (of definitions, of truth). We have seen at the end of the previous section that, as expected, revision-theoretic supervaluation yields an ω-inconsistent theory of truth, as well standard revision does in its "near stability" version. Analogously, revision-theoretic supervaluation, in the form here presented, does not handle in a satisfactory way implicit and inductive definitions, suffering from the same bias standard revision suffers. In [25] I tried to overcome the latter difficulty by introducing a variant of revision-theoretic supervaluation which agrees with the mathematical practice on implicit and inductive definitions. I believe that this theme of exploring variants of revision-theoretic supervaluation which are suggested by the intended applications should be investigated further: in particular, it could be followed in looking for a more reliable revision-theoretic supervaluational theory of self-referential truth.

Backing to this latter application of revision-theoretic supervaluation — the only one I dealt with in the present paper — still remains a further concern about the philosophical interpretation of the formalism. Taking σ^ω as an "evaluation scheme" we might have an argument against Hellman's claim that revision does not provide "languages with their own truth predicates" [14, p. 1071] as fixed-point semantics actually does[38]. However, I do not wish to endorse this reading of revision-theoretic supervaluation: the fixed-point version

[37] Clearly, in Theorem 3, we have to assume the operator δ to be hyperarithmetical, as τ is.

[38] See also P. Kremer [18, p. 372] and Gupta and Belnap [11, p. 60].

of revision is a useful mathematical fact, but it does not carry any philosophical "natural" interpretation of σ^ω as an evaluation scheme. The fact that the resulting partial extension might be a plausible interpretation of the truth predicate does not imply that the formal evaluation scheme we have employed in building the partial extension is also a plausible scheme in the partial-logic sense.

A similar conclusion can be maintained regarding the so-called *Gupta's challenge*. Referring to one variant of Gupta's puzzle, Gupta says:

> No natural scheme, as far as I know, yields a least-fixed-point theory that is free from problems of this sort [Gupta's puzzle] [10, p. 433].

Formally σ^ω is used *as* a valuation scheme to built up a jump operator, and the least fixed point of such operator handles Gupta's puzzle in the expected way. But σ^ω is not regarded as a valuation scheme for a three-valued *account* for truth. The approach is still *revision-theoretic*. So I acknowledge that Gupta's challenge still holds.

Acknowledgements The application of the *revision-theoretic supervaluation* technique to the theory of self-referential truth was first presented at the 15th Congress on Logic Methodology and Philosophy of Science, held in Helsinki in August 2015. I wish to thank the Department of Mathematics of the University of Torino and, in particular, Matteo Viale, for funding and supporting my participation in this event. I want also remember the audience of this talk and, in particular, Lorenzo Rossi and Johannes Stern for their stimulating questions. A preliminary version of this work was also presented at Barcelona, in December 2015, and at Firenze, in February 2016: I wish to thank José Martinez and Riccardo Bruni, respectively, for inviting me to speak at these two events, as well as Andrea Cantini and all other participants. Finally, let me mention Volker Halbach, for helping me to understand some subtleties of the Tarskian operator over the standard model of arithmetic, and two anonymous referees who helped me with their illuminating comments and suggestions.

A The duality between sieve and jump

In the subsequent Corollary 1, we will prove an order-theoretic proposition about monotone operators on partially ordered sets from which Theorem 1 easily follows.

Let $\mathbb{P} = \langle P, \preceq \rangle$ and $\mathbb{Q} = \langle Q, \preceq \rangle$ be two partially ordered sets. We say that a function $F : P \to Q$ is *antitone* if and only if

$$p \preceq p' \implies F(p') \preceq F(p),$$

for every $p, p' \in P$. We say that a pair (F, G) of functions $F : P \to Q$ and $G : Q \to P$ is a *Galois connection* if and only if

(1) Both F and G are antitone maps.
(2) $p \preceq G(F(p))$ and $q \preceq F(G(q))$, for every $p \in P$ and $q \in Q$.

Given a function $f : P \to P$ and an element $p \in P$, we say that

- p is *f-sound* if and only if $p \preceq f(p)$.

– p is *f-replete* if and only if $f(p) \preceq p$.
– p is a *fixed point* of f if and only if $f(p) = p$.

The set of all fixed points of f will be denoted by $\mathrm{fix}(f)$.

Proposition 1 *[26, thm. 3.16] Let (F, G) be a Galois connection and let $\Gamma = G \circ F$ and $\Lambda = F \circ G$. Then the maps $F : \mathrm{fix}(\Gamma) \to \mathrm{fix}(\Lambda)$ and $G : \mathrm{fix}(\Lambda) \to \mathrm{fix}(\Gamma)$ are inverse bijections. Thus $\mathrm{fix}(\Gamma) = \{G(q) \mid q \in Q\}$ and $\mathrm{fix}(\Lambda) = \{F(p) \mid p \in P\}$.*

Lemma 1 *Let $F : P \to Q$ and $G : Q \to P$ be two antitone maps and let $\Gamma = G \circ F$ and $\Lambda = F \circ G$. Then*

(1) Both Γ and Λ are monotone operators.
(2) $G \circ \Lambda = \Gamma \circ G$ and $F \circ \Gamma = \Lambda \circ F$.
(3) For every $p \in P$ and $q \in Q$: (a) if p is Γ-sound, then $F(p)$ is Λ-replete; (b) if q is Λ-replete, then $G(q)$ is Γ-sound; (c) if p is Γ-replete, then $F(p)$ is Λ-sound; (d) if q is Λ-sound, then $G(q)$ is Γ-replete. In particular, if p is a fixed point of Γ, then $F(p)$ is a fixed point of Λ, and if q is a fixed point of Λ, then $G(q)$ is a fixed point of Γ.
(4) (F, G) is an antitone Galois connection between the sets $\mathrm{fix}(\Gamma)$ and $\mathrm{fix}(\Lambda)$.

Proof (1) Obvious, since both Γ and Λ are compositions of antitone maps. ⊣

(2) $G(\Lambda(q)) = G((F \circ G)(q)) = (G \circ F)(G(q)) = \Gamma(G(q))$ and $F(\Gamma(p)) = F((G \circ F)(p)) = (F \circ G)(F(p)) = \Lambda(F(q))$. ⊣

(3) (a) Let p be Γ-sound, i.e., $p \preceq \Gamma(p)$. Hence, since F is antitone and by (2), $\Lambda(F(p)) = F(\Gamma(p)) \preceq F(p)$, so $F(p)$ is Λ-replete. (b) Similarly, if q is Λ-replete, i.e., $\Lambda(q) \preceq q$ then, since G is antitone and by (2), $G(q) \preceq G(\Lambda(q)) = \Gamma(G(q))$, so $G(q)$ is Γ-sound. (c) and (d) are proved by symmetric arguments.⊣

(4) Both F and G are antitone when restricted to $\mathrm{fix}(\Gamma)$ and $\mathrm{fix}(\Lambda)$, respectively. By (3), $\mathrm{ran}(F \restriction \mathrm{fix}(\Gamma)) \subseteq \mathrm{fix}(\Lambda)$ and $\mathrm{ran}(G \restriction \mathrm{fix}(\Lambda)) \subseteq \mathrm{fix}(\Gamma)$. We have only to show that the defining condition of antitone Galois connection holds.

Let $p \in \mathrm{fix}(\Gamma)$ and $q \in \mathrm{fix}(\Lambda)$. On one direction, assume $p \preceq G(q)$. Hence $q = \Lambda(q) = F(G(q)) \preceq F(p)$. On the other direction, assume $q \preceq F(p)$. Hence $p = \Gamma(p) = G(F(p)) \preceq G(q)$.

Corollary 1 *F and G are inverse bijections between $\mathrm{fix}(\Gamma)$ and $\mathrm{fix}(\Lambda)$. Hence they are order isomorphisms between $\mathrm{fix}(\Gamma)$ and $\mathrm{fix}(\Lambda)^{\mathsf{op}}$, the latter denoting the set $\mathrm{fix}(\Lambda)$ ordered by reversing $\preceq$. In particular, if Γ has a least fixed point $\mathsf{lfp}(\Gamma)$, then $F(\mathsf{lfp}(\Gamma))$ is the greatest fixed point of Λ and, conversely, if Λ has a greatest fixed point $\mathsf{gfp}(\Lambda)$, then $G(\mathsf{gfp}(\Lambda))$ is the least fixed point of Γ.*

Proof By Lemma 1, (F', G') is an antitone Galois connection between $\mathrm{fix}(\Gamma)$ and $\mathrm{fix}(\Lambda)$, where $F' = F \restriction \mathrm{fix}(\Gamma)$ and $G' = G \restriction \mathrm{fix}(\Lambda)$. Moreover, $G' \circ F' = \Gamma \restriction \mathrm{fix}(\Gamma)$ and $F' \circ G' = \Lambda \restriction \mathrm{fix}(\Lambda)$ hence, by Proposition 1, F' and G' are inverse bijections. Since F is antitone, $p \preceq p'$ implies $F(p') \preceq F(p)$ for every $p, p' \in \mathrm{fix}(\Gamma)$. Conversely, $p, p' \in \mathrm{fix}(\Gamma)$ and $F(p') \preceq F(p)$ implies $p = \Gamma(p) =$

$GF(p) \preceq GF(p') = \Gamma(p') = p'$. So, F and G are order isomorphisms between the sets $\mathrm{fix}(\Gamma)$ and $\mathrm{fix}(\Lambda)$, the latter ordered by reversing $\preceq$.

Suppose $\bar{p} = \mathsf{lfp}(\Gamma)$ and let $\bar{q} = F(\bar{p})$. Since $\bar{p} \in \mathrm{fix}(\Gamma)$, $\bar{q} \in \mathrm{fix}(\Lambda)$. Let $q \in \mathrm{fix}(\Lambda)$. Since $G(q) \in \mathrm{fix}(\Gamma)$ and $\bar{p} = \mathsf{lfp}(\Gamma)$ it follows $\bar{p} \preceq G(q)$, hence $q = \Lambda(q) = FG(q) \preceq F(\bar{p}) = \bar{q}$. Thus $\bar{q} = \mathsf{gfp}(\Lambda)$, namely, $F(\mathsf{lfp}(\Gamma)) = \mathsf{gfp}(\Lambda)$. The converse claim that $G(\mathsf{gfp}(\Lambda)) = \mathsf{lfp}(\Lambda)$ follows by a symmetric argument.

Let us now back to the proof of Theorem 1. Consider the set P of all partial hypotheses p and the set Q of all sets of (total) hypotheses $\mathcal{H}$, both ordered by inclusion. Clearly, the two maps $J : p \mapsto \{h \mid p \subseteq h\}$ and $K : \mathcal{H} \mapsto \bigcap\{\tau^\omega(h) \mid h \in \mathcal{H}\}$, defined in Section 3, are antitone maps between P and Q. Moreover, for the sieve Δ it holds $\Delta = J \circ K$, and for the jump σ^ω it holds $\sigma^\omega = K \circ J$. Hence, by Corollary 1, J and K are order isomorphisms between the set of all fixed points of σ^ω and the set of all fixed points of Δ (this latter ordered by reverse inclusion). Because we already know that both $\mathsf{lfp}(\sigma^\omega)$ and $\mathsf{gfp}(\Delta)$ exist, Corollary 1 yields that $\mathsf{lfp}(\sigma^\omega) = K(\mathsf{gfp}(\Delta))$, and that $\mathsf{gfp}(\Delta) = J(\mathsf{lfp}(\sigma^\omega))$. Therefore, Theorem 1 is proved.

B Revision-theoretic and standard supervaluation

We want to prove in this section that (a) for each σ-sound partial interpretation p (i.e., such that $p \subseteq \sigma(p)$) there exists the least fixed point of σ^ω above p, and (b) the least fixed point of σ^ω is exactly the least fixed point of σ^ω above the least fixed point of σ.

For a partial interpretation p and an operator θ on partial interpretations, let $\mathsf{lfp}(\theta, p)$ denote the least fixed point of θ above p (when it exists). Under this notation, our goal becomes to prove the following

Theorem 2 *For each σ-sound partial interpretation p, $\mathsf{lfp}(\sigma^\omega, p)$ exists. Moreover,*

$$\mathsf{lfp}(\sigma^\omega, \mathsf{lfp}(\sigma)) = \mathsf{lfp}(\sigma^\omega).$$

By an easy induction on ω we see that whenever p is σ-sound, p is σ^ω-sound too. Hence $\mathsf{lfp}(\sigma^\omega, p)$ exists. The second part of Theorem 2 will follow from the dual fact that (a) for each fixed point q of σ^ω there exists the greatest fixed point of σ below q, denoted by $\mathsf{gfp}(\sigma, q)$, and (b) the maps $d : p \mapsto \mathsf{lfp}(\sigma^\omega, p)$ and $e : q \mapsto \mathsf{gfp}(\sigma, q)$ form a monotone Galois connection[39] between the sets of all fixed points of σ and σ^ω ordered by inclusion.

To prove the existence of $\mathsf{gfp}(\sigma, q)$ for every $q \in \mathrm{Fix}(\sigma^\omega)$ we define an auxiliary jump operator σ^* as follows:

$$\sigma^*(p) = \bigcap\{\bigcap \tau^\frown(h) \mid p \subseteq h\}.$$

It is not difficult to check the following key properties of σ^*:

[39] We say that a pair (F, G) of functions $F : P \to Q$ and $G : Q \to P$ between two partially ordered sets $\mathbb{P}$ and $\mathbb{Q}$ is a *monotone* Galois connection if and only if (1) Both F and G are monotone maps; and (2) $p \preceq G(F(p))$ and $F(G(q)) \preceq q$, for every $p \in P$ and $q \in Q$.

1. Each σ^ω-replete q is σ^*-replete too, namely, $\sigma^\omega(q) \subseteq q$ implies $\sigma^*(q) \subseteq q$.
2. $\mathrm{Fix}(\sigma^*) = \mathrm{Fix}(\sigma)$.

From these two properties immediately follows that for each fixed point q of σ^ω (which obviously is also σ^ω-replete), by (1) there exists the greatest fixed point of σ^* below q which, by (2), coincides with $\mathsf{gfp}(\sigma, q)$.

Lemma 2 *Let* $d : p \mapsto \mathsf{lfp}(\sigma^\omega, p)$ *and* $e : q \mapsto \mathsf{gfp}(\sigma, q)$*. Then* (d, e) *is a monotone Galois connection between* $\mathrm{Fix}(\sigma)$ *and* $\mathrm{Fix}(\sigma^\omega)$*, both ordered by inclusion.*

Proof Clearly, both d and e are monotone maps. Let $p \in \mathrm{Fix}(\sigma)$ and $q \in \mathrm{Fix}(\sigma^\omega)$. By definition of e, $e(d(p))$ is the largest $p' \in \mathrm{Fix}(\sigma)$ below $d(p)$. Since $p \subseteq d(p)$ and $p \in \mathrm{Fix}(\sigma)$, it follows $p \subseteq e(d(p))$. By definition of d, $d(e(q))$ is the least $q' \in \mathrm{Fix}(\sigma^\omega)$ above $e(q)$. Since $e(q) \subseteq q$ and $q \in \mathrm{Fix}(\sigma^\omega)$, it follows $e(d(q)) \subseteq q$.

We are now ready to prove the second part of Theorem 2, namely that $\mathsf{lfp}(\sigma^\omega) = \mathsf{lfp}(\sigma^\omega, \mathsf{lfp}(\sigma))$. Let $\bar{p} = \mathsf{lfp}(\sigma)$ and $\bar{q} = \mathsf{lfp}(\sigma^\omega)$. By definition of $\bar{p}$ and e, $\bar{p} \subseteq e(\bar{q}) \subseteq \bar{q}$. By definition of $\bar{q}$ and d, and by Lemma 2, $\bar{q} \subseteq d(\bar{p}) \subseteq d(e(\bar{q})) \subseteq \bar{q}$. Hence $\bar{q} = d(\bar{p})$, i.e., $\mathsf{lfp}(\sigma^\omega) = \mathsf{lfp}(\sigma^\omega, \mathsf{lfp}(\sigma))$.

C Complexity of revision-theoretic supervaluation

Theorem 3 *The set* V *of all (Gödel codes of) sentences declared true by the least fixed point of revision-theoretic supervaluation is* Π^1_1*.*

For the purposes of this section, let us back to the official definition of τ as an operator on subsets of ω. Let $\tau^\frown(X) = \{\tau^n(X) \mid n \in \omega\}$ denote the trajectory of τ starting with X. Accordingly, we have the following definitions:

- $\mathsf{stab}^+(X) = \{k \in \omega \mid \exists m \, \forall n \geq m \, (k \in \tau^n(X))\}$.
- $\mathsf{stab}^-(X) = \{k \in \omega \mid \exists m \, \forall n \geq m \, (k \notin \tau^n(X))\}$.
- $\Theta^+(X^+, X^-) = \bigcap\{\mathsf{stab}^+(X) \mid X^+ \subseteq X \;\;\&\;\; X \cap X^- = \emptyset\}$.
- $\Theta^-(X^+, X^-) = \bigcap\{\mathsf{stab}^-(X) \mid X^+ \subseteq X \;\;\&\;\; X \cap X^- = \emptyset\}$.
- $\Theta(X^+, X^-) = (\Theta^+(X^+, X^-), \Theta^-(X^+, X^-))$.

It is straightforward to see that

1. Θ is a monotone operator on partial interpretations.
2. If (Z^+, Z^-) denotes the least fixed point of Θ, then $Z^+ = \mathsf{V}$.

Moreover, it is clear that the definition of Θ fits the same template of the definition of the supervaluational jump operator J_{vF} in Burgess [3, p. 666], with the Tarskian operator J_{T} (Burgess' notation for our τ) and its complement just replaced by the operators stab^+ and stab^- defined above. It is well known that the relation $\{(X, n) \mid n \in J_{\mathsf{T}}(X)\}$ and its complement are Δ^1_1. Therefore, by mimicking Burgess' computation of the complexity of 0^+_{vF} (Burgess' notation for V^σ) in [3, p. 670], if we can show that both relations $\{(X, n) \mid n \in \mathsf{stab}^+(X)\}$

and $\{(X,n) \mid n \in \mathsf{stab}^-(X)\}$ are Δ^1_1, then the upper bound of the complexity of V will result to be Π^1_1 as well the upper bound of the complexity of V^σ.

To prove that the relations $\{(X,n) \mid n \in \mathsf{stab}^+(X)\}$ and $\{(X,n) \mid n \in \mathsf{stab}^-(X)\}$ are Δ^1_1 all we need to show is that the relation $\{(k,n,X) \mid k \in \tau^n(X)\}$ is Δ^1_1.

In order to emphasise the generality of the result we replace the Tarskian operator τ with a generic operator Ψ on subsets of ω and prove the following

Proposition 2 *Let Ψ be an operator on subsets of ω. If Ψ is Δ^1_1, so is its iteration $\Psi^\frown$.*

We will prove Proposition 2 by a series of lemmata and remarks.

Let $\mathcal{P}(A)$ denote the set of all subsets of some set A. In general, we can code any function $f : B \to \mathcal{P}(A)$ by a binary relation $R \subseteq B \times A$, by putting

$$R(y,x) \iff x \in f(y),$$

for all $y \in B$ and $x \in A$. Indeed, for all $y \in B$, we have

$$f(y) = \{x \in A \mid R(y,x)\}.$$

We call R the *relation associated to* f.

Lemma 3 *Let $\Psi : \mathcal{P}(A) \to \mathcal{P}(A)$. For all $S \subseteq \omega \times A$ and $\forall i,j \in \omega$, the following are equivalent:*

- *(a)* $\exists Y\,(\forall z\,(S(i,z) \iff z \in Y)\ \&\ \forall w\,(S(j,w) \iff w \in \Psi(Y)))$.
- *(b)* $\forall Y\,(\forall z\,(S(i,z) \iff z \in Y) \implies \forall w\,(S(j,w) \iff w \in \Psi(Y)))$.

Proof (a) $\implies$ (b). Let $Y \subseteq A$ be such that $\forall z\,(S(i,z) \iff z \in Y)$ holds and let Y' be a subset of A satisfying (a). Then, $\forall z\,(z \in Y' \iff S(i,z) \iff z \in Y)$, so $Y = Y'$. Thus $\forall w\,(S(j,w) \iff w \in \Psi(Y))$ follows.

(b) $\implies$ (a). Assume (b) and define $Y = \{z \in A \mid S(i,z)\}$. Then $\forall z\,(S(i,z) \iff z \in Y)$ holds, so, by (b), also $\forall w\,(S(j,w) \iff w \in \Psi(Y)))$ holds.

We denote by $R(i,j,S)$ the relation equivalently defined by either the condition (a) or (b) of Lemma 3. Clearly, if S is the relation associated to the sequence $s : \omega \to \mathcal{P}(A)$, then $R(i,j,S)$ holds if and only if $s(j) = \Psi(s(i))$, for all $i,j \in \omega$.

Remark 1 Let S be the relation associated to the sequence $s : \omega \to \mathcal{P}(A)$. Then the relation

$$R'(i,j,S) \iff \Psi(s(i)) \subseteq s(j),$$

admits the following equivalent definitions

- (a) $\exists Y\,(\forall z\,(S(i,z) \iff z \in Y)\ \&\ \forall w\,(w \in \Psi(Y) \implies S(j,w)))$.
- (b) $\forall Y\,(\forall z\,(S(i,z) \iff z \in Y) \implies \forall w\,(w \in \Psi(Y) \implies S(j,w)))$.

Lemma 4 *We can give two equivalent explicit definitions of the relation $x \in \Psi^n(X)$ associated to the trajectory $\Psi^\frown(X)$ of X in Ψ as follows:*

1. $Q(n,x) \iff \exists S\,(\forall y\,(S(0,y) \iff y \in X)\ \&\ \forall i < n\, R(i,i+1,S)\ \&\ S(n,x))$.
2. $Q'(n,x) \iff \forall Y\,(Y = \Psi^n(X) \implies x \in Y)$.

Proof (1) We will show, by induction on n, that $\forall x \in A\,(Q(n,x) \iff x \in \Psi^n(X))$.

Let $n = 0$. Suppose $x \in \Psi^0(X) = X$ and define $S = \{\langle k,y\rangle \mid k = 0\ \&\ y \in X\}$. Then, $\forall y\,(S(0,y) \iff y \in X$ holds by definition, $\forall i < 0\, R(i,i+1,S)$ vacuously holds and $S(0,x)$ holds, since $x \in X$. Conversely, suppose there exists S such that $\forall y\,(S(0,y) \iff y \in X$ and $S(0,x)$. Then $x \in X = \Psi^0(X)$.

Let $n = m+1$. Suppose $x \in \Psi^{m+1}(X) = \Psi(\Psi^m(X))$. By the inductive hypothesis, there exists S' such that $\forall y \in A\,(Q(m,y) \iff y \in \Psi^m(X))$. Define $Y = \{z \mid S'(m,z)\}$ and $S = S' \restriction (m+1) \cup \{\langle m+1, z\rangle \mid z \in \Psi(Y)\}$. Then, $\forall y\,(S(0,y) \iff y \in X)$ and $R(i,i+1,S)$ for all $i < m$ hold since $S' \restriction (m+1) \subseteq S$. Let $i = m$. Then, by definition of Y and S, $\forall z\,(S(m,z) \iff S'(m,z) \iff z \in Y)\ \&\ \forall w\,(S(m+1,w) \iff w \in \Psi(Y))$, hence $R(m,m+1,S)$ holds. By the inductive hypothesis, $\forall z\,(z \in Y \iff S'(m,z) \iff z \in \Psi^m(X))$, hence $Y = \Psi^m(X)$. Thus $x \in \Psi^{m+1}(X) \implies x \in \Psi(\Psi^m(X)) \implies x \in \Psi(Y) \implies S(m+1,x)$.

Suppose, conversely, that there exists $S \subseteq \omega \times A$ such that $\forall y\,(S(0,y) \iff y \in X)\ \&\ \forall i < m+1\, R(i,i+1,S)\ \&\ S(m+1,x)$ holds.

Claim: Let $P(T,k)$ be the condition $\forall y\,(T(0,y) \iff y \in X)\ \&\ \forall i < k\, R(i,i+1,T)$. Then, $\forall T' T'\,(P(T,k)\ \&\ P(T',k) \implies \forall i \le k\, \forall y\,(T(i,y) \iff T'(i,y)))$.

Proof of the claim. By induction on k. If $k = 0$, then $\forall y\,(T(0,y) \iff y \in X \iff T'(0,y))$. Let $k = j+1$. By the inductive hypothesis, $\forall u\,(T(j,u) \iff T'(j,u))$. Hence, from $R(j,j+1,T)$ and $R(j,j+1,T')$ it follows:

$\exists Y\,(\forall z\,(T(j,z) \iff z \in Y)\ \&\ \forall w\,(T(j+1,w) \iff w \in \Psi(Y)))$ and $\exists Y'\,(\forall z\,(T'(j,z) \iff z \in Y')\ \&\ \forall w\,(T'(j+1,w) \iff w \in \Psi(Y')))$. Hence $\forall z\,(z \in Y \iff T(j,z) \iff T'(j,z) \iff z \in Y')$, so $Y = Y'$. Thus $\forall w\,(T(j+1,w) \iff w \in \Psi(Y) = \Psi(Y') \iff T'(j+1,w))$. ⊣

By the claim, we can prove that $\exists S(P(S,m+1)\ \&\ S(m+1,x) \implies x \in \Psi^{m+1}(X))$. By the hypothesis, there exists Y such that $\forall z\,(S(m,z) \iff z \in Y)\ \&\ \forall w\,(S(m+1,w) \iff w \in \Psi(Y))$. Since $S(m+1,x)$ holds, it follows $x \in \Psi(Y)$. Thus, it remains to show that $Y = \Psi^m(X)$. If $y \in Y$, then $S(m,Y)$. Since $P(S,m)$ holds, by the inductive hypothesis $y \in \Psi^m(X)$. Conversely, let $y \in \Psi^m(X)$. By the inductive hypothesis, $\exists S'\,(P(S',m)\ \&\ S'(m,y))$. By the claim, $\forall z\,(S(m,z) \iff S'(m,z))$, so $S(m,y)$ holds, hence $y \in Y$. Since $Y = \Psi^m(X)$, $x \in \Psi(Y) = \Psi(\Psi^m(X)) = \Psi^{m+1}(X)$. ⊣

(2) Immediate from the definition.

Remark 2 The graph $Q = \{\langle n,y\rangle \mid y = f^\frown(x)(n)\} = \{\langle n,y\rangle \mid y = f^n(x)\}$ of the trajectory $f^\frown(x)$ admits the following *inductive definition*:

(a) $\langle 0,x\rangle \in Q$,
(b) if $\langle m,z\rangle \in Q$ then $\langle m+1, f(z)\rangle \in Q$.
(c) Q is the intersection of all relations satisfying (1) and (2)

which can be converted into the following explicit definition:

$$\langle n, y\rangle \in Q \iff \forall R\,((\langle 0, x\rangle \in R \;\&\; \forall m, z\,(\langle m, z\rangle \in R \implies \langle m{+}1, f(z)\rangle \in R)) \implies \langle n, y\rangle \in R).$$

Proof of Proposition 2 More precisely, we will prove the following statement. Let $D \subseteq \mathcal{P}(\omega) \times \omega$ and let $\Psi(Y) = \{n \in \omega \mid D(Y, n)\}$ for all $Y \subseteq \omega$. Let $Q \subseteq \mathcal{P}(\omega) \times \omega^2$ be defined by

$$Q(X, n, k) \iff k \in \Psi^n(X),$$

for all $n, k \in \omega$ and $X \subseteq \omega$. If D is Δ^1_1, so is Q.

Given a relation $S \subseteq \omega \times \omega$, define $s : \omega \to \mathcal{P}(\omega)$ by $s(i) = \{j \in \omega \mid S(i, j)\}$, for all $i \in \omega$.

Let $R \subseteq \omega \times \omega \times {}^{\omega\times\omega}2$ be the relation

$$R(i, j, S) \iff s(j) = \Psi(s(i)).$$

Since $n \in \Psi(Y) \iff D(Y, n)$ and D is Δ^1_1, then, by Lemma 3, R is Δ^1_1. Let $P \subseteq {}^{\omega\times\omega}2 \times \omega \times \mathcal{P}(\omega)$ be the relation

$$P(S, k, X) \iff \forall j\,(S(0, j) \iff j \in X) \;\&\; \forall i < k\, R(i, i+1, S).$$

The relation P is Δ^1_1 since R is. By Lemma 4, $Q(X, n, k)$ holds if and only if $\exists S\,(P(S, k, X) \;\&\; S(n, k))$ holds. Hence Q is Σ^1_1.

On the other hand, by Lemma 4, $Q(X, n, k)$ holds if and only if $\forall Y\,(Y = \Psi^n(X) \implies k \in Y)$. By Remark 2,

$$Y = \Psi^n(X) \iff \forall R\,((\langle 0, X\rangle \in R \;\&\; \forall m, Z\,(\langle m, Z\rangle \in R \implies \langle m+1, \Psi(Z)\rangle \in R)) \implies \langle n, Y\rangle \in R).$$

Hence $Y = \Psi^n(X)$ is Π^1_1, so also $Q(X, n, k)$ is Π^1_1, and thus Δ^1_1.

Since the relation $Q(X, n, k)$ is Δ^1_1, also the relation $Y = \Psi^n(X)$ is Δ^1_1. For, $Y = \Psi^n(X) \iff \forall z\,(z \in Y \iff Q(X, n, z))$. □

Now we can complete the proof of Theorem 3 as follows. Since τ is a Δ^1_1 operator on subsets of ω, by Proposition 2 the ternary relations $k \in \tau^n(X)$ and $k \notin \tau^n(X)$ are Δ^1_1. It follows that the binary relations $n \in \mathsf{stab}^+(X)$ and $n \in \mathsf{stab}^-(X)$ are Δ^1_1 too. Hence, as observed at the beginning of this section, by mimicking Burgess [3, p. 670], the set $Z^+ = \mathsf{V}$ is Π^1_1.

Question 1 Is V a Π^1_1-complete set?[40]

[40] An anonymous referee suggested me look at a recent paper by Thomas Schindler [27] to answer this question. Schindler's Proposition 8 shows that whenever a valuation scheme is nice then the set of (the codes of) all sentences declared true by its least fixed point is Π^1_1-hard. A valuation scheme is *nice* when it is mononotonic and satisfies eight conditions labelled in Schindler's paper by V1-V5 and N1-N3. Actually, σ^ω is not a nice scheme, because it fails to satisfies the condition V2, namely, in our notation: $\mathsf{T}t \in \mathsf{Dom}(\sigma^\omega(p)) \Leftrightarrow$

D Revision-theoretic supervaluation and standard revision

Theorem 4 $\mathsf{lfp}(\sigma^\omega)$ *is compatible with* $\mathsf{stab}^\#$.

We will show that $\mathsf{lfp}(\sigma^\omega)$ and $\mathsf{stab}^\#$ are compatible as partial interpretations (namely, that they agree on the common part of their respective domains) by proving that they are compatible in the order-theoretic sense, namely that there exists a partial interpretation which extends both. This partial interpretation will be defined as the set of stabilities of a variant of revision based (as well as $\mathsf{stab}^\#$) on the concept of near stability.

The notion of revision sequence was defined by imposing on each limit stage of the sequence a coherence condition stated in terms of the notion of stability: h coheres with $S \upharpoonright \beta$ if and only if $\mathsf{stab}(S \upharpoonright \beta) \subseteq h_\beta$. We can do the same replacing the notion of stability with the notion of near stability:

Definition 2 A *revision*$^\#$ *sequence* $S = \langle h_\alpha \mid \alpha \in \mathsf{On}\rangle$ is an ordinal-length sequence of hypotheses satisfying the following two requirements:

1. $h_{\alpha+1} = \tau(h_\alpha)$, for every $\alpha \in \mathsf{On}$.
2. $\mathsf{stab}^\#(S \upharpoonright \beta) \subseteq h_\beta$, for every β limit.

The partial interpretation yielded by using near stability in the above sense is the set

$$\mathsf{stab}^{\#\#} = \bigcap\{\mathsf{stab}^\#(S) \mid S \text{ is a revision}^\# \text{ sequence}\}$$

It is straightforward to see that, for every sequence S, $\mathsf{stab}(S) \subseteq \mathsf{stab}^\#(S)$, hence every revision$^\#$ sequence is also a revision sequence and $\mathsf{stab}^\# \subseteq \mathsf{stab}^{\#\#}$.

To prove that also $\mathsf{lfp}(\sigma^\omega) \subseteq \mathsf{stab}^{\#\#}$, we will reformulate the partial interpretation $\mathsf{stab}^{\#\#}$ as the least fixed point of a suitable monotone operator:

Definition 3 $\sigma^{\#\#}$ denotes the monotone operator defined by

$$\sigma^{\#\#}(p) = \bigcap\{\mathsf{stab}^\#(S) \mid p \subseteq S(0) \;\;\&\;\; S \text{ is a revision}^\# \text{ sequence}\}.$$

Obviously, $\sigma^{\#\#}(\emptyset) = \mathsf{stab}^{\#\#}$.

Given a revision sequence S, let $\mathsf{Cf}(S)$ denote the set of all hypotheses which occur cofinally many times in S. A key feature of near stability we will use in the following is that

$$\mathsf{stab}^\#(S) = \bigcap\{\tau^\omega(h) \mid h \in \mathsf{Cf}(S)\}.$$

Lemma 5 $\mathsf{stab}^{\#\#} = \mathsf{lfp}(\sigma^{\#\#})$.

$t^{\mathbb{N}} \in \mathsf{Dom}(p)$ and $\sigma^\omega(p)(\mathsf{T}t) = p(t^{\mathbb{N}})$, for every partial interpretation p. Schindler [27, p. 462] remarks that, in order to prove Proposition 8, "it is sufficient to assume that [the conditions] hold merely for those partial models that arise in the construction of the least fixed point". However, for σ^ω, the left-to-right direction of the condition V2 fails even for the empty interpretation (a counterexample is given, for instance, by $t = \ulcorner 0 = 0 \urcorner$), so neither Proposition 8 nor its proof can be directly applied to the valuation scheme σ^ω.

Proof Let $\bar{p} = \sigma^{\#\#}(\emptyset)$. By monotonicity, $\bar{p} \subseteq \sigma^{\#\#}(\bar{p})$.

Conversely, let $\langle \phi, v \rangle \in \sigma^{\#\#}(\bar{p})$. Hence $\langle \phi, v \rangle \in \mathsf{stab}^{\#}(S)$ for any revision$^{\#}$ sequence S such that $\bar{p} \subseteq S(0)$. For any sequence S and any ordinal $\alpha < \mathsf{lh}(S)$, let S^{α} denote the *final segment* of S at α, namely the sequence S' defined by

1. $\mathsf{lh}(S') = \mathsf{lh}(S) - \alpha$.
2. $S'(\xi) = S(\alpha + \xi)$, for every $\xi < \mathsf{lh}(S')$.

Let S be any revision$^{\#}$ sequence. Since S is ordinal-length, there exists a limit ordinal γ such that $\mathsf{Cf}(S) = \mathrm{ran}(S^{\gamma})$: let $g = S(\gamma)$. Since $\bar{p} \subseteq \mathsf{stab}^{\#}(S) = \bigcap\{\tau^{\omega}(h) \mid h \in \mathsf{Cf}(S)\}$ and $g \in \mathsf{Cf}(S)$, it follows $\bar{p} \subseteq \tau^{\omega}(g)$. Let $\delta = \gamma + \omega$. Since $\mathsf{stab}^{\#}(S \restriction \delta) = \tau^{\omega}(g)$, by the coherence condition $\bar{p} \subseteq \tau^{\omega}(g) \subseteq S(\delta)$. Since S^{δ} is a revision$^{\#}$ sequence and $\bar{p} \subseteq S^{\delta}(0) = S(\delta)$, it follows that $\langle \phi, v \rangle \in \mathsf{stab}^{\#}(S^{\delta}) = \mathsf{stab}^{\#}(S)$. Hence, we have showed that for every revision$^{\#}$ sequence S, $\langle \phi, v \rangle \in \mathsf{stab}^{\#}(S)$. Therefore $\langle \phi, v \rangle \in \bar{p}$, thus $\sigma^{\#\#}(\bar{p}) \subseteq \bar{p}$.

It follows that $\bar{p} = \sigma^{\#\#}(\emptyset) = \mathsf{stab}^{\#\#}$ is a fixed point of $\sigma^{\#\#}$. Let q be any fixed point of $\sigma^{\#\#}$. By monotonicity, $\mathsf{stab}^{\#\#} = \sigma^{\#\#}(\emptyset) \subseteq \sigma^{\#\#}(q) = q$, hence $\mathsf{stab}^{\#\#} = \mathsf{lfp}(\sigma^{\#\#})$.

Lemma 6 *$\sigma^{\omega}(p) \subseteq \sigma^{\#\#}(p)$, for every σ^{ω}-sound p. In particular, every σ^{ω}-sound p is also sound for $\sigma^{\#\#}$.*

Proof Let p be σ^{ω}-sound and let S be a revision$^{\#}$ sequence such that $p \subseteq S(0)$. We will show, by induction on γ, that $p \subseteq S(\gamma)$ for every γ limit.

For $\gamma = \omega$, $p \subseteq \sigma^{\omega}(p) \subseteq \tau^{\omega}(S(0)) = \mathsf{stab}^{\#}(S \restriction \omega) \subseteq S(\omega)$. Let $\gamma = \delta + \omega$. By the inductive hypothesis, $p \subseteq S(\delta)$. Hence, by definition of σ^{ω}, $p \subseteq \sigma^{\omega}(p) \subseteq \tau^{\omega}(S(\delta)) = \mathsf{stab}^{\#}(S \restriction \gamma) \subseteq S(\gamma)$. Finally, let γ be a limit of limits. By the inductive hypothesis, $p \subseteq S(\delta)$ for every $\delta \in \gamma \cap \mathrm{Lim}$. Hence $p \subseteq \sigma^{\omega}(p) \subseteq \tau^{\omega}(S(\delta))$ for every $\delta \in \gamma \cap \mathrm{Lim}$. Hence $p \subseteq \sigma^{\omega}(p) \subseteq \bigcap\{\tau^{\omega}(S(\delta)) \mid \delta \in \gamma \cap \mathrm{Lim}\} \subseteq \liminf\{\tau^{\omega}(S(\delta)) \mid \delta \in \gamma \cap \mathrm{Lim}\} = \mathsf{stab}^{\#}(S \restriction \gamma) \subseteq S(\gamma)$.

Since $p \subseteq S(\gamma)$ for every γ limit, it follows $\sigma^{\omega}(p) \subseteq \bigcap\{\tau^{\omega}(S(\gamma)) \mid \gamma \in \mathrm{Lim}\} \subseteq \liminf\{\tau^{\omega}(S(\gamma)) \mid \gamma \in \mathrm{Lim}\} = \mathsf{stab}^{\#}(S)$. Hence $p \subseteq \sigma^{\omega}(p) \subseteq \sigma^{\#\#}(p)$.

Finally, we can conclude the proof of Theorem 4 as follows. From Lemma 6 it follows, by transfinite induction, that $\mathsf{lfp}(\sigma^{\omega}) \subseteq \mathsf{lfp}(\sigma^{\#\#})$. Hence, by Lemma 5, $\mathsf{lfp}(\sigma^{\omega}) \subseteq \mathsf{stab}^{\#\#}$. We already saw that $\mathsf{stab}^{\#} \subseteq \mathsf{stab}^{\#\#}$ too. Hence $\mathsf{stab}^{\#\#}$ is a common upper bound of both $\mathsf{lfp}(\sigma^{\omega})$ and $\mathsf{stab}^{\#}$ which, therefore, are order-theoretically compatible.

References

1. Antonelli, G. A. *The Complexity of Revision.* Notre Dame Journal of Formal Logic, Vol. 35(1), 67–72 (1994).
2. Belnap, N. *Gupta's Rule of Revision Theory of Truth.* Journal of Philosophical Logic, Vol. 11(1), 103–116 (1982).
3. Burgess, J. P. *The Truth is Never Simple.* The Journal of Symbolic Logic, Vol. 51, 663–681 (1986).

4. Chapuis, A. *Alternative revision theories of truth.* Journal of Philosophical Logic, Vol. 25, 399–423 (1996).
5. Fischer, M. et al. *Axiomatizing semantic theories of truth?* The Review of Symbolic Logic, Vol. 8(2), 257–278 (2015).
6. Fitting, M. *Notes on the mathematical aspects of Kripke's theory of truth.* Notre Dame Journal of Formal Logic, Vol. 27, 75–88 (1986).
7. Gupta, A. *Truth and Paradox.* Journal of Philosophical Logic, Vol. 11(1), 1–60 (1982).
8. Gupta, A. *The Meaning of Truth.* In Lepore, E. (ed.), *New Directions in Semantics*, 453–480 (1987).
9. Gupta, A. *Remarks on Definitions and the Concept of Truth.* Proceedings of the Aristotelian Society, Vol. 89, 227–246 (1989).
10. Gupta, A. *Definition and Revision: A Response to McGee and Martin.* Philosophical Issues, Vol. 8, 419–443 (1997).
11. Gupta, A. and Belnap, N. *The Revision Theory of Truth.* A Bradford Book. MIT Press, Cambridge, MA (1993).
12. Halbach, V *Axiomatic Theories of Truth.* Cambridge University Press, Cambridge, UK (2011).
13. Hansen, C. S. *Supervaluation on trees for Kripke's theory of truth.* The Review of Symbolic Logic, Vol. 8, 46–74 (2015).
14. Hellman, G. *Reviews of Gupta, 1982 and Herzberger, 1982.* The Journal of Symbolic Logic, Vol. 50(4), 1068–1071 (1985).
15. Herzberger, H. G. *Notes on Naive Semantics.* Journal of Philosophical Logic, Vol. 11(1), 61–102 (1982).
16. Horsten, L. and Halbach, V. *Truth and Paradox.* In Horsten, L. and Pettigrew R. (eds.), *The Bloomsbury Companion to Philosophical Logic*, Bloomsbury Publishing (2014).
17. Kremer, P. *The Gupta-Belnap Systems* $\mathsf{S}^{\#}$ *and* S^{*} *are not Axiomatisable.* Notre Dame Journal of Formal Logic, Vol. 34, 583–596 (1993).
18. Kremer, P. *Comparing Fixed-Point and Revision Theories of Truth.* Journal of Philosophical Logic, Vol. 38, 363–403 (2009).
19. Kripke, S. *Outline of a theory of truth.* Journal of Philosophy, Vol. 72, 690–716 (1975).
20. Löwe, B. *Revision Forever!.* In Schärfe, H. et al. (eds.), *Conceptual Structures: Inspiration and Application, 14th International Conference on Conceptual Structures, ICCS 2006, Aalborg, Denmark, July 16–21*, Lecture Notes in Artificial Intelligence, 4068, 22–36 (2006).
21. Löwe, B. and Welch, P. *Set-Theoretic Absoluteness and the Revision Theory of Truth.* Studia Logica, Vol. 68(1), 21–41 (2001).
22. Martin, D. *Revision and its rivals.* Philosophical Issues, Vol. 8, 407–418 (1997).
23. McGee, V. *Revision.* Philosophical Issues, Vol. 8, 387–406 (1997).
24. Meadows, T. *Truth, Dependence and Supervaluation: Living with the Ghost.* Journal of Philosophical Logic, 42, 221–240 (2013).
25. Rivello, E. *Revision without revision sequences: Circular definitions.* Journal of Philosophical Logic, to appear (2018).

26. Roman, S. *Lattices and Ordered Sets*. Springer (2008).
27. Schindler, T. *Some notes on truth and comprehension*. The Journal of Philosophical Logic (forthcoming).
28. Shapiro, S. *The rationale behind revision-rule semantics*. Philosophical studies, Vol. 129, 477–515 (2006).
29. Sheard, M. *Truth, Provability and Naive Criteria*. In Halbach V. and Horsten L. (eds.), *Principles of Truth*, 169–182, Dr Hänsel, Hohenhausen (2002).
30. Visser, A. *Semantics and the Liar Paradox*. In Gabbay D. M. and Günthner F. (eds.), *Handbook of Philosophical Logic*, IV, 617–706, Reidel, Dordrecht (1989).
31. Visser, A. *Semantics and the Liar Paradox*. In Gabbay D. M. and Günthner F. (eds.), *Handbook of Philosophical Logic*, 11, (2nd ed. of [30]), Kluwer Academic Publishers, The Netherlands (2004).
32. Welch, P. *On Gupta-Belnap Revision Theories of Truth, Kripkean Fixed Points, and the Next Stable Set*. The Bulletin of Symbolic Logic, Vol. 7(3), 345–360 (2001).
33. Yaqub, A. M. *The Liar Speaks the Truth. A Defense of the Revision Theory of Truth*. Oxford University Press, Oxford (1993).