THE REVIEW OF SYMBOLIC LOGIC
Volume 19, Number 3, September 2026

REVISION THEORY WITHOUT ω-INCONSISTENCY

JOHN SCHINDLER
Department of Philosophy, University of Pittsburgh

Abstract. An apparent issue for the Revision Theory of definitions has long been that its most plausible versions engender ω-inconsistencies. In this paper I develop a new ω-consistent revision theory and use it to argue that revision theorists can and should embrace ω-consistency. I show how my theory, called $\mathbf{S}^{\#N}$, withstands the theoretical pressures towards ω-inconsistency and moreover compares favorably to the best ω-inconsistent theories vis-à-vis several important desiderata. I tentatively conclude that $\mathbf{S}^{\#N}$ is the best known revision theory.

The Revision Theory grows out of the idea that circular definitions can be legitimate. Revision theorists have claimed that ordinary language expresses circularly defined concepts, including truth, rational choice, and property exemplification.[1] The heart of the Revision Theory consists in providing a fully general semantics and logic for circular definitions. As presented in its *locus classicus*, Gupta and Belnap [14] (henceforth RTT), the Revision Theory is not just a single theory of definitions, but a genus of semantical theories sharing a common philosophical motivation and technical apparatus. (I will refer to the individual species as 'revision theories', in lowercase.) Of the known revision theories, arguably the most attractive have a curious feature: they are *ω-inconsistent*, in the sense that, relative to certain legitimate definitions, ω-inconsistencies become categorically true in the standard model of arithmetic.[2] These ω-inconsistencies are not restricted to artificial definitions: even the circular definition of truth produces ω-inconsistencies in languages with sufficiently expressive syntactic resources (see §7). Gupta [12] goes so far as to argue that ω-inconsistency is non-optional: the Revision Theory—and indeed any logic of definitions—must tolerate it. To firmly committed revision theorists, true ω-inconsistencies may present themselves as a novel logical discovery. To other philosophers, though, they may look more like a thorn in the side of the Revision Theory.

Received: February 6, 2026.
2020 *Mathematics Subject Classification*: Primary 03A05, Secondary 03F03, 03B65.
Key words and phrases: revision theory, truth, circular definitions, omega-inconsistency.

[1] For application of the Revision Theory to truth, see Gupta [9], Herzberger [17], Gupta and Belnap [14], Yaqub [29], and Gupta [13]; to rational choice, see Chapuis [4], Gupta [10], Chapuis [5], and Bruni and Sillari [2]; and to property exemplification, see Orilia [21].

[2] Recall that an *ω-inconsistency* is a set of sentences $\{\exists x \neg A(x), A(0), A(0'), \ldots\}$, for some formula A. The key semantical notion of categorical truth will be explained in §1.2. §2 illustrates the sense in which the standard model could yet give rise to ω-inconsistencies.

© The Author(s), 2026. Published by Cambridge University Press on behalf of The Association for Symbolic Logic. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use.

doi:10.1017/S1755020326101142

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

Is the Revision Theory indeed committed to ω-inconsistency? I argue no. In this paper I develop a collection of new ω-consistent semantical and proof systems, and use them to articulate a new revision theory that I call $\mathbf{S}^{\#N}$. I then use $\mathbf{S}^{\#N}$ to show why the Revision Theory can and should remain ω-consistent. I explain how $\mathbf{S}^{\#N}$, unlike previous revision theories, withstands Gupta's [12, 13] arguments for ω-inconsistency, which claim that only ω-inconsistent systems have viable proof systems and can preserve the expressive functions of truth. Beyond meeting these challenges for ω-consistency, I argue that $\mathbf{S}^{\#N}$ compares favorably to the best ω-inconsistent theories vis-à-vis a variety of important desiderata. I tentatively conclude that $\mathbf{S}^{\#N}$ is the best known revision theory.

As my exemplar of an ω-inconsistent theory I will take $\mathbf{S}^{\#}$, the stronger of the two ω-inconsistent theories recently defended by Gupta [12, 13]. The other, $\mathbf{S}$, differs from $\mathbf{S}^{\#}$ only in small details. On the whole my considerations remain the same when comparing to $\mathbf{S}$; for the relevant differences see fnn. 22 and 27. My comparison revolves around nine desiderata for revision theories, which I highlight below. I will argue that $\mathbf{S}^{\#N}$ fares no worse than $\mathbf{S}^{\#}$ on each of these considerations, and strictly better than $\mathbf{S}^{\#}$ on some. This means that I need not establish how much weight to give to each desideratum—only that it contributes something to a theory's plausibility—in order to conclude that my theory is preferable to $\mathbf{S}^{\#}$.[3]

To clarify my claims, we should make a further terminological distinction. Besides distinguishing between the Revision Theory (uppercase) and specific revision theories (lowercase), we should also distinguish revision *theories* from revision *systems*. A revision system is a mathematical object given by a definition of categorical truth in that system. We can identify the revision system $\mathbf{S}^{\#}$, for instance, with the relation $M \models^{\mathcal{D}}_{\#} A$ that holds between a model M, set of definitions $\mathcal{D}$, and a sentence A iff A is categorically true in $\mathbf{S}^{\#}$ in M relative to $\mathcal{D}$. Importantly, the relation $\models^{\mathcal{D}}_{\#}$, as a mere mathematical object, cannot be correct or incorrect—any more than the empty set is correct or incorrect.

A revision theory, on the other hand, uses revision systems to make genuine claims about the semantics of languages with circular definitions. The theory associated with the system $\mathbf{S}^{\#}$ says, for instance: "a sentence A of a language $\mathcal{L}^{+}$ interpreted according to the ground model M and supplemented with definitions $\mathcal{D}$ is categorically true iff $M \models^{\mathcal{D}}_{\#} A$." Such a claim can be correct or incorrect. (Indeed, this paper argues the latter.) In the case of $\mathbf{S}^{\#}$ one can directly read off the revision theory from the corresponding revision system. Matters are less simple for my alternative theory $\mathbf{S}^{\#N}$, which is given in terms of a sequence of revision systems $\mathbf{S}^{\#0}$, $\mathbf{S}^{\#1}$, ... §3 will provide details. I'll use the name '$\mathbf{S}^{\#}$' both for the revision system $\mathbf{S}^{\#}$ as well as the corresponding revision theory, trusting context to disambiguate when I am talking about a mathematical object versus a semantical theory. Similarly, context will clarify when 'N' is a variable, so that '$\mathbf{S}^{\#N}$' ranges over individual revision systems $\mathbf{S}^{\#0}$, $\mathbf{S}^{\#1}$, ..., and when '$\mathbf{S}^{\#N}$' is a closed term referring to a single revision theory.

[3] In general, Bayesians do not accept the inference "E_1, ..., E_n individually confirm H; therefore, their conjunction confirms H." They do, however, under certain independence assumptions that intuitively are satisfied by §10's desiderata. Most straightforwardly, one can note that, where E_i = "desideratum i speaks in favor of $\mathbf{S}^{\#N}$" and H = "$\mathbf{S}^{\#N}$ is superior to $\mathbf{S}^{\#}$," each E_i intuitively continues to confirm H even after updating on E_1, ..., E_{i-1}.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

I'll begin by presenting the formal systems $\mathbf{S}^{\#}$ and $\mathbf{S}^{\#N}$, followed by a few examples. §3 will then introduce the semantical theories $\mathbf{S}^{\#}$ and $\mathbf{S}^{\#N}$ based on these systems. This section will give a philosophical interpretation of *revision strengths*, a new parameter appearing in my systems. In subsequent sections I compare the theories $\mathbf{S}^{\#}$ and $\mathbf{S}^{\#N}$ regarding categorical truth (§4), entailment (§5), proof systems (§6), truth theories (§7), and theoretical simplicity (§9). Along the way, I respond in §§6.3 and 7.2 to Gupta's arguments for ω-inconsistency. §8 extends the considerations of §7, and §10 summarizes my conclusions with a "scorecard" for the two theories.

§1. The semantical systems $\mathbf{S}^{\#}$ and $\mathbf{S}^{\#N}$.

1.1. Generalized schematic definitions. Our aim in offering a revision semantics is to enrich a classical ground language with circularly defined predicates.[4] For our ground language $\mathcal{L}$ we may take any first-order language with identity and the standard classical connectives.[5] We will call a classical model M $(= \langle D, I \rangle)$ for $\mathcal{L}$ a *ground model*. D is the *domain* of M. I, the *interpretation function* of M, assigns each n-place predicate a set of n-tuples from D^n, each individual constant an element of D, and each n-place function symbol a function $D^n \to D$. For a set of sentences Γ, I'll use $M \models \Gamma$ to say that $M \models A$ for all $A \in \Gamma$.

We expand $\mathcal{L}$ to $\mathcal{L}^+$ by introducing a countable (finite or infinite) set of new unary predicates $G_1, G_2, \ldots$, each defined by a corresponding definition $\mathcal{D}_i$. Let $\mathcal{D}$ be the set of all such definitions $\mathcal{D}_i$. We may assume that all definitions in $\mathcal{D}$ conform to the highly flexible form of *generalized schematic definitions* [13, sec. 145].[6] A generalized schematic definition of G uses a *(generalized) scheme* ϕ, which is a function:

(i) that maps none, some, or all *closed* terms t of $\mathcal{L}$ to sentences ϕ_t of $\mathcal{L}^+$ (which may contain any of $G_1, G_2, \ldots$); and
(ii) that maps one variable x of $\mathcal{L}$ to a formula ϕ_x of $\mathcal{L}^+$ with only x free.

The scheme ϕ is a function in our metalanguage, and need not (though may) be represented in $\mathcal{L}^+$. Given a scheme ϕ_i for one of our G_i, we may write a generalized schematic definition as

$$G_i t =_{Df} \phi_{i,t}; \text{ otherwise, } G_i x =_{Df} \phi_{i,x}. \qquad (\mathcal{D}_i)$$

The first clause is called the *partial-definition clause* of $\mathcal{D}_i$; the second is the *explicit definition clause*. G_i is the *defined term*. We say that $\phi_{i,t}$ [respectively, $\phi_{i,x}$] is the *definiens* for $G_i t$ [$G_i x$] in ϕ_i (and also in $\mathcal{D}_i$). When the domain of ϕ_i contains finitely many closed terms and $\phi_{i,x} = x \neq x$, we obtain the *partial definition*

$$G_i t_0 =_{Df} A_0; \quad G_i t_1 =_{Df} A_1; \quad \ldots; \quad G_i t_n =_{Df} A_n. \qquad (\mathcal{D}_{partial})$$

Alternatively, we can let the domain of ϕ_i contain only x to obtain the *explicit definition*

$$G_i x =_{Df} \phi_{i,x}. \qquad (\mathcal{D}_{exp})$$

[4] My presentation of the revision-theoretic framework in this section follows Gupta [12], with additional material on sets of multiple definitions and on generalized schematic definitions drawn from RTT and Gupta [13], respectively.

[5] For definiteness, take the primitive logical constants of $\mathcal{L}$ to be $\neg$, $\wedge$, $\forall$, and $=$; we may take $\vee$, $\rightarrow$, $\leftrightarrow$, and $\exists$ to be defined.

[6] The format of generalized schematic definitions presented here can easily be extended to accommodate n-place predicates and function symbols.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

When ϕ_i contains terms other than x in its domain, the definition $\mathcal{D}_i$ carries a *presupposition*, notated $Pr_{\mathcal{D}_i}$, that distinct closed terms in the domain of ϕ_i refer to different things. We may identify $Pr_{\mathcal{D}_i}$ with the set of sentences of the form $t_1 \neq t_2$, where t_1 and t_2 are distinct closed terms in the domain of ϕ_i. A ground model M is *normal for* $\mathcal{D}_i$ iff $M \models A$ for all $A \in Pr_{\mathcal{D}_i}$. According to the semantical system presented below, when a model is not normal for a definition $\mathcal{D}_i$ of G_i, G_i will be categorically false of everything in the domain.

1.2. Categorical truth in $\mathbf{S}^{\#}$ ***and*** $\mathbf{S}^{\#N}$**.** We will define our systems' key semantical notion of categorical truth by isolating a set of "optimal" hypotheses, or interpretations of the defined terms.

Let $\mathcal{L}$ be a ground language and M $(= \langle D, I \rangle)$ a ground model for $\mathcal{L}$. A *hypothesis* for a set of defined predicates $G_1, G_2, \ldots$ is a function h that assigns a candidate extension $h(G_i) \subseteq D$ to each defined predicate G_i. The set of hypotheses is the set of all such functions, which we will denote Hyp. A ground model M may be supplemented with any $h \in Hyp$ to produce the model $M + h$ that is just like M except that each G_i is assigned the extension $h(G_i)$. When there is only one defined term, we'll abuse notation by treating h simply as the extension of G_1.

We can view circular definitions as instructions for refining hypotheses. Relative to a ground model M, the set of definitions $\mathcal{D}$ yields a *revision rule* $\delta_{\mathcal{D},M} : Hyp \rightarrow Hyp$, defined as follows. For all i, hypotheses h, and $d \in D$: $d \in \delta_{\mathcal{D},M}(h)(G_i)$ iff either

(i) M is normal for $\mathcal{D}_i$ and there is a term t in the domain of ϕ_i such that $I(t) = d$ and $\phi_{i,t}$ is true in $M + h$; or
(ii) M is normal for $\mathcal{D}_i$, there is no such term t, and d satisfies $\phi_{i,x}$ in $M + h$.

It follows that when M is not normal for $\mathcal{D}_i$ the extension $\delta_{\mathcal{D},M}(h)(G_i)$ is empty. Informally speaking, given a hypothesis h, the revision rule produces revised extensions $\delta_{\mathcal{D},M}(h)(G_1)$, $\delta_{\mathcal{D},M}(h)(G_2), \ldots$ comprising the members of the domain that would satisfy the definientia according to h.

We can distinguish optimal hypotheses using *revision sequences*. Let On be the class of all ordinals, let $\mathcal{S}$ be an On-long sequence of elements of Hyp, and let $\mathcal{S}_\beta$ be the βth member of $\mathcal{S}$.[7,8] If α is a limit ordinal, then an element $d \in D$ is *stably* G_i [*stably not* G_i] *in* $\mathcal{S}$ *at* α iff

$$\exists\gamma < \alpha \; \forall\beta \; (\text{if } \gamma \leq \beta < \alpha \text{ then } d \in [\notin]\mathcal{S}_\beta(G_i)).$$

We say that d is *stable for* G_i *in* $\mathcal{S}$ *at* α iff d is either stably G_i or stably not G_i in $\mathcal{S}$ at α; otherwise, d is *unstable for* G_i *in* $\mathcal{S}$ *at* α. A hypothesis h *coheres with* $\mathcal{S}$ *at* a limit ordinal α iff for all G_i:

(i) if $d \in D$ is stably G_i in $\mathcal{S}$ at α then $d \in h(G_i)$, and
(ii) if $d \in D$ is stably not G_i in $\mathcal{S}$ at α then $d \notin h(G_i)$.

Finally, $\mathcal{S}$ is a *revision sequence* for a revision rule ρ iff $\mathcal{S}$ is an On-long sequence of hypotheses such that, for all ordinals α and β,

[7] While quantification over proper classes aids presentation, the notions of $\mathbf{S}^{\#}$ may all be defined in **ZFC** with urelements. See page 175 of RTT for details.

[8] Unless otherwise indicated, throughout I use the letters i through n to range over natural numbers; α, β, and γ to range over ordinals; and h to range over hypotheses.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

(i) if $\alpha = \beta + 1$ then $\mathcal{S}_\alpha = \rho(\mathcal{S}_\beta)$, and
(ii) if α is limit then $\mathcal{S}_\alpha$ coheres with $\mathcal{S}$ at α.

A hypothesis h is *cofinal* in a revision sequence $\mathcal{S}$ iff for all α some $\beta \geq \alpha$ is such that $\mathcal{S}_\beta = h$. Cardinality considerations imply that all revision sequences contain cofinal hypotheses (RTT, Theorem 5C.7). A hypothesis h is *recurring for* a revision rule ρ iff h is cofinal in some revision sequence for ρ. Since the revision process never eliminates recurring hypotheses, we can reasonably consider these optimal for ρ.

The definitions of categorical truth in $\mathbf{S}^\#$ and $\mathbf{S}^{\#N}$ are most transparent when set alongside the corresponding definition of a weaker system, $\mathbf{S}^*$, presented in RTT. This system is the most well-known ω-consistent alternative to $\mathbf{S}^\#$. It defines categorical truth as truth on all recurring hypotheses.

Definition 1.1 (Categorical truth in $\mathbf{S}^*$). *Let A be a sentence of $\mathcal{L}^+$ and M a ground model. A is* categorically true, *or simply* categorical,[9] *in M relative to $\mathcal{D}$ in the system $\mathbf{S}^*$ (notation: $M \models^{\mathcal{D}}_* A$) iff*

$$\forall h \textit{ recurring for } \delta_{\mathcal{D},M}, M + \delta_{\mathcal{D},M}(h) \models A.$$

In $\mathbf{S}^*$, as well as the systems presented below, a sentence is *categorically false* iff its negation is categorically true, and *pathological* iff it is neither categorically true nor categorically false.[10]

Unfortunately, $\mathbf{S}^*$'s definition of categorical truth is unsatisfactory, for many of the recurring hypotheses contain disruptive artifacts that creep in at limit stages (see Example 6.3). We can clean the recurring hypotheses up by revising them finitely many more times. This motivates Definition 1.2. In this definition, $\rho^n(h)$ denotes the result of applying the revision rule ρ n times to h, with $\rho^0(h) = h$.

Definition 1.2 (Categorical truth in $\mathbf{S}^\#$). *Let A be a sentence of $\mathcal{L}^+$ and M a ground model. A is* categorically true, *or* categorical, *in M relative to $\mathcal{D}$ in the system $\mathbf{S}^\#$ (notation: $M \models^{\mathcal{D}}_\# A$) iff*

$$\exists n \forall h \textit{ recurring for } \delta_{\mathcal{D},M}, M + \delta^n_{\mathcal{D},M}(h) \models A. \tag{1}$$

Again, a sentence is categorically false *iff its negation is categorically true.*[11]

My new systems $\mathbf{S}^{\#N}$ offer a different way to improve upon $\mathbf{S}^*$: rather than revising the recurring hypotheses an *arbitrary* number of times, as in Definition 1.2, we revise them only up to some fixed natural number N. We may call the parameter N a *revision strength*.[12] I'll postpone further philosophical discussion of revision strengths to §3. For now, we have the following definition.

[9] My use of the term 'categorical' differs from RTT's, where it means *either categorically true or categorically false.*

[10] Further semantical distinctions among the pathological sentences can be drawn, including a notion of paradoxicality. See RTT, Definition 5D.15.

[11] Definition 1.2 is equivalent to the definition of $\mathbf{S}^\#$ in [12, 13]. RTT originally used the name $\mathbf{S}^\#$ for a slightly stronger system, which switches the order of the '$\forall h$' and '$\exists n$' quantifiers.

[12] One may also devise systems in which each of several interacting circular definitions is assigned its own revision strength. One can construct corresponding variations of the proof systems $\mathbf{C}^N$ (§6) that are sound for these multi-strength semantics. I've chosen not to present these systems because they add considerable complexity for a slender philosophical payoff.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

DEFINITION 1.3 (Categorical truth in the systems $\mathbf{S}^{\#N}$). *Let A be a sentence of $\mathcal{L}^+$, M a ground model, and $N \in \mathbb{N}$. A is* categorically true, *or* categorical, *in M relative to $\mathcal{D}$ in the system $\mathbf{S}^{\#N}$ (notation: $M \models^{\mathcal{D}}_{\#N} A$) iff*

$$\forall h \text{ recurring for } \delta_{\mathcal{D},M},\ M + \delta^N_{\mathcal{D},M}(h) \models A.^{13}$$

I wish to stress that, since N is a parameter that may vary, Definition 1.3 defines a sequence of revision systems $\mathbf{S}^{\#0}$, $\mathbf{S}^{\#1}$, etc.

The term "revision strength" is motivated by the fact that increasing N results in more and more sentences becoming categorical in $\mathbf{S}^{\#N}$. $\mathbf{S}^{\#0}$ is just $\mathbf{S}^*$; increasing N results in progressively stronger systems, with $\mathbf{S}^{\#}$ the limit of this sequence. Relatedly, we have the following proposition, which will be important for my subsequent discussion.

PROPOSITION 1.4. *Let Γ be a finite set of sentences of a language $\mathcal{L}^+$ whose circular predicates are defined by $\mathcal{D}$, and let M be a ground model. Then there is a strength N such that for all $A \in \Gamma$, $M \models^{\mathcal{D}}_{\#} A$ iff $M \models^{\mathcal{D}}_{\#N} A$.*

Proof. For each $A \in \Gamma$ such that $M \models^{\mathcal{D}}_{\#} A$, let $N(A)$ be the least n witnessing (1) of Definition 1.2. Choose $N = \max(\{N(A) : A \in \Gamma\} \cup \{0\})$. □

§2. Examples of circular definitions.

EXAMPLE 2.1 (A finite definition). *Let the ground language $\mathcal{L}$ contain individual constants a and b, which the ground model M interprets as* 0 *and* 1, *respectively. The domain D may contain other objects. We define H explicitly as follows*:

$$Hx =_{Df} (x = a \vee x = b) \wedge \neg Ha \wedge \neg Hb. \qquad (\mathcal{D}_0)$$

For all hypotheses $h \subseteq D$, the revision rule $\delta_{\mathcal{D}_0,M}$ is as follows:

$$\delta_{\mathcal{D}_0,M}(h) = \begin{cases} \{0,1\} & \text{if } 0 \notin h \text{ and } 1 \notin h; \\ \varnothing & \text{otherwise.} \end{cases}$$

The recurring hypotheses are exactly $\varnothing$, $\{0\}$, $\{1\}$, and $\{0,1\}$. Table 1 illustrates how $\mathbf{S}^*$, $\mathbf{S}^{\#}$, and $\mathbf{S}^{\#N}$ treat some pertinent sentences.

Table 1. *Semantical statuses of sentences in M relative to $\mathcal{D}_0$ in $\mathbf{S}^*$, $\mathbf{S}^{\#N}$, and $\mathbf{S}^{\#}$*

	Ha	Hb	$\forall x(x \neq a \wedge x \neq b \rightarrow \neg Hx)$	$Ha \leftrightarrow Hb$
$\mathbf{S}^*$ $(= \mathbf{S}^{\#0})$	—	—	✓	—
$\mathbf{S}^{\#N}, N \geq 1$	—	—	✓	✓
$\mathbf{S}^{\#}$	—	—	✓	✓

Note: '✓' indicates a categorical truth; '—' indicates pathologicality.

[13] The definiens of Definition 1.3 is equivalent to the following:

$$\exists n \leq N\ \forall h \text{ recurring for } \delta_{\mathcal{D},M},\ M + \delta^n_{\mathcal{D},M}(h) \models A,$$

which shows how $\mathbf{S}^{\#N}$ can be obtained by bounding the existential quantifier in Definition 1.2.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

The sentence $Ha \leftrightarrow Hb$ is pathological in $\mathbf{S}^*$ due to the recurring hypotheses $\{0\}$ and $\{1\}$. These occur only at the limit stages of revision sequences, after which they are immediately revised away. The fact that $\mathbf{S}^{\#1}$, like $\mathbf{S}^{\#}$, eliminates these imperfections in the revision process already illustrates an improvement over $\mathbf{S}^*$. §6.3 will expand upon this theme.

EXAMPLE 2.2 (Truth). *By far the most important applications of the Revision Theory are to semantic predicates such as truth. In certain models, truth can be given a generalized schematic definition, as follows. Let us expand a ground language $\mathcal{L}$ to $\mathcal{L}^+$ by adding a one-place predicate T. Let a* full sentential frame *for $\mathcal{L}^+$ be a pair $\langle M, \ulcorner\ \urcorner\rangle$, where M is a ground model whose domain contains every sentence of $\mathcal{L}^+$ and $\ulcorner\ \urcorner$ is an injection that*:

(i) *maps every sentence $B \in \mathcal{L}^+$, to a closed term $\ulcorner B \urcorner$ of $\mathcal{L}$ that refers in M to B; and*
(ii) *maps the formula $x \neq x$ to x.*

In a full sentential frame, we may use $(\ulcorner\ \urcorner)^{-1}$ as a scheme to define T by the truth definition

$$T\ulcorner A\urcorner =_{Df} A;\ \textit{otherwise},\ Tx =_{Df} x \neq x. \qquad (\mathcal{T})$$

Its revision rule in M is the Tarski jump, *τ_M. For any hypothesis h,*

$$\tau_M(h) = \{A \in \mathcal{L}^+ : M + h \models A\}.$$

That is, τ_M takes the sentences that were true according to h and puts them in the extension of $\tau_M(h)$.

For now I will discuss only the basic outlines of how $\mathbf{S}^*$, $\mathbf{S}^{\#}$, and the systems $\mathbf{S}^{\#N}$ handle self-referential sentences, on which they all agree. (All statements about the systems $\mathbf{S}^{\#N}$ in this example hold for all values of N.) Later, in §7, I'll study the differences between the systems. In all three systems, $T\ulcorner A\urcorner \leftrightarrow A$ is categorical for any sentence of $\mathcal{L}$, i.e., not containing T. In fact, a stronger result holds: when there is no vicious self-reference in $\mathcal{L}^+$, $T\ulcorner A\urcorner \leftrightarrow A$ is categorical for all A of the *expanded* language.[14] Now suppose that M and $\ulcorner\ \urcorner$ are such that $\mathcal{L}^+$ contains a *liar sentence* $\lambda = \ulcorner \neg T\lambda \urcorner$. Then both $T\lambda$ and $\neg T\lambda$ are pathological, and $T\ulcorner \neg T\lambda \urcorner \leftrightarrow \neg T\lambda$ is categorically false. But, since revision semantics preserves logical truths, $T\lambda \vee \neg T\lambda$ is categorically true. If $\mathcal{L}^+$ contains a *truth-teller* $\theta = \ulcorner T\theta \urcorner$, $T\theta$ and $\neg T\theta$ are pathological, but $T\ulcorner T\theta \urcorner \leftrightarrow T\theta$ is categorical (indeed, a logical truth).

My final example, a focal point of the subsequent discussion, highlights the crucial differences between $\mathbf{S}^{\#}$ and the systems $\mathbf{S}^{\#N}$.[15]

EXAMPLE 2.3 (An ω-inconsistency in $\mathbf{S}^{\#}$). *Let $\mathcal{L}$ contain both*

(i) *arithmetical symbols*: *an individual constant* 0, *a one-place function symbol* $'$, *and a two-place predicate* $<$; *and*
(ii) *syntactical symbols*: *an infinity of constants $c_0, c_1, \ldots$ (to be used as names of sentences), plus a one-place function symbol g.*

[14] See RTT, Theorem 6A.5. The condition of "no vicious self-reference" is made precise by the notion of S-neutrality (RTT, Definition 6A.3).

[15] A stripped-down version of Example 2.3, without the truth predicate, first appeared in RTT, pp. 155–156.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

We will write $\overline{n}$ for 0 followed by n applications of $'$; $\geq$ and $\leq$ can be defined from $<$ and $=$ in the ordinary way. Further, for a formula A, let $\forall iA$, $\forall jA$, ... abbreviate $\forall x(0 \leq x \rightarrow A), \forall y(0 \leq y \rightarrow A), \ldots$. $\mathcal{L}^+$ is obtained by adding to $\mathcal{L}$ one-place predicates G and T. Let $\langle M, \ulcorner\ \urcorner\rangle$ be a full sentential frame for $\mathcal{L}^+$, in which we stipulate the following:

- $D = \mathbb{N} \cup \{A : A$ *is a sentence of* $\mathcal{L}^+\}$.
- I *gives the mathematical symbols their standard interpretations. In particular,* $\langle d_1, d_2\rangle \in I(<)$ *only if* $d_1, d_2 \in \mathbb{N}$, *and* $I(')$ *maps sentences to the dummy object* $\forall x\ x = x$ *(which is in* D*).*
- $I(g)$ *maps each natural number n to its G-predication* $G\overline{n}$, *and maps sentences to* $\forall x\ x = x$.

We define G explicitly by

$$Gx =_{Df} 0 \leq x\ \wedge\ \neg(G \textit{ is closed}) \wedge \forall j(j < x \rightarrow Gj),$$

where 'G is closed' abbreviates $G0 \wedge \forall y(Gy \rightarrow Gy')$. *T is defined as in the previous example*:

$$T\ulcorner A\urcorner =_{Df} A;\ \textit{otherwise},\ Tx =_{Df} x \neq x.$$

Let $\mathcal{D}_1$ *be the set consisting of the definitions for G and T. Notice that these definitions are interdependent*: *some definienda of the truth definition, such as the sentence* $T\ulcorner G0\urcorner$, *are themselves defined in terms of the circular predicate G.*

The revision rule $\delta_{\mathcal{D}_1,M}$ *is as follows. For any hypothesis h,*

$$\delta_{\mathcal{D}_1,M}(h)(G) = \begin{cases} \{n \in \mathbb{N} : n \leq m\} & \textit{if } m = \min\{k \in \mathbb{N} : k \notin h(G)\}; \\ \varnothing & \textit{if } \mathbb{N} \subseteq h. \end{cases}$$

$$\delta_{\mathcal{D}_1,M}(h)(T) = \{A \in \mathcal{L}^+ : M + h \models A\}.$$

One can verify that every subset of $\mathbb{N}$ *occurs as* $h(G)$ *for some recurring hypothesis h.*

The following sentences are all categorical in M relative to $\mathcal{D}_1$ in $\mathbf{S}^{\#}$:

$$G0,\ G0',\ G0'', \ldots,\ \exists i\neg Gi;\ T\ulcorner G0\urcorner,\ T\ulcorner G0'\urcorner, \ldots,\ \exists i\neg Tg(i). \tag{2}$$

Of those, $\mathbf{S}^{\#N}$ $(N \geq 3)$ regards only these as categorical:

$$\begin{aligned} &G0,\ G0', \ldots,\ G\overline{(N-2)},\ \exists i\neg Gi; \\ &T\ulcorner G0\urcorner,\ T\ulcorner G0'\urcorner, \ldots,\ T\ulcorner G\overline{(N-3)}\urcorner,\ \exists i\neg Tg(i). \end{aligned} \tag{3}$$

It regards these as pathological:

$$G\overline{(N-1)},\ G\overline{N}, \ldots;\ T\ulcorner G\overline{(N-2)}\urcorner,\ T\ulcorner G\overline{(N-1)}\urcorner, \ldots. \tag{4}$$

Note that the truth predications become pathological one numeral earlier than the corresponding G-predications.

Standardly, an ω-inconsistency is defined as a set of sentences $\{\exists x\neg A(x), A(0), A(0'), \ldots\}$ for some formula A. For languages that can talk about non-mathematical objects, an ω-inconsistency should probably be defined instead as a set of sentences $\{\exists x(0 \leq x \wedge \neg A(x)), A(0), A(0'), \ldots\}$. On the revised definition, the sentences $\exists i\neg Gi$, $G0$, $G0'$, ... form an ω-inconsistency that is categorically true in $\mathbf{S}^{\#}$. As (2) indicates,

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

this ω-inconsistency is mirrored by a second one, $\{\exists i \neg Tg(i), Tg(0), Tg(0'), \ldots\}$, at the level of truth predications. Turning now to $\mathbf{S}^{\#N}$, we can see in Example 2.3 the role of the bound N in avoiding ω-inconsistency: bounding the revision process prevents some of the sentences $G\overline{n}$ from becoming categorically true.

§3. Setting revision strengths. So much for the technical apparatus. What philosophical claims about semantics can we make with these revision systems? As above, we may identify the $\mathbf{S}^{\#}$ *theory* with the claim

($\mathbf{S}^{\#}$) A sentence A of a language $\mathcal{L}^{+}$ interpreted according to the ground model M and supplemented with definitions $\mathcal{D}$ is categorically true iff $M \models^{\mathcal{D}}_{\#} A$.

We could similarly formulate semantical theories corresponding to each of the systems $\mathbf{S}^{\#0}$, $\mathbf{S}^{\#1}$, and so on. An arguably more attractive theory, however, results from letting different revision strengths apply to different languages.[16] The result is the following $\mathbf{S}^{\#N}$ *theory*:

($\mathbf{S}^{\#N}$) A sentence A of a language $\mathcal{L}^{+}$ interpreted according to the ground model M and supplemented with definitions $\mathcal{D}$ is categorically true iff $M \models^{\mathcal{D}}_{\#N} A$, where N is the revision strength of $\mathcal{L}^{+}$.

As stated, the $\mathbf{S}^{\#N}$ theory must be supplemented with an account of what it is for N to *be the revision strength of* $\mathcal{L}^{+}$. I'll present such an account below, and expand upon it in §8.

In determining the revision strength of a language, there are two cases to consider. In Case 1, a speaker stipulates the revision strength for her idiolect, or the speakers of a dialect collectively stipulate its revision strength. Of course, this case is extraordinarily rare: it presupposes that the speakers are not only revision theorists but proponents of $\mathbf{S}^{\#N}$ (or some closely related theory). Usually, we are in Case 2: no revision strength is successfully stipulated, either because no speaker attempted to or because they did not have the authority to perform the stipulation. (For instance, I cannot by myself stipulate the revision strength of the English language for the same reason that I cannot by mere stipulation change the conventional meanings of English words.) The most prominent instances of Case 2 are natural languages containing a truth predicate. According to the Revision Theory, truth has a circular semantics. But clearly no natural language's speakers ever collectively decided upon their language's revision strength.

A language's revision strength in Case 2 must, then, be set to some "default" strength. The following principle provides an attractive way of determining default strengths for a large class of languages:

[16] One can think of a language either as a mathematically defined abstract entity (an *abstract language*) or as an instance of such an entity as used by a group of speakers (a *language-in-use*). The latter could perhaps be defined as an ordered pair of an abstract language and a set of speakers. Strictly speaking, semantical theories apply to languages-in-use, not to abstract languages. Likewise, it only makes sense to assign revision strengths to languages-in-use. These points aside, I will use 'language' and names like $\mathcal{L}^{+}$ to refer to either kind of entity, depending on context.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

(Agreement) If the circular predicates of $\mathcal{L}^+$ are defined by $\mathcal{D}$, and the totality of $\mathcal{L}^+$-sentences used by its speakers comprises a finite set Γ, then the default revision strength for $\mathcal{L}^+$ is the least N satisfying Proposition 1.4 for Γ and $\mathcal{D}$.

In the sense relevant for Agreement, a speaker *uses* a sentence B of $\mathcal{L}^+$ roughly iff she uses (in the ordinary sense) some particular token of B in an act of communication.[17] The upshot of Agreement is that the theory $\mathbf{S}^{\#N}$ agrees with $\mathbf{S}^{\#}$ on the categorical truth of all sentences actually in use—which presumably are finite in number. In the dialectical context of comparing the two theories, this fact is very effective. It promises to preempt any complaints that $\mathbf{S}^{\#N}$ lacks certain desirable features of $\mathbf{S}^{\#}$. One would respond that, as far as concerns the sentences actually in use, $\mathbf{S}^{\#N}$ yields the *very same* semantics as $\mathbf{S}^{\#}$. We will see some applications of this idea in §6.3 and §7.

So far, Agreement only applies to languages with finite usage. In §8, I generalize Agreement to languages with infinite usage and continue my comparison for such languages.

§4. Desiderata on categorical truth. We can now evaluate $\mathbf{S}^{\#}$ and $\mathbf{S}^{\#N}$ according to two desiderata on categorical truth. First, both theories satisfy Strong Conservativeness:

(Strong Conservativeness) For all sets $\mathcal{D}$ of definitions, models M, and sentences A of ground languages $\mathcal{L}$: if $\mathcal{L}$ is interpreted according to M and expanded with $\mathcal{D}$, then A is categorically true iff $M \models A$.[18]

In other words, adding definitions leaves the semantics of the ground language untouched. I think we should, with Gupta [12, p. 185], regard Strong Conservativeness as "non-negotiable." Strong Conservativeness of $\mathbf{S}^{\#}$ and $\mathbf{S}^{\#N}$ follows from the fact that every revision rule has at least one recurring hypothesis in every model (p. 7).

Our second desideratum is ω-consistency:

(ω-consistency) Where $\mathcal{L}$ is the language of arithmetic interpreted by its standard model and possibly expanded with a set $\mathcal{D}$ of definitions, no ω-inconsistent set of $\mathcal{L}$-sentences is categorically true.

In Example 2.3 we already saw that $\mathbf{S}^{\#}$ makes ω-inconsistencies categorically true, and it is clear that an ω-inconsistency would remain if one restricted the ground model's domain to the natural numbers.

That the theory $\mathbf{S}^{\#N}$ satisfies ω-consistency follows from the fact that each system $\mathbf{S}^{\#N}$ is ω-consistent. To see why the latter is true, it is informative to consider the related thesis of *weak reductionism about circular concepts*. Weak reductionism says that, for any set of definitions $\mathcal{D}$ and ground model M, there is a set of hypotheses

[17] My definition of 'use' implies that, in using a schematic sentence or other generalizing device, one does not thereby use all of the instances generalized upon. One only uses the sentences whose tokens one writes out or utters in full.

[18] See RTT, Definition 5A.13.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

$H \subseteq Hyp$ such that for all sentences A of $\mathcal{L}^+$, A is categorical in M iff $M + h \models A$ for all $h \in H$ [13, sec. 174]. If weak reductionism holds, categorical truth can always be understood as supervaluation over some set H of "acceptable" hypotheses. Weak reductionism fails in the system $\mathbf{S}^{\#}$, but in the systems $\mathbf{S}^{\#N}$ it holds for $H = \{h \in Hyp : \exists h'\ h'$ is recurring and $h = \delta^N_{\mathcal{D},M}(h')\}$. While neither weak reductionism nor its negation strikes me as intrinsically desirable, weak reductionism is one way of achieving ω-consistency. The next proposition makes this observation precise and establishes that all the systems $\mathbf{S}^{\#N}$ are ω-consistent.

Proposition 4.1. *Every revision system that satisfies both Strong Conservativeness and weak reductionism about circular concepts is ω-consistent.*

Proof. Straightforward. □

In setting up ω-consistency—or any other feature—as a desideratum, I don't claim that this feature is decisive, only that it's attractive to some extent. In the case of ω-consistency, I suspect this claim is uncontroversial—even among those who ultimately favor ω-inconsistent theories. However, it's worthwhile to briefly dispel a potential misunderstanding about why ω-inconsistency is unattractive. The standard objection to ω-inconsistency is that models of ω-inconsistent theories are arithmetically nonstandard.[19] In the context of the Revision Theory this objection loses its force (RTT, p. 156). The objection assumes, crucially, that 'model' is to be understood as a traditional, classical model—what I've been calling a ground model. Now, it is true that the theory consisting of the sentences categorically true in $\mathbf{S}^{\#}$ in Example 2.3 has no arithmetically standard models in *this* sense. But that fact is irrelevant: revision theorists think that circular concepts call for a revision semantics, not a classical model theory. The objection attempts, as it were, to fit a square semantics onto circular concepts. Meanwhile, the revision semantics in Example 2.3 *does* employ an arithmetically standard ground model.

In my view there are still two legitimate issues with ω-inconsistent revision theories. First, the incredulous stare: "What do you mean by saying that some number isn't G when you also say, of each individual number, that it *is*?" Stares may not make for eloquent arguments, but being counterintuitive surely counts against a theory. Second, rejecting ω-consistency requires denying a number of other attractive principles. Suppose we expand the language $\mathcal{L}^+$ of Example 2.3 to $\mathcal{L}^+_K$ by adding a new (non-circular) predicate $K_{\mathbf{t}}(x)$. We interpret this predicate to be true of precisely the $\mathcal{L}^+$ sentences categorically true according to $\mathbf{S}^{\#N}$, for a given N.[20] We can then show that

$$M \models^{\mathcal{D}_1}_{\#N} \forall i(K_{\mathbf{t}}(g(i)) \rightarrow Gi). \tag{K}$$

Roughly, if G is "categorically true of" i, then i is G. We could suitably generalize the claim to apply to circular definitions in any language. Due to $\mathbf{S}^{\#}$'s ω-inconsistency, however, the claim for $\mathbf{S}^{\#}$ corresponding to (K) fails to be assertible. Indeed, $\mathbf{S}^{\#}$ asserts that while the G-predication of some number is categorical, that number is not G. As

[19] Barrio and Picollo [1] offer a variant of the traditional objection by pointing out that the union of an ω-inconsistent theory with the second-order Peano axioms has *no* full second-order models.

[20] Since no sentence of the newly expanded language is in the extension of $K_{\mathbf{t}}$, it can be given a fully classical interpretation. If we wanted, we could alternatively let $K_{\mathbf{t}}$ apply to sentences of $\mathcal{L}^+_K$ and give it a circular semantics in the manner of RTT, chap. 6.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

I show in a different paper, ω-inconsistency in $\mathbf{S}^{\#}$ also results in the categorical falsity of reflection principles for the theory consisting of Peano Arithmetic closed under $\mathbf{S}^{\#}$'s proof system **C** [23]. Since reflection principles for a theory express its soundness, I argue that $\mathbf{S}^{\#}$ pays a heavy price for rejecting these principles.

§5. Entailment. In this section I compare the entailment relations of $\mathbf{S}^{\#}$ and $\mathbf{S}^{\#N}$. To define entailment in the Revision Theory, one begins with the notion of a *logical truth*: a sentence that is categorically true in all models. We then say that $A_1, \ldots, A_n$ *entail* B iff the conditional $A_1 \wedge \ldots \wedge A_n \rightarrow B$ is a logical truth. Definition 5.1 generalizes this idea to possibly infinite premise sets. Part (i) is Gupta's [12, 13] definition of entailment for $\mathbf{S}^{\#}$.

DEFINITION 5.1 (Entailment). *Let $\mathcal{L}^+$ be a ground language expanded with a set $\mathcal{D}$ of definitions, and let $N \in \mathbb{N}$.*

(i) *Sentences Γ of $\mathcal{L}^+$ entail the $\mathcal{L}^+$-sentence B in $\mathbf{S}^{\#}$ relative to $\mathcal{D}$ (notation: $\Gamma \models^{\mathcal{D}}_{\#} B$) iff for all ground models M,*

$$\exists n \forall h (\textit{if } h \textit{ is recurring for } \delta_{\mathcal{D},M}, \textit{ then } M + \delta^n_{\mathcal{D},M}(h) \models \Gamma \textit{ implies } M + \delta^n_{\mathcal{D},M}(h) \models B).$$

(ii) *Γ entails B in the system $\mathbf{S}^{\#N}$ relative to $\mathcal{D}$ (notation: $\Gamma \models^{\mathcal{D}}_{\#N} B$) iff for all ground models M,*

$$\exists n \leq N \, \forall h (\textit{if } h \textit{ is recurring for } \delta_{\mathcal{D},M}, \textit{ then } M + \delta^n_{\mathcal{D},M}(h) \models \Gamma \textit{ implies } M + \delta^n_{\mathcal{D},M}(h) \models B).$$

It's clear from these definitions that every entailment in $\mathbf{S}^{\#N}$ is also an entailment in $\mathbf{S}^{\#}$. As Example 5.1 will illustrate, some entailments in $\mathbf{S}^{\#}$ are invalid in $\mathbf{S}^{\#N}$ for any value of N. In this respect, the situation regarding entailment differs from that for categorical truth.

As with categorical truth, one should distinguish *entailment-in-a-system*, of the sort defined in Definition 5.1, from a *theory of entailment*. $\mathbf{S}^{\#N}$'s theory of entailment, for instance, says:

($\mathbf{S}^{\#N}$, Entailment) Let $\mathcal{L}^+$ be an interpreted language whose logical constants receive their ordinary classical interpretations and which is supplemented with definitions $\mathcal{D}$. Then a set Γ of $\mathcal{L}^+$-sentences entails an $\mathcal{L}^+$-sentence A iff $\Gamma \models^{\mathcal{D}}_{\#N} A$, where N is the revision strength of $\mathcal{L}^+$.

The corresponding theory of entailment for $\mathbf{S}^{\#}$ is straightforward. The rest of this section compares these two theories.

When the logic of the ground language is first-order classical logic, both theories of entailment have the following attractive feature:

(Ground Logic Preservation) Forms of argument that are valid in the ground language continue to be valid after the introduction of circular definitions.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

In our present setup, the ground logic is first-order classical logic, but Ground Logic Preservation will continue to hold when $\mathbf{S}^{\#}$ and $\mathbf{S}^{\#N}$ are defined for ground languages with a range of alternative logics. Gupta [12] regards Ground Logic Preservation, like Strong Conservativeness, as non-negotiable and in Gupta [13, sec. 49] provides two arguments for Ground Logic Preservation. I am not sure these considerations establish such a strong requirement, but even so, Ground Logic Preservation is clearly attractive. By the end of this section it will emerge that *requiring* this feature strengthens the considerations in favor of $\mathbf{S}^{\#N}$.

I'll now demonstrate a surprising feature of $\mathbf{S}^{\#}$: it regards some collections of categorical truths as contradictory.

Example 5.1 (Contradictory categorical truths in $\mathbf{S}^{\#}$). *Define $\mathcal{L}_<$ as the ground language with the non-logical symbols* 0, $<$, *and* $'$. *In its* standard model, $\mathbb{N}_<$ $(= \langle \mathbb{N}, I_< \rangle)$, *$I_<$ interprets* 0, $<$, *and* $'$ *as standard. $\mathcal{L}_<^+$ adds the predicate G to $\mathcal{L}_<$ and gives it the explicit definition*

$$Gx =_{Df} (G \textit{ is closed} \wedge \exists y \neg Gy \wedge Gx) \vee [\neg(G \textit{ is closed}) \wedge \forall y(y < x \rightarrow Gy)]. \qquad (\mathcal{D}_2)$$

'G is closed' still abbreviates $G0 \wedge \forall y(Gy \rightarrow Gy')$. We will use AX to abbreviate the conjunction of axioms stating that $<$ is a strict linear order,[21] *together with*

$$\forall x \; x \not< 0 \qquad \text{(AX1)}$$

$$\forall x \; (x = 0 \vee \exists y \; x = y') \qquad \text{(AX2)}$$

$$\forall x \; x < x' \qquad \text{(AX3)}$$

$$\forall x \forall y (x < y' \rightarrow x < y \vee x = y). \qquad \text{(AX4)}$$

In $\mathbb{N}_<$, the revision rule for $\mathcal{D}_2$ has the same effect as the rule for $\mathcal{D}_1$ on hypothetical extensions for G, so that in revision sequences its extension fluctuates endlessly. But now consider a nonstandard model M satisfying AX; such a model must contain "nonstandard numbers" larger than all the standard numbers. Revision sequences for $\mathcal{D}_2$ in M always culminate in a fixed point $h = \delta_{\mathcal{D}_2, M}(h)$, after which the extension never changes. At such fixed points, G *is closed* comes out true.

Where $\Gamma = \{AX, G0, G0', G0'', \ldots, \neg(G \textit{ is closed})\}$, we can use the above facts to show that $\Gamma \underset{\#}{\overset{\mathcal{D}_2}{\not\models}} \forall x \; x \neq x$. Nevertheless, all the sentences in Γ are categorically true in the standard model relative to $\mathcal{D}_2$ in $\mathbf{S}^{\#}$. (See Proposition A.1 in the Appendix.) Note that, since $\neg(G \textit{ is closed})$ classically entails $\exists x \neg Gx$, in $\mathbb{N}_<$ $\mathcal{D}_2$ generates an ω-inconsistency, which is partly responsible for the strange result.[22]

---

[21] These axioms state that $<$ is irreflexive, transitive, and connected $[\forall x \forall y (x < y \vee x = y \vee x > y)]$.

[22] $\mathbf{S}^{\#}$ and its weaker relative $\mathbf{S}$ differ somewhat on whether they regard sets of categorical truths as contradictory. The example's entailment from Γ to $\forall x \; x \neq x$ doesn't hold in $\mathbf{S}$, and more generally $\mathbf{S}$ satisfies Categorical Truth Preservation for sets of explicit definitions. However, relative to generalized schematic definitions, categorical truths nonetheless entail contradictions in $\mathbf{S}$. For instance, expand $\mathcal{L}_<$ with the following set of definitions:

$$\begin{aligned} &Gx =_{Df} \neg(G \textit{ is closed}) \wedge \forall y(y < x \rightarrow Gy) \\ &Ht =_{Df} \phi_t; \text{ otherwise, } Hx =_{Df} \phi_x, \end{aligned} \qquad (\mathcal{D})$$

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

Example 5.1 looks disastrous for $\mathbf{S}^{\#}$'s entailment theory. When we interpret $\mathcal{L}_<$ standardly, Γ is a correct description of how the world actually is, according to $\mathbf{S}^{\#}$. $\mathbf{S}^{\#}$ then goes on to say that this description is trivial: it entails everything. But surely the actual world is not trivial! $\mathbf{S}^{\#N}$ corrects these problems by eliminating ω-inconsistency. One can easily show that it satisfies a stronger desideratum:

(Categorical Truth Preservation) Entailment preserves categorical truth: in an interpreted classical language expanded with circular definitions, whenever Γ entails A, then A is categorically true if all members of Γ are.

A defender of $\mathbf{S}^{\#}$ has a couple of responses available. The less attractive option is to restrict the theory of entailment to finite premise sets. We can then show that finite entailment in $\mathbf{S}^{\#}$ preserves categorical truth. This move still places $\mathbf{S}^{\#}$ at a significant disadvantage to $\mathbf{S}^{\#N}$. We can apply the Revision Theory to languages whose ground logic is not compact—for instance, second-order languages under the full semantics. In such cases, $\mathbf{S}^{\#N}$ will capture all the entailments of the ground language (i.e., preserve ground logic), whereas $\mathbf{S}^{\#}$ will miss out on interesting entailments. Worse, the restriction to finite entailments seems to merely push the problem out of view rather than solve it. If the entailment relation on all premise sets was implausible, that seems to indicate that it was the wrong *kind* of entailment relation. We should still think it is the wrong kind of entailment relation when restricted to the finite case.

A second way to defend $\mathbf{S}^{\#}$ is to employ a different notion of entailment. We could *define* entailment in $\mathbf{S}^{\#}$ as categorical truth preservation in all models, and so achieve Categorical Truth Preservation by fiat. We could call the old and new relations *local* and *global entailment*, respectively.[23]

DEFINITION 5.2 (Global entailment in $\mathbf{S}^{\#}$). *Let $\mathcal{L}$ be expanded to $\mathcal{L}^+$ with a set $\mathcal{D}$ of definitions. Sentences Γ of $\mathcal{L}^+$* globally entail *the $\mathcal{L}^+$-sentence B in $\mathbf{S}^{\#}$ relative to $\mathcal{D}$ (notation: $\Gamma \models^{\mathcal{D}}_{\#,g} B$) iff, for all ground models M, $M \models^{\mathcal{D}}_{\#} \Gamma$ implies $M \models^{\mathcal{D}}_{\#} B$.*

The trouble with global entailment is now Ground Logic Preservation. The new definition still preserves all valid entailments of first-order logic with identity, but it ceases to preserve a number of very useful meta-rules. The following, all preserved by local entailment in $\mathbf{S}^{\#}$ and $\mathbf{S}^{\#N}$, are no longer admissible under global entailment:

(Conditional Proof) If $\Gamma, A \models B$, then $\Gamma \models A \rightarrow B$.

(Proof by Cases) If $\Gamma, A \models C$ and $\Delta, B \models C$, then $\Gamma, \Delta, A \vee B \models C$.

(*Reductio ad absurdum*) If $\Gamma, A \models B$ and $\Delta, A \models \neg B$, then $\Gamma, \Delta \models \neg A$.

(Contraposition) If $\Gamma, A \models B$, then $\Gamma, \neg B \models \neg A$.[24]

---

where ϕ maps each numeral $\overline{n}$ to $\forall x\ x = x$ and maps the variable x to $x \neq x$. In $\mathbf{S}$ relative to $\mathcal{D}$, $\{AX, G0, G0', G0'', \ldots, \forall x Hx\}$ is categorical in $\mathbb{N}_<$, though these sentences entail $\forall x\ x \neq x$.

[23] The terminology mimics a similar distinction in the supervaluationism literature (see [25]). Yaqub [29, chap. 4.3] considers revision-theoretic definitions of entailment similar to global entailment but rejects them on the grounds that they treat pathological sentences in the same way as contradictions and that they disallow classically admissible meta-rules. Field [7, chap. 11], in his treatment of revision systems, considers only the global option.

[24] Counterexamples to these rules are easily found by using a liar sentence. Williamson [26–28] has been an especially fierce critic of supervaluational theories of vagueness for their rejection of these four meta-rules.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

If the ground logic is instead second-order logic with the full semantics, global entailment in $\mathbf{S}^{\#}$ ceases to preserve even ordinary entailments. Where PA^2 is the conjunction of the second-order Peano axioms, the following argument is semantically valid in second-order logic:

$$PA^2, \neg(G \textit{ is closed}), G0, G0', \ldots \models \forall x\ x \neq x. \tag{5}$$

But when G is defined by $\mathcal{D}_2$, the argument is globally invalid in $\mathbf{S}^{\#}$. As before, I think giving up these aspects of ground logic could perhaps be managed, though at considerable cost.

Note that since the premises of (5) are all categorically true in $\mathbf{S}^{\#}$ on the standard model of $\mathcal{L}_<$, second-order logic declares a set of categorical truths contradictory. Hence, no matter how $\mathbf{S}^{\#}$ defines entailment, it is forced to reject either Categorical Truth Preservation or Ground Logic Preservation for second-order logic. On the other hand, $\mathbf{S}^{\#N}$ achieves both desiderata at once. Either way, $\mathbf{S}^{\#N}$ comes out ahead.

§6. Proof systems. Perhaps the most pressing challenge for an ω-consistent revision theory is providing a satisfactory way of reasoning with circular definitions. Beyond the minimal requirement of soundness, such a proof system should be strong and well-motivated, with intuitive rules regarding definitions. Proponents of ω-inconsistent revision semantics may adopt RTT's calculus $\mathbf{C}$, a very natural logic of definitions that I present below. But since $\mathbf{C}$ derives ω-inconsistencies from the Peano axioms given certain definitions, it is not available to ω-consistent revision theorists. Lack of a suitable proof system is a major flaw with all previously studied ω-consistent revision theories, such as $\mathbf{S}^*$. Fortunately for $\mathbf{S}^{\#N}$, however, $\mathbf{C}$ can be weakened to various proof systems $\mathbf{C}^N$, each sound for the corresponding system $\mathbf{S}^{\#N}$. Users of a language with revision strength N may reason with the corresponding calculus $\mathbf{C}^N$.

One of Gupta's [12] arguments for ω-inconsistency is that adopting $\mathbf{C}$, or some equivalent, appears necessary if one's proof system is to preserve the particularly compelling class of entailments known as *finite natural implication*. After presenting $\mathbf{C}$ and the calculi $\mathbf{C}^N$, I will explain and respond to Gupta's argument. Following Gupta, I propose to use finite natural implication as a test of proof systems' strength; however, I'll argue that, once we have a collection of proof systems $\mathbf{C}^N$ at our disposal, we can pass Gupta's test without committing to ω-inconsistency.

***6.1. The calculus* C.** The characteristic feature of deductions in $\mathbf{C}$ is that to each formula is attached an index $j \in \mathbb{Z}$, written as a superscript; e.g., A^j.[25] One can think of the indices as distinguishing between revision stages: a move from index j to $j+1$ indicates a move to the next stage in a revision sequence.

The inference rules for definitions illustrate the index policy. If G_i is defined by the explicit definition

$$G_i x =_{Df} \phi_{i,x} \tag{$\mathcal{D}_i$}$$

we have the following *Definiendum Introduction* and *Elimination* rules, for all indices j[26]:

[25] See pp. 157–162 of RTT for a Fitch-style presentation of $\mathbf{C}$, which is there called $\mathbf{C}_0$. Gupta [13, sec. 57] extends $\mathbf{C}$ for reasoning with generalized schematic definitions.

[26] The subscript in $\mathbf{Dfl}_r$ is for 'revision'.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

(**Dfl$_r$**) $\phi_{i,x}[t/x]^j$; therefore $G_i t^{j+1}$.
(**DfE$_r$**) $G_i t^{j+1}$; therefore $\phi_{i,x}[t/x]^j$.

Here $\phi_{i,x}[t/x]$ is the result of uniformly substituting a term t for all free occurrences of x in $\phi_{i,x}$. One can notice a tight correspondence between the definition rules and the definition's rule of revision.

All deductions of classical logic are legitimate in **C**, so long as the premises and conclusion bear the same index. Moreover, one may use versions of *Reductio* and Universal Generalization that relax the restrictions on indices. The rule supplying classical logic is:

Classical Logic (CL) (i) If the rules of classical first-order logic with identity allow the derivation of B from $A_1, \ldots, A_n$, then B^j may be derived from $A_1^j, \ldots, A_n^j$, for any integer j.
(ii) If $[B \wedge \neg B]^{j_{n+2}}$ is derivable from $A_1^{j_1}, \ldots, A_{n+1}^{j_{n+1}}$, then $[\neg A_{n+1}]^{j_{n+1}}$ is derivable from $A_1^{j_1}, \ldots, A_n^{j_n}$.
(iii) If $[A_{n+1}(z)]^{j_{n+1}}$ is derivable from $A_1^{j_1}, \ldots, A_n^{j_n}$ and the variable z is not free in $A_1, \ldots, A_n$, then $[\forall z A_{n+1}(z)]^{j_{n+1}}$ is derivable from $A_1^{j_1}, \ldots, A_n^{j_n}$.

Finally, we may shift the index of any formula not containing defined terms:

Index Shift (IS) If A does not contain any defined term, then, for any integers j_1 and j_2, A^{j_2} may be deduced from A^{j_1}.

We now define **C** as the proof system consisting of **Dfl$_r$**, **DfE$_r$**, **CL**, and **IS**. To define derivability in the non-indexed language, say that a sentence B is *derivable in* **C** from sentences $A_1, A_2, \ldots$ on the basis of definition $\mathcal{D}$ (notation: $A_1, A_2, \ldots \vdash^{\mathcal{D}} B$) iff, for some integer j, B^j can be derived from $A_1^j, A_2^j, \ldots$ using the rules of **C** (notation: $A_1^j, A_2^j, \ldots \vdash^{\mathcal{D}} B^j$). It is straightforward to check that **C** is sound for $\mathbf{S}^{\#}$ for explicit definitions.[27]

For an example of a deduction in **C**, we can use definition $\mathcal{D}_0$ from Example 2.1 to prove the theorem $Ha \rightarrow Hb$:

6.2. The calculi $\mathbf{C}^N$. For reasoning with explicit definitions in the system $\mathbf{S}^{\#N}$, we can use the calculus $\mathbf{C}^N$, a weakening of **C**. As before, an index j is attached to each

[27] A major difference between **S** and $\mathbf{S}^{\#}$ is that **C** is also complete for the former for explicit definitions, whereas the latter's entailment relation is not axiomatizable [19]. The main tradeoff between **S** and $\mathbf{S}^{\#}$ is thus between the axiomatizability of **S** and the added strength of $\mathbf{S}^{\#}$, which gives a richer and more determinate content to circularly defined predicates. The results in [19] also imply that entailment in $\mathbf{S}^{\#N}$ is not axiomatizable for any value N. Thus if one highly values a complete proof system, one may have reason to prefer **S** to both $\mathbf{S}^{\#N}$ and $\mathbf{S}^{\#}$.

I myself don't know of a good reason to think the logic of definitions is axiomatizable. Even if I did, I think the added strength of $\mathbf{S}^{\#N}$ and $\mathbf{S}^{\#}$ could easily outweigh the value of a complete proof system. My reason is that entailment in **S** relative to many schematic definitions—including $\mathcal{T}$ when syntax is arithmetized—is still not axiomatizable. Since truth is the main application of the Revision Theory, it seems reasonable to place more weight on the strength of the truth theory than on having a complete proof system for other kinds of definitions.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

1.	$\mid [Ha \wedge \neg Hb]^0$	Ass.
2.	$\mid Ha^0$	1; **CL**
3.	$\mid [(a = a \vee a = b) \wedge \neg Ha \wedge \neg Hb]^{-1}$	2; $\mathbf{DfE}_r$
4.	$\mid [(b = a \vee b = b) \wedge \neg Ha \wedge \neg Hb]^{-1}$	3; **CL**
5.	$\mid Hb^0$	4; $\mathbf{DfI}_r$
6.	$[\neg(Ha \wedge \neg Hb)]^0$	1-5; **CL**
7.	$[Ha \rightarrow Hb]^0$	6; **CL**

formula in a $\mathbf{C}^N$-deduction. Now, however, *no negative indices are allowed*; indices are always natural numbers. The rules for deductions in $\mathbf{C}^N$ are otherwise exactly as in **C**. The revision strength N determines how liberal our definition of derivability will be. We say that the sentence B of $\mathcal{L}^+$ is *derivable in* $\mathbf{C}^N$ from $A_1, A_2, \ldots$ on the basis of definition $\mathcal{D}$ (notation: $A_1, A_2, \ldots \overset{\mathcal{D}}{\underset{N}{\vdash}} B$) iff, for some natural number $j \leq N$, B^j can be derived using the rules of $\mathbf{C}^N$ from $A_1^j, A_2^j, \ldots$ (notation: $A_1^j, A_2^j, \ldots \overset{\mathcal{D}}{\underset{N}{\vdash}} B^j$). It is again straightforward to check that $\mathbf{C}^N$ is sound for $\mathbf{S}^{\#N}$ for explicit definitions.

If we wished to carry out the above sample derivation in $\mathbf{C}^N$, we would only need to increment each index by 1. (We assume $N \geq 1$; $N = 0$ is just classical logic.)

6.3. A test case: Finite natural implication. Gupta's desideratum of preserving finite natural implication concerns the specially well-behaved class of finite definitions. Our main example of a finite definition will be $\mathcal{D}_0$ from Example 2.1, which is in fact the only finite definition appearing in this paper. We characterize these definitions as follows. A hypothesis h is *finitely reflexive* for a revision rule ρ iff there is $n > 0$ (called h's *period*) such that $\rho^n(h) = h$. In the model M of Example 2.1, the hypotheses $\varnothing$ and $\{0, 1\}$ (and only these) are finitely reflexive, with period 2. A definition $\mathcal{D}$ is *finite* iff, for all ground models M there is a number n such that, for all hypotheses h, $\delta^n_{\mathcal{D},M}(h)$ is finitely reflexive. Note that all partial definitions (§1.1) are finite.

An attractive feature of finite definitions is that we can eliminate all "suboptimal" hypotheses after finitely many revision stages. In view of this fact, finite definitions can be given a natural semantics: we declare a sentence A categorically true in M relative to a finite definition $\mathcal{D}$ iff it is true on all hypotheses finitely reflexive for $\delta_{\mathcal{D},M}$. *Finite natural implication* is the corresponding implication relation restricted to finite definitions. That is, where $\mathcal{D}$ is finite, B is a finite natural implication of Γ (notation: $\Gamma \overset{\mathcal{D}}{\underset{f}{\models}} B$) iff for all ground models M,

$$\forall h \,(\text{if } h \text{is finitely reflexive for } \delta_{\mathcal{D},M}, \text{ then } M + h \models \Gamma \text{ implies } M + h \models B)\,.$$

As I understand it, Gupta's desideratum for a proof system $\mathbf{C}^\circ$ is:

(FNI) $\mathbf{C}^\circ$ preserves (i.e., is complete for) finite natural implication.

The next proposition, proved in the Appendix, compares **C** and the calculi $\mathbf{C}^N$ with respect to FNI.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

PROPOSITION 6.1 (Finite natural implication in $\mathbf{C}$ and $\mathbf{C}^N$).

(i) $\mathbf{C}$ *is complete for finite natural implication*;
(ii) *none of the individual calculi* $\mathbf{C}^N$ *is complete for finite natural implication*; *but*
(iii) *for each finite definition* $\mathcal{D}$, *some calculus* $\mathbf{C}^N$ *is complete for all finite natural implications relative to* $\mathcal{D}$.

I'll unpack the significance of these three claims in turn. Gupta's argument for ω-inconsistency from FNI proceeds by observing (i) and that FNI is not satisfied by any known proof system sound for an ω-consistent semantics. To illustrate the point, consider Example 2.1 again. The sentence $Ha \leftrightarrow Hb$ is a finite natural implication of the empty premise set, yet Table 1 implies that no proof system for $\mathbf{S}^*$ can preserve this entailment. Elsewhere Gupta [13, sec. 176] refers to this weakness of $\mathbf{S}^*$ as a "fatal flaw."

As stated, it is unclear whether the calculi $\mathbf{C}^N$ satisfy FNI: individually, they don't (claim (ii)), but, collectively, they do (claim (iii)). I suggest that the situation is ultimately favorable for these calculi: given my proposed understanding of revision strengths, claim (iii) is all we should desire. If this is true, finite natural implication does not in fact favor $\mathbf{C}$ over the calculi $\mathbf{C}^N$.

Whether this is plausible depends on the considerations motivating FNI. First, Gupta claims that a proof system should preserve finite natural implications because they are "intuitively compelling" [12, p. 190]. (Witness the straightforward proof of $Ha \rightarrow Hb$ on p. 17.) I agree that it would be a red flag if the entire apparatus of proof calculi $\mathbf{C}^N$ couldn't capture some finite natural implication. But when we have multiple calculi at our disposal, I see no problem with providing, for each finite definition, the revision strength appropriate to capture its finite natural implications.

Gupta's second motivation for FNI points out that finite definitions are needed for applications of the Revision Theory, such as to rational choice, and the applications require that we can reason with these definitions.[28] But when stipulating a set $\mathcal{D}$ of finite definitions for the purposes of some application, we can always stipulate a revision strength N for which $\mathbf{C}^N$ preserves finite natural implication. (We just need to say, "Let the revision strength of our dialect be the minimum strength for which finite natural implication is preserved.") Moreover, the finite definitions employed in applications can be easily transformed into the especially manageable form mentioned by Gupta [10]. His observations there show that from the syntactic form of such definitions one can read off an explicit revision strength that preserves finite natural implications. Even when users of the definition don't stipulate a strength, Agreement implies that, on the default strength N, $\mathbf{S}^{\#N}$ will agree with $\mathbf{S}^{\#}$ on categorical truth for sentences in use. From this fact one can show that finite natural implications on sentences used in the application are guaranteed to preserve categorical truth.[29] In summary: the calculi $\mathbf{C}^N$ provide for each job a tool that can handle it. We need not wish further for a single "supertool" that can perform every job.

[28] See fn. 1 for relevant citations.

[29] Agreement doesn't, as far as I know, guarantee that all finite natural implications will be *valid entailments* at the language's default strength. But Proposition 6.1 (iii) shows that one could modify Agreement so that default strengths indeed make all finite natural implications valid.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

§7. Truth. In this section I compare $\mathbf{S}^{\#}$'s and $\mathbf{S}^{\#N}$'s truth theories, that is, their semantical theories for truth predicates. We will see that both of these theories support strong compositional and disquotational principles, which in my view already make them among the best classical theories of truth. I argue that by meeting a third desideratum—preserving truth's generalization function— $\mathbf{S}^{\#N}$ pulls ahead of $\mathbf{S}^{\#}$ for languages with finite usage. Along the way, I will respond to Gupta's [12, 13] second argument for ω-inconsistency.

7.1. Compositional truth principles. *Compositional truth principles* say that a conjunction is true iff its conjuncts are; that the negation of a sentence is true iff the sentence is not true; and so on. These principles are key features of how the concept of truth functions in semantics. Unlike other leading classical truth theories, $\mathbf{S}^{\#}$ and $\mathbf{S}^{\#N}$ both validate all the compositional principles.[30] To make this claim precise, suppose $\langle M, \ulcorner\ \urcorner \rangle$ is a full sentential frame for $\mathcal{L}^{+}$, which contains a truth predicate. Further suppose that the ground language $\mathcal{L}$ contains the predicates $Neg(x, y)$, $Conj(x, y, z)$, $Ui(x, y)$, and $Ref(x, y)$, which M interprets as follows. For $d_1, d_2, d_3 \in D$,

- $\langle d_1, d_2 \rangle \in I(Neg)$ iff $d_1, d_2 \in \mathcal{L}^{+}$ and d_1 is the negation of d_2;
- $\langle d_1, d_2, d_3 \rangle \in I(Conj)$ iff $d_1, d_2, d_3 \in \mathcal{L}^{+}$ and d_1 is a conjunction of d_2 and d_3;
- $\langle d_1, d_2 \rangle \in I(Ui)$ iff $d_1, d_2 \in \mathcal{L}^{+}$ and d_1 is a universal quantification of d_2;
- $\langle d_1, d_2 \rangle \in I(Ref)$ iff d_1 is a closed term of $\mathcal{L}^{+}$ and d_1 refers to d_2 in M.

Then we have the desideratum:

(Compositional Principles) Where Neg, $Conj$, Ui, and Ref are interpreted as above in a full sentential frame $\langle M, \ulcorner\ \urcorner \rangle$, the following principles are categorical:

$$\forall x \forall y [Neg(x, y) \rightarrow (Tx \leftrightarrow \neg Ty)] \tag{C$\neg$}$$

$$\forall x \forall y \forall z [Conj(x, y, z) \rightarrow (Tx \leftrightarrow Ty \wedge Tz)] \tag{C$\wedge$}$$

$$\forall y \exists x Ref(x, y) \rightarrow \forall x [\exists y Ui(x, y) \rightarrow (Tx \leftrightarrow \forall y (Ui(x, y) \rightarrow Ty))]. \tag{C$\forall$}$$

Both $\mathbf{S}^{\#}$ and $\mathbf{S}^{\#N}$ satisfy Compositional Principles—provided, in the latter case, that the language's revision strength is at least 1. In a language where $\neg$, $\wedge$, and $\forall$ are the primitive logical constants (apart from identity), (C$\neg$)-(C$\forall$) are the core compositional principles. If we take some other combination of logical constants as primitive, both theories will render compositional principles for them categorical, too.

7.2. Blind truth ascriptions. As many philosophers have stressed, the truth predicate serves important expressive functions. One of these is to allow us to make so-called *blind truth ascriptions*, in which one ascribes truth to some sentence one isn't in a position to explicitly assert. As Horwich [18, pp. 2–3] writes:

> On occasion we wish to adopt some attitude towards a proposition—for example, believing it, assuming it for the sake of argument, or desiring that it be the case—but find ourselves thwarted by ignorance of what exactly the proposition is. We might know it only as 'what

[30] $\mathbf{S}^{*}$ and the axiomatic theories **KF** [6] and **VF** [3] fail to preserve compositional principles. In fact, none of (C$\neg$)-(C$\forall$) is categorical in $\mathbf{S}^{*}$—another indication of this theory's weakness. The revision theory **S** and the axiomatic theory **FS** [8] validate the compositional principles.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

> Oscar thinks' or 'Einstein's principle'; perhaps it was expressed, but not clearly or loudly enough, or in a language we don't understand; or—and this is especially common in logical and philosophical contexts—we may wish to cover infinitely many propositions (in the course of generalizing) and simply can't have all of them in mind. In such situations the concept of truth is invaluable. For it enables the construction of another proposition, intimately related to the one we can't identify, which is perfectly appropriate as the alternative object of our attitude.

What Horwich says about propositions also holds for sentences, the bearers of truth considered in this section. For now let us focus on the case of blind ascriptions of truth to single sentences; we will consider the case of generalizations in the next subsection.

In order for my blind ascription 'What Oscar said is true' to serve its intended purpose, it must be in some sense equivalent to the sentence Oscar said. A very weak form of equivalence is *co-affirmability*: we require that a blind ascription of truth to B is affirmable iff B itself is affirmable (cf. [15]). In the Revision Theory, it is plausible to say that a sentence is affirmable iff it is categorically true. Thus the following informal condition should be a desideratum for a revision theory:

(Blind Ascription) Where t_B is a singular term referring to the sentence B, any use of Tt_B is categorically true iff B is.

Blind Ascription corresponds to a formal constraint on revision systems. Let $\mathcal{L}^+$ be a language with a set of definitions $\mathcal{D}$ that includes the truth definition, and which is interpreted via a full sentential frame $\langle M, \ulcorner\ \urcorner\rangle$. Using $\models^{\mathcal{D}}_{\circ}$ to indicate categorical truth in an unspecified revision system, we have

$(T\models)$ $M \models^{\mathcal{D}}_{\circ} B$ iff $M \models^{\mathcal{D}}_{\circ} Tt_B$, where t_B is any singular term referring to B in M.

$\mathbf{S}^{\#}$ satisfies $(T\models)$ for all sentences B (RTT, Theorem 6C.1). Once again, the situation is more complicated with $\mathbf{S}^{\#N}$, but I will argue that it comes out in $\mathbf{S}^{\#N}$'s favor. The systems $\mathbf{S}^{\#N}$ do not satisfy $(T\models)$ in full generality. For a counterexample, we can use the language, ground model, and definitions from Example 2.3. As recorded in (3) and (4), $M \models^{\mathcal{D}}_{\#N} G\overline{(N-2)}$ even though $M \not\models^{\mathcal{D}}_{\#N} T\ulcorner G\overline{(N-2)}\urcorner$, whenever $N \geq 2$. Nevertheless, $\mathbf{S}^{\#N}$ can salvage a restricted principle. If a proponent of $\mathbf{S}^{\#N}$ accepts Agreement, she may claim that, in a language with default strength N, the system $\mathbf{S}^{\#N}$ satisfies $(T\models)$ for all truth ascriptions Tt_B that are ever used. (This fact follows from the more general Proposition 7.1, below.) As I formulated Blind Ascription, this is all the desideratum does and should demand. The expressive functions of truth only require $(T\models)$ for the sentences that are in use: any further instances don't help anyone express anything. An analogy is helpful: a refrigerator light has the function of making the contents of the refrigerator visible. But that function only requires the light to be on when the door is open. No fridge is defective because its light doesn't stay on 24/7.

As with refrigerator lights, economy in performing functions proves valuable in the case of truth: cutting down on excess instances of $(T\models)$ allows $\mathbf{S}^{\#N}$ to dodge Gupta's second argument for ω-inconsistency. McGee's Theorem states that any classical truth theory satisfying $(T\models)$ and (C$\neg$)-(C$\forall$) is ω-inconsistent when syntax is arithmetized

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

[20].[31] Gupta [13] claims that all of these principles are needed for truth to fully serve its expressive functions; hence, truth gives rise to ω-inconsistencies. But, as we have seen, the full principle $(T\models)$ is far stronger than needed for truth's expressive functions. In my view, $\mathbf{S}^{\#N}$ achieves the optimal response to McGee's Theorem: we maintain classical logic, ω-consistency, the semantical laws, and every instance of $(T\models)$ needed for truth's expressive functions.[32]

7.3. The generalization function. As already noted, truth's expressive functions allow us not just to blindly affirm single sentences, but to affirm infinitely many sentences in a single breath. If I want to affirm all the axioms of Peano Arithmetic, it's clearly impossible for me to do so by individually asserting each of the infinitely many axioms. I have no option but to use a generalizing device. The obvious solution is to say "all the axioms of Peano Arithmetic are true."

We face the same expressive problem in trying to affirm any infinite set of sentences. It is plausible that the truth predicate provides an all-purpose solution. By all appearances, we can use the truth predicate to affirm any infinite set of sentences, provided some way of specifying those sentences. This expressive capacity is one aspect of what is often called truth's *generalization function*.[33]

Suppose Γ is a set of sentences that we can specify with some predicate $\gamma(x)$ true of all and only the sentences in Γ. If truth performs the generalization function, saying

$$\text{Every sentence with the property } \gamma \text{ is true.} \qquad \text{(G)}$$

should be in some sense equivalent to individually affirming each sentence of Γ. Again, we can let co-affirmability be our guide in formulating a desideratum, yielding:

(Generalization) Where γ is a non-circular predicate true of precisely the sentences Γ, any use of (G) is categorically true iff all the sentences Γ are categorically true.

To arrive at a formal condition, let $\mathcal{L}$ be a language that is expanded with a truth predicate T (and perhaps other new predicates) to $\mathcal{L}^+$, and let $\langle M, \ulcorner\ \urcorner\rangle$ be a full sentential frame for $\mathcal{L}^+$. Then the generalization function requires of a revision system $\models_{\circ}^{\mathcal{D}}$:

---

31 These conditions are actually considerably stronger than those McGee uses.

32 $\mathbf{S}^{\#N}$ shares its approach to McGee's Theorem with the axiomatic truth theories $\mathbf{FS}_n$, which are ω-consistent weakenings of the ω-inconsistent truth theory **FS**. In fact, there is a very close relation between revision semantics and these truth theories: for each $n \in \mathbb{N}$, all the axioms of $\mathbf{FS}_n$ are categorical in $\mathbf{S}^{\#n}$—and all inference rules preserve categorical truth—when syntax is arithmetized and truth is defined by $\mathcal{T}$ in the standard model of arithmetic. All axioms and inference rules of **FS** are categorical (preserve categorical truth) in $\mathbf{S}^{\#}$, as well as in **S**. See Halbach [16, chap. 14] for details.

33 See Quine [22, p. 12] for a famous statement of truth's expressive capacity that I am calling attention to. As Quine observes, the truth predicate more generally grants us the expressive effects of quantifying into sentence position. Here I only investigate the paradigm case of universal generalizations, in which $\mathbf{S}^{\#N}$ already surpasses $\mathbf{S}^{\#}$. Sometimes the generalization function is associated with considerably stronger claims—for instance, that the capacity to affirm infinite numbers of sentences *requires* a truth predicate, that truth is nothing more than a generalizing device, or that some equivalence stronger than co-affirmability obtains. I won't take a stand on these claims here.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

(Generalization (Formal)) If the $\mathcal{L}$-formula $\gamma(x)$ is true of precisely the set of $\mathcal{L}^+$-sentences Γ in M, then $M \models^{\mathcal{D}}_{\circ} \Gamma$ iff $M \models^{\mathcal{D}}_{\circ} \forall x(\gamma(x) \rightarrow Tx)$.

Truth in $\mathbf{S}^{\#}$ does not always satisfy Generalization (Formal), on account of ω-inconsistencies. To see this we can again use the language $\mathcal{L}^+$, ground model M, and definitions $\mathcal{D}_1$ of Example 2.3. Someone who uses $\mathcal{L}^+$ might notice that all the sentences in the set $\Gamma = \{G0, G0', \ldots\}$ are categorically true. She might then reasonably wish to affirm all of Γ. Of course, to do so she needs some generalizing device. If the truth predicate T performs the generalization function, then she can affirm all these sentences by asserting

$$\forall x(\exists j\ x = g(j) \rightarrow Tx). \tag{GG}$$

Colloquially, we'd say, "the G-predication of every number is true." Since the formula $\exists j\ x = g(j)$ is true of precisely the sentences in Γ, (GG) is a way of saying that the sentences in Γ are true. However, (GG) is categorically *false* according to $\mathbf{S}^{\#}$; this is a failure of Generalization (Formal).

It is easy to see that the systems $\mathbf{S}^{\#N}$, too, don't always satisfy Generalization (Formal). Whenever $N \geq 2$, all the sentences $G0, G0', \ldots, G\overline{(N-2)}$ are categorically true (see (3)); and the formula $\exists y\ (y \leq \overline{(N-2)}\ \wedge\ x = g(y))$ is true in M of precisely these sentences. But the generalization

$$\forall x[\exists y(y \leq \overline{(N-2)} \wedge x = g(y)) \rightarrow Tx]$$

("all G-predications of numbers less than or equal to $N-2$ are true") is pathological according to $\mathbf{S}^{\#N}$. The crucial advantage of the theory $\mathbf{S}^{\#N}$ over $\mathbf{S}^{\#}$, though, is that the informal requirement Generalization only ought to apply to sentences that are used by speakers to make generalizations. By appealing to Agreement, $\mathbf{S}^{\#N}$ can once again salvage a restricted principle and save the generalization function. If finitely many sentences of $\mathcal{L}^+$ are used and Agreement yields default strength N, then $\mathbf{S}^{\#N}$ satisfies Generalization (Formal) for every generalization that is ever used. Proposition 7.1 makes this claim precise.

PROPOSITION 7.1. *Let $\mathcal{L}^+$ be a ground language expanded with any set $\mathcal{D}$ of definitions that includes the truth definition $\mathcal{T}$, and let $\langle M, \ulcorner\ \urcorner\rangle$ be a full sentential frame for $\mathcal{L}^+$. Suppose the $\mathcal{L}$-formula $\gamma(x)$ defines the set Γ of $\mathcal{L}^+$-sentences in M, and suppose N is such that $\mathbf{S}^{\#N}$ and $\mathbf{S}^{\#}$ agree on whether $\forall x(\gamma(x) \rightarrow Tx)$ is categorically true in M relative to $\mathcal{D}$. Then also*

$$M \models^{\mathcal{D}}_{\#N} \Gamma \textit{ iff } M \models^{\mathcal{D}}_{\#N} \forall x(\gamma(x) \rightarrow Tx).$$

Proof. Straightforward, using the fact that $M \models^{\mathcal{D}}_{\#N} B$ iff $M \models^{\mathcal{D}}_{\#(N+1)} T\ulcorner B\urcorner$. □

At first one might have thought that, even if Agreement ensures that truth in $\mathbf{S}^{\#N}$ performs well on the sentences actually in use, once we start generalizing over unused sentences the theory will run into problems. But Proposition 7.1 shows that exactly the opposite is true: it is precisely on generalizations like (GG) that $\mathbf{S}^{\#N}$ performs well and $\mathbf{S}^{\#}$ does not.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

§8. Infinite usage.

8.1. A generalization of Agreement. The previous section's argument involved the assumption that linguistic usage is finite. In this section I briefly discuss default strengths and the expressive functions of truth for languages whose usage is infinite.

First, I'll show that these languages present no particular problem for setting default strengths, by suitably generalizing Agreement. To settle on a principle, I'll begin with some examples. The most obvious test case is the language $\mathcal{L}^+$ from Example 2.3 with the defined predicates G and T. Imagine now that its speakers use all the sentences $\exists i \neg Gi, G0, G0', \ldots$. If these are the only sentences used, the most natural choice of a revision strength seems simply to be 0. Trying to maximize the number of sentences whose semantical status agrees with $\mathbf{S}^{\#}$ no longer singles out any revision strength. In the absence of good reasons to pick one revision strength over any other, we break the tie by choosing the least candidate.

Now imagine that $\mathcal{L}^+$ contains, in addition to G and T, a new predicate H, which is given definition $\mathcal{D}_0$ from Example 2.1. (This time use '0' and '$\overline{1}$' for 'a' and 'b', respectively.) The speakers of $\mathcal{L}^+$ use all the sentences $\exists i \neg Gi, G0, G0', \ldots$, plus the sentence $H0 \leftrightarrow H\overline{1}$. Now it seems that $\mathcal{L}^+$'s revision strength should be at least 1, so as to yield optimal results both for $\exists i \neg Gi$ and $H0 \leftrightarrow H\overline{1}$. Beyond strength 1, however, there is complete symmetry between the other candidate strengths; we break the tie by choosing the least candidate, 1.

We can generalize the idea as follows. Define the *stabilization point* $st(B)$ of a sentence B as the least N for which B is either categorically true or categorically false in $\mathbf{S}^{\#N}$, if such an N exists; and 0 otherwise. Even in cases of infinite usage, speakers presumably have a tendency to use sentences with low stabilization points. Our guiding thought is that all sentences reflecting this tendency should stabilize. There may be several ways to cash this idea out. Here is one: we set the revision strength as the largest number at which "especially many" sentences stabilize. More precisely, we stipulate that the strength is the largest n such that, for all $m > n$, more of the used sentences stabilize at n than at m; and it is 0 if there is no such n. Formally, this is:

(Infinite Agreement) Let Γ be the totality of $\mathcal{L}^+$-sentences used by its speakers, and let $\Gamma_n = \{B \in \Gamma : st(B) = n\}$. Then the default revision strength for $\mathcal{L}^+$ is $\max(\{0\} \cup \{n \in \mathbb{N} : \forall m > n \ |\Gamma_n| > |\Gamma_m|\})$.

Infinite Agreement entails both Agreement and the strengths suggested in the previous two scenarios. Alternative or more sophisticated versions of Infinite Agreement may be desirable, but for now we at least have a promising proof of concept.

8.2. Truth in languages with infinite usage. When we consider infinite cases, it is less plausible that $\mathbf{S}^{\#N}$ performs truth's functions strictly better than $\mathbf{S}^{\#}$. One can easily construct cases where $\mathbf{S}^{\#}$ but not $\mathbf{S}^{\#N}$ satisfies the expressive functions for all truth ascriptions in use. Consider again $\mathcal{L}^+$ with defined predicates G and T, and suppose the only sentences used are $G0, G0', \ldots$ and $T^{\ulcorner}G0^{\urcorner}, T^{\ulcorner}G0'^{\urcorner}, \ldots$. $\mathbf{S}^{\#}$ satisfies Blind Ascription for all the truth predications, but for no choice of N can the system $\mathbf{S}^{\#N}$ match this feat. This fact somewhat dampens §7's argument.

Nevertheless, there is a reasonable case that $\mathbf{S}^{\#}$ does not pull ahead of $\mathbf{S}^{\#N}$ even when we consider infinite usage. Failures of the expressive functions in the two theories go hand in hand, in a certain sense. The next proposition shows that, whenever $\mathbf{S}^{\#N}$

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

inevitably fails to satisfy Blind Ascription on some infinite set of sentences Δ, there is a natural generalization about Δ for which $\mathbf{S}^{\#}$ fails Generalization. The upshot is that, if we were to devise some way of "counting" failures of expressive functions, there's no reason to think that $\mathbf{S}^{\#N}$ would have the higher count.

PROPOSITION 8.1. *Let* $\langle M, \ulcorner\ \urcorner\rangle$ *be a full sentential frame for a language* $\mathcal{L}$. *Expand* $\mathcal{L}$ *to* $\mathcal{L}^{+}$ *with predicates defined by the set of definitions* $\mathfrak{D}$; *in particular,* $\mathfrak{D}$ *defines the predicate* T *by the truth definition* $\mathcal{T}$. *Suppose* Δ *is a set of* $\mathcal{L}^{+}$*-sentences such that, for all* $N \in \mathbb{N}$, *it's not the case that for all* $B \in \Delta$,

$$M \models^{\mathfrak{D}}_{\#N} B \text{ iff } M \models^{\mathfrak{D}}_{\#N} T\ulcorner B\urcorner.$$

Then there is a set $\Gamma \subseteq \Delta$ *such that, if an* $\mathcal{L}$*-formula* $\gamma(x)$ *defines* Γ *in* M, *then it's not the case that*

$$M \models^{\mathfrak{D}}_{\#} \Gamma \text{ iff } M \models^{\mathfrak{D}}_{\#} \forall x(\gamma(x) \rightarrow Tx).$$

Proof. Suppose $(B_n)_{n\in\mathbb{N}}$ is such that $M \models^{\mathfrak{D}}_{\#n} B_n$ and $M \models^{\mathfrak{D}}_{\#n} T\ulcorner B_n\urcorner$ have different truth values for all n. Since $M \models^{\mathfrak{D}}_{\#n} T\ulcorner B\urcorner$ implies $M \models^{\mathfrak{D}}_{\#n} B$, we have, for all n, (i) $M \models^{\mathfrak{D}}_{\#n} B_n$ and (ii) $M \not\models^{\mathfrak{D}}_{\#n} T\ulcorner B_n\urcorner$. Where $\Gamma = \{B_n : n \in \mathbb{N}\}$, (i) implies $M \models^{\mathfrak{D}}_{\#} \Gamma$. However, if $\gamma(x)$ defines Γ in M, (ii) implies $M \not\models^{\mathfrak{D}}_{\#} \forall x(\gamma(x) \rightarrow Tx)$. □

In the end, I don't think infinite cases significantly alter my verdict about truth. If one is inclined to place more weight on the case of finite usage, one should still judge $\mathbf{S}^{\#N}$'s truth theory superior to $\mathbf{S}^{\#}$'s. Even if not, one should probably judge the theories to be on a par.

§9. Simplicity. At this point, $\mathbf{S}^{\#}$'s last defense may seem to lie in the charge that $\mathbf{S}^{\#N}$ is, in some way, unnecessarily complicated. There are a few possible routes to this conclusion. Most obviously, the $\mathbf{S}^{\#N}$ theory commits to a new notion of revision strengths, and if we include claims like Agreement in its formulation, $\mathbf{S}^{\#N}$ takes longer to state. But no one has proposed a good measure of the number of "notions" in a theory, and a theory's length is highly language-dependent.

Perhaps more promising is to claim that $\mathbf{S}^{\#}$ is more parsimonious in its theory of propositions. On certain fine-grained views of propositions, adding revision strengths seems to complicate the space of propositions. Consider how to develop a theory of propositions, properties, and relations (PRPs) for $\mathbf{S}^{\#}$. The most natural view will probably say, for instance, that a liar sentence $\lambda = \ulcorner\neg T\lambda\urcorner$ expresses the proposition *that* λ *is not true*. This proposition, like the sentence $\neg T\lambda$, is neither categorically true nor categorically false. On this view, truth will be a "circular" property, in the sense that it cannot be assigned a classical extension.[34] Suppose we adopt a very fine-grained view of propositions, on which the structure and individuation of propositions roughly

[34] Orilia [21], the most carefully developed revision theory of PRPs, lends itself to this view. Orilia's theory endows PRPs with maximally fine-grained identity conditions, though the identity conditions are not essential to his proposal. An alternative view would hold that paradoxicality is linguistic in nature. Sentences may fail to have a categorical truth value, but every proposition is either categorically true or categorically false. I'm not sure of the philosophical promise of this approach. As far as I'm aware, it hasn't been pursued in the Revision Theory literature.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

mirror the structure and individuation of sentences of a formal language that express them. On this view, for each circular definition of G there corresponds one—but only one—proposition expressed by '$G0$'.

By contrast, proponents of $\mathbf{S}^{\#N}$ should probably say that, even holding the definition of G fixed, '$G0$' could express any of infinitely many propositions—one for each revision strength. To illustrate why, suppose G is defined according to $\mathcal{D}_1$ (Example 2.3) in both $\mathcal{L}_1^+$ and $\mathcal{L}_2^+$, and suppose the revision strengths of these languages are 1 and 2, respectively. Then '$G0$' will be categorically true in $\mathcal{L}_2^+$ but not in $\mathcal{L}_1^+$, so that this sentence cannot express the same proposition in both languages. On the most natural view, given $\mathcal{D}_1$, the language's revision strength determines which of infinitely many distinct propositions '$G0$' expresses. These propositions will also be distinct from the propositions expressed by '$G0$' under any other definition, since we are individuating propositions finely. The result is to multiply our space of circular propositions infinitely many times over, once for each revision strength.

In some sense this proliferation in propositions strikes me as an increase in $\mathbf{S}^{\#N}$'s complexity. But I should temper the point with two remarks. First, it is still rather obscure in what sense $\mathbf{S}^{\#N}$'s theory of propositions would be more complex. Of course, the *cardinality* of the class of all propositions is the same on both views. Second, I assumed quite a fine-grained individuation of propositions. At the other extreme, one could identify propositions with functions from possible worlds to values in the set {categorically true, categorically false, pathological}. In that case we could use the same space of functions as propositions for both $\mathbf{S}^{\#}$ and $\mathbf{S}^{\#N}$.

Other aspects of $\mathbf{S}^{\#N}$ may lead us to judge that it is simpler after all. $\mathbf{S}^{\#N}$ tends to support simple and pleasing generalizations where $\mathbf{S}^{\#}$ lacks them. Most obviously, only $\mathbf{S}^{\#N}$ is ω-consistent and satisfies principles like (K) (from §4), weak reductionism, and Categorical Truth Preservation. Simplicity in the principles entailed by a theory seems to indicate simplicity in the theory itself.

When I step back, my judgments of simplicity exhibit a kind of bistability: by attending to different features, I can make myself feel that either theory is simpler than the other. And there is not, in my view, any sufficiently plausible general theory of simplicity that could decide between the competing factors. As things stand, we should not give much weight to such precarious judgments in choosing a semantical theory.[35] As far as I can tell, the desideratum of simplicity comes out as a wash.

§10. Summary: The scorecard. Table 2 summarizes my conclusions about my nine desiderata: Strong Conservativeness, ω-consistency, Ground Logic Preservation, Categorical Truth Preservation, FNI (finite natural implication), Compositional Principles, Blind Ascription, Generalization, and simplicity. I've marked Ground Logic Preservation and Categorical Truth Preservation together with $>$, since $\mathbf{S}^{\#}$ will fail one or the other, depending on whether entailment is defined locally or globally. My markings for FNI and Blind Ascription and Generalization reflect my conclusions in §§6.3, 7, and 8.2. The '$\approx$' marking for Simplicity reflects the previous section's

[35] Even when judgments of simplicity *are* clear, one might be skeptical that simplicity is an important consideration in choosing a semantic theory, on the grounds that we should not expect highly non-fundamental sciences like semantics to have simple laws.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

Table 2. *A scorecard for* $\mathbf{S}^{\#N}$ *and* $\mathbf{S}^{\#}$

		$\mathbf{S}^{\#N}$? $\mathbf{S}^{\#}$
Categorical truth	Strong Conservativeness	=
	ω-consistency	>
Entailment	Ground Logic Preservation and Categorical Truth Preservation	>
Proof system	FNI (finite natural implication)	=
Truth	Compositional Principles	=
	Blind Ascription and Generalization	≥
Simplicity		≈

uncertainty. My comparison of these two theories naturally hasn't been exhaustive. But as things stand, $\mathbf{S}^{\#N}$ is more plausible than $\mathbf{S}^{\#}$ in some respects and at least as plausible in the others. Tentatively, I conclude that $\mathbf{S}^{\#N}$ is the superior revision theory. Since a case can be made for $\mathbf{S}^{\#}$ as the best ω-inconsistent theory—for instance, it attributes the most content to circular predicates—a case can also be made that $\mathbf{S}^{\#N}$ is the best known revision theory.

Stepping back further, I think $\mathbf{S}^{\#N}$ strengthens the overall case for the Revision Theory. If there were not a viable ω-consistent revision theory, I think this should decrease one's confidence in the overall view. But the Revision Theory need not shoulder the burden of ω-inconsistency. Not only does the move to an ω-consistent revision theory come at no significant cost, one even finds unexpected benefits—such as in the areas of truth and entailment.

§A. Proofs. Throughout the Appendix, I'll employ several abbreviations. Given a ground model $M = \langle D, I\rangle$ for $\mathcal{L}_<$, I'll write: $d_1 < d_2$ for $\langle d_1, d_2\rangle \in I(<)$, $[d_1, d_2]$ for $\{d \in D : d_1 \leq d \leq d_2\}$, 0 for $I(0)$, n_M for $I(')^n(0)$, d' for $I(')(d)$, and $d^{(n)}$ for $I(')^n(d)$. Let us say that $X \subseteq D$ is *closed* iff $0 \in X$ and $(\forall d \in X)(d' \in X)$. When writing a revision rule $\delta_{\mathcal{D},M}$, I'll omit the subscripts when they are clear from context.

Proposition A.1 (Contradictory categorical truths in $\mathbf{S}^{\#}$). *Let* $\mathcal{L}_<$, $\mathbb{N}_<$, $\mathcal{D}_2$, *and* AX *be as in Example* 5.1, *and let* $\Gamma = \{AX, G0, G0', G0'', \ldots, \neg(G \text{ is closed})\}$. *Then*

(i) $\Gamma \models^{\mathcal{D}_2}_{\#} \forall x\ x \neq x$; *and*
(ii) $\mathbb{N}_< \models^{\mathcal{D}_2}_{\#} B$ *for all* $B \in \Gamma$.

Before proving the proposition, we'll need to prove a few lemmas.

Fact A.2. *Let* $M(= \langle D, I\rangle)$ *be a model for* $\mathcal{L}_<$. *If* h *is closed and* $h \neq D$, h *is a fixed point of* $\delta_{\mathcal{D}_2,M}$.

Lemma A.3. *Let* M *be a model for* $\mathcal{L}_<$ *such that* $M \models AX$, *and let* $\mathcal{S}$ *be a revision sequence for* $\delta_{\mathcal{D}_2,M}$. *Suppose* $\mathcal{S}_\alpha$ *is not closed, and there is* $d_0 \in D$ *such that*

(a) $d_0 \in \mathcal{S}_\alpha$ *and* $d_0' \notin \mathcal{S}_\alpha$; *and*
(b) $\forall d \in D\ (d < d_0 \implies d \in \mathcal{S}_\alpha)$.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

Where

$$X = \{d \in D : d < d_0\}$$
$$Y = \{d \in D : \exists n\ d = d_0^{(n)}\},$$

$\mathcal{S}_{\alpha+\omega} = X \cup Y$. *Moreover,* $\mathcal{S}_{\alpha+\omega}$ *is closed.*

Proof. Let M, $\mathcal{S}$, α, and d_0 be as specified. First, we show by induction on n that for all n, $1 \leq n < \omega$,

$$\mathcal{S}_{\alpha+n} = X \cup [d_0, d_0^{(n)}], \text{ and} \tag{A.1}$$
$$\forall d \in D\ (d < d_0^{(n)} \implies d \in \mathcal{S}_{\alpha+n}). \tag{A.2}$$

- *Base*: $n = 1$. First we show (A.1). From (AX4),
$$d < d_0' \implies d < d_0 \vee d = d_0 \tag{A.3}$$
for all $d \in D$. Then (a) and (b) imply that all $d \in X \cup \{d_0, d_0'\}$ satisfy G's definiens. Hence $X \cup \{d_0, d_0'\} \subseteq \mathcal{S}_{\alpha+1}$. From AX we can deduce that if $d \notin X \cup \{d_0, d_0'\}$, then $d_0' < d$. Consequently, d doesn't satisfy G's definiens in $M + \mathcal{S}_\alpha$, so $d \notin \mathcal{S}_{\alpha+1}$. So $\mathcal{S}_{\alpha+1} = X \cup \{d_0, d_0'\}$. Statement (A.2) follows from (A.3).
- *Inductive step*. Suppose (A.1) and (A.2) hold for n. We first show (A.1) for $n+1$. Suppose $d \in X \cup [d_0, d_0^{(n)}]$, and let $d_1 < d$. By (AX3), $d_0^{(i)} < d_0^{(j)}$ whenever $i < j$; hence transitivity implies $d_1 < d_0^{(n)}$. Equation (A.2) then implies $d_1 \in \mathcal{S}_{\alpha+n}$. This suffices to show that $d \in \mathcal{S}_{\alpha+n+1}$. Now consider $d_0^{(n+1)}$. By (AX4),
$$d_1 < d_0^{(n+1)} \implies d_1 < d_0^{(n)} \vee d_1 = d_0^{(n)},$$
and thus $d_1 \in \mathcal{S}_{\alpha+n}$. This shows that $d_0^{(n+1)} \in \mathcal{S}_{\alpha+n+1}$.

  Now suppose $d \notin X \cup [d_0, d_0^{(n+1)}]$. From AX we can deduce that $d_0^{(n+1)} < d$. But since $d_0^{(n+1)} \notin \mathcal{S}_{\alpha+n}$, d doesn't satisfy the definiens in $M + \mathcal{S}_{\alpha+n}$; so $d \notin \mathcal{S}_{\alpha+n+1}$. We've thus shown that $\mathcal{S}_{\alpha+n+1} = X \cup [d_0, d_0^{(n+1)}]$.

  Second, we show (A.2). According to (AX4), $d < d_0^{(n+1)}$ implies that either $d < d_0^{(n)}$ or $d = d_0^{(n)}$. In either case $d \in \mathcal{S}_{\alpha+n}$ by the inductive hypothesis, and $\mathcal{S}_{\alpha+n} \subset \mathcal{S}_{\alpha+n+1}$. This completes the induction.

From (A.1) it follows that all $d \in X \cup Y$ are stably in $\mathcal{S}$ at $\alpha + \omega$, and all $d \notin X \cup Y$ are stably out of $\mathcal{S}$ at $\alpha + \omega$. So $\mathcal{S}_{\alpha+\omega} = X \cup Y$.

To show that $\mathcal{S}_{\alpha+\omega}$ is closed, suppose first that $d \in X$. Then either $d' < d_0$ or $d' = d_0$, both of which imply $d' \in \mathcal{S}_{\alpha+\omega}$. Alternatively, if $d \in Y$, then $d' \in Y$ by definition of Y. □

The next three lemmas illustrate the different behaviors of the revision rules for $\mathcal{D}_2$ in standard and nonstandard models.

Lemma A.4. *Let* M $(= \langle D, I\rangle)$ *be a model for* $\mathcal{L}_<$ *such that* $M \models AX$ *and* M *is not isomorphic to* $\mathbb{N}_<$ $(M \not\cong \mathbb{N}_<)$. *If* $\mathcal{S}_\alpha = \{0\}$ *in a revision sequence* $\mathcal{S}$ *for* $\delta_{\mathcal{D}_2, M}$, *then* $\mathcal{S}_{\alpha+\omega}$ *is closed and is a proper subset of* D.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

Proof. Let M, $\mathcal{S}$, and α satisfy the above conditions. Letting $d_0 = 0$, we use Lemma A.3 to find that $\mathcal{S}_{\alpha+\omega} = \{d \in D : \exists n \ \ d = n_M\}$, which is closed. It is straightforward to show that any nonstandard model of AX contains nonstandard numbers larger than all elements n_M. Thus $\mathcal{S}_{\alpha+\omega} \subset D$. □

LEMMA A.5. *Let M $(= \langle D, I\rangle)$ be a model for $\mathcal{L}_<$ such that $M \models AX$ and $M \ncong \mathbb{N}_<$. Then h is recurring for $\delta_{\mathcal{D}_2,M}$ only if h is closed and $h \neq D$.*

Proof. We show that if h is either not closed or $h = D$, then h is not recurring. Suppose $h = \mathcal{S}_\alpha$ in some revision sequence $\mathcal{S}$ for δ.

- *Case 1:* h is not closed. We break into two sub-cases.
  - *Case 1.1:* There is $d_0 \in D$ such that
    (a) $d_0 \in h$ and $d_0' \notin h$; and
    (b) $\forall d \in D\ (d < d_0 \implies d \in h)$.
    By Lemma A.3, $\mathcal{S}_{\alpha+\omega}$ is closed (and $\neq \mathcal{S}_\alpha$). If $\mathcal{S}_{\alpha+\omega} \neq D$, then it is a fixed point (Fact A.2), which implies that $h = \mathcal{S}_\alpha$ is not recurring. If $\mathcal{S}_{\alpha+\omega} = D$, then $\mathcal{S}_{\alpha+\omega+1} = \varnothing$, and $\mathcal{S}_{\alpha+\omega+2} = \{0\}$. Lemma A.4 tells us that $\mathcal{S}_{\alpha+\omega\cdot 2}$ is closed and is a proper subset of D. By Fact A.2, it is a fixed point, and therefore h is not recurring.
  - *Case 1.2:* There is no $d_0 \in D$ satisfying (a) and (b). If $0 \notin \mathcal{S}_\alpha$, we find that $\mathcal{S}_{\alpha+1} = \{0\}$. Lemma A.4 implies that $\mathcal{S}_{\alpha+\omega}$ is closed and a proper subset of D; it is therefore a fixed point, and h is not recurring. So suppose that $0 \in \mathcal{S}_\alpha$. We will now show that $\mathcal{S}_{\alpha+1}$, which is
    $$\{d \in D : \forall d_1 (d_1 < d \implies d_1 \in \mathcal{S}_\alpha)\},$$
    is closed and a proper subset of D. It will then follow that $\mathcal{S}_{\alpha+1}$ is a fixed point, and h is not recurring.
    To show that $\mathcal{S}_{\alpha+1}$ is closed, note first that 0 and $0'$ are both in $\mathcal{S}_{\alpha+1}$. Now suppose $d \in \mathcal{S}_{\alpha+1}$, $d \neq 0$, and suppose $d_1 < d'$. By (AX4), either (i) $d_1 = d$ or (ii) $d_1 < d$. If (i), d_1 must be in $\mathcal{S}_\alpha$, else d_1's predecessor would satisfy (a) and (b). (Use the fact that $d_1 = d \in \mathcal{S}_{\alpha+1}$.) If (ii), $d \in \mathcal{S}_{\alpha+1}$ implies $d_1 \in \mathcal{S}_\alpha$. Hence for all $d_1 < d'$, $d_1 \in \mathcal{S}_\alpha$; therefore $d' \in \mathcal{S}_{\alpha+1}$. This establishes that $\mathcal{S}_{\alpha+1}$ is closed. To show that $\mathcal{S}_{\alpha+1} \neq D$, suppose otherwise for contradiction. Then the revision rule implies $\mathcal{S}_\alpha = D$, contradicting the assumption that h is not closed.
- *Case 2:* $h = D$. Then $\mathcal{S}_{\alpha+1} = \varnothing$, and $\mathcal{S}_{\alpha+2} = \{0\}$. Lemma A.4 implies that $\mathcal{S}_{\alpha+\omega}$ is closed and a proper subset of D. By Fact A.2 it is a fixed point, so h is not recurring.

□

LEMMA A.6. *If M $(= \langle D, I\rangle)$ is isomorphic to $\mathbb{N}_<$, for all hypotheses h and $n \geq 2$, there exists k such that $\delta^n_{\mathcal{D}_2,M}(h) = [0, k_M]$.*

Proof. Suppose $M \cong \mathbb{N}_<$. Let h be an arbitrary hypothesis, and suppose first that $h = D$. Using the isomorphism, one verifies that $\delta(h) = \varnothing$, and $\delta^n(h) = [0, (n-2)_M]$ for all $n \geq 2$.

Now suppose that $h \neq D$. Then h must not be closed. We may then deduce that $\delta(h) = [0, k_M]$ for some $k \geq 0$. We can further verify that for all $n \geq 2, \delta^n(h) = [0, (k + n - 2)_M]$. □

We're finally in a position to prove Proposition A.1.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

Proof.

(i) This is equivalent to the claim that for all M

$$\exists n \forall h \text{ recurring for } \delta \ \left(M + \delta^n(h) \not\models \Gamma\right).$$

There are two cases, depending on M.

- *Case 1:* $M \cong \mathbb{N}_<$. Let h be recurring for δ, and let $n \geq 2$. By Lemma A.6, $\delta^n(h) = [0, k_M]$ for some k. Thus $M + \delta^n(h) \not\models G\overline{k}'$.
- *Case 2:* $M \ncong \mathbb{N}_<$. Assume $M \models AX$. Let h be recurring, which by Lemma A.5 implies that h is closed and $h \neq D$. By Fact A.2, for all $n \geq 0, \delta^n(h) = h$; hence $M + \delta^n(h) \not\models \neg(G \textit{ is closed})$.

(ii) Clearly $\mathbb{N}_< \models AX$. Let h be an arbitrary hypothesis for δ. By Lemma A.6, $\delta^2(h) = [0, k]$ for some k. This already establishes that $\mathbb{N}_< \models_{\#}^{\mathfrak{D}_2} \neg(G \textit{ is closed})$. We can further calculate that for all $n \geq 2$, $\delta^n(h) = [0, k + n - 2]$. Consequently, for all $n \geq m + 2$, $\mathbb{N}_< + \delta^n(h) \models G\overline{m}$, and thus $\mathbb{N}_< \models_{\#}^{\mathfrak{D}_2} G\overline{m}$ for all m.

□

Proposition 6.1 (Finite natural implication in $\mathbf{C}$ and $\mathbf{C}^N$).

(i) $\mathbf{C}$ *is complete for finite natural implication*;
(ii) *none of the individual calculi* $\mathbf{C}^N$ *is complete for finite natural implication*; *but*
(iii) *for each finite definition* $\mathfrak{D}$, *some calculus* $\mathbf{C}^N$ *is complete for all finite natural implications relative to* $\mathfrak{D}$.

Proof. (i): $\mathbf{C}$ is sound and complete for the weak revision system $\mathbf{S}_0$ (RTT, Theorem 5B.1), which is easily seen to preserve all finite natural implications.

(ii): Fix $N \in \mathbb{N}$; we construct a counterexample to FNI in $\mathbf{C}^N$. We will use the language $\mathcal{L}_<$ and abbreviation AX from Example 5.1, as well as the notation from the proof of Proposition A.1. Define the new predicate G explicitly by

$$Gx =_{Df} AX \wedge x \leq \overline{N} \wedge \forall y(y < x \rightarrow Gy). \qquad (\mathfrak{D}_3)$$

We claim that

(a) $\mathfrak{D}_3$ is finite;
(b) $\models_f^{\mathfrak{D}_3} AX \rightarrow G\overline{N}$; and
(c) $\not\models_N^{\mathfrak{D}_3} AX \rightarrow G\overline{N}$.

For (a) and (b), we investigate the behavior of the revision rule $\delta_{\mathfrak{D}_3, M}$ in models M of AX. One can verify the following:

(d) If $[0, N_M] \subseteq h$, then $\delta(h) = [0, N_M]$.
(e) Otherwise, let n_M be the least element of D excluded from h. Then $\delta(h) = [0, n_M]$.

From (d) and (e), we can see that for any h, $[0, N_M] \subseteq \delta^{N+1}(h)$, and that for all $n \geq N + 1, \delta^{N+1}(h) = [0, N_M]$. The hypothesis $[0, N_M]$ is finitely reflexive (with period 1). If M is not a model of AX, then $\delta(h) = \varnothing$ for all h; and $\varnothing$ is finitely reflexive (again with period 1). Thus we conclude that $\mathfrak{D}_3$ is finite.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

In models M of AX, $[0, N_M]$ is in fact the only reflexive hypothesis. Thus, in these models, $M \models_f^{\mathcal{D}_3} AX \rightarrow G\overline{N}$. If M is not a model of AX, then also $M \models_f^{\mathcal{D}_3} AX \rightarrow G\overline{N}$. This establishes (b).

For (c), we appeal to the semantical systems $\mathbf{S}_{1a(N)}$, which stand to $\mathbf{S}_0$ as the systems $\mathbf{S}^{\#N}$ stand to $\mathbf{S}^{\#}$. Implication in $\mathbf{S}_{1a(N)}$ is defined as follows, for $N \in \mathbb{N}$: $\Gamma \models_{1a(N)}^{\mathcal{D}} B$ iff, for all ground models M,

$$\forall h \left(M + \delta^N_{\mathcal{D},M}(h) \models \Gamma \text{ implies } M + \delta^N_{\mathcal{D},M}(h) \models B \right).$$

See [24] for a characterization of $\mathbf{S}_{1a(N)}$. In particular, that paper uses a strategy similar to RTT's Theorem 5B.1 to prove that for each $N \geq 0$, $\mathbf{C}^N$ is sound and complete for $\mathbf{S}_{1a(N)}$ for explicit definitions. Thus for (c) it suffices to show that $\not\models_{1a(N)}^{\mathcal{D}_3} AX \rightarrow G\overline{N}$.

For our counterexample, take the standard model $\mathbb{N}_<$ and the hypothesis $\varnothing$. We calculate

$$\begin{aligned} \delta(\varnothing) &= \{0\} \\ &\vdots \\ \delta^N(\varnothing) &= [0, N-1]. \end{aligned}$$

Hence, $\mathbb{N}_< + \delta^N(\varnothing) \not\models AX \rightarrow G\overline{N}$.

(iii): Appealing to the completeness of $\mathbf{C}^N$ for $\mathbf{S}_{1a(N)}$, it suffices to show that, given $\mathcal{D}$, there is N such that $\mathbf{S}_{1a(N)}$ preserves finite natural implications relative to $\mathcal{D}$. We'll also use a fact shown by Gupta [11]: we may reverse the $\forall M$ and $\exists n$ quantifiers in the definition of a finite definition. Thus if $\mathcal{D}$ is finite, there exists an n such that for all models M and hypotheses h, $\delta^n_{\mathcal{D},M}(h)$ is finitely reflexive. Given a finite definition $\mathcal{D}$, let us choose such an N as our strength.

Now suppose $\Gamma \models_f^{\mathcal{D}} B$, and let M be arbitrary. Since $\delta^N_{\mathcal{D},M}(h)$ is finitely reflexive for all h,

$$\forall h \left(M + \delta^N_{\mathcal{D},M}(h) \models \Gamma \text{ implies } M + \delta^N_{\mathcal{D},M}(h) \models B \right);$$

thus $\Gamma \models_{1a(N)}^{\mathcal{D}} B$. □

Acknowledgments. For helpful conversations, I would like to thank Douglas Blue, Patrick Chandler, Alexandre da Eira, James Hodgson, and James Shaw, who provided detailed comments on an earlier draft. Thanks also to an anonymous reviewer for several helpful suggestions. Finally, I owe a special debt of gratitude to Anil Gupta, not only for his careful reading and incisive questions, but for his generous mentorship and his vigorous encouragement to write this paper.

BIBLIOGRAPHY

[1] Barrio, E., & Picollo, L. (2013). Notes on ω-inconsistent theories of truth in second-order languages. *Review of Symbolic Logic* **6** (4), 733–741. https://doi.org/10.1017/s1755020313000269.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

[2] Bruni, R., & Sillari, G. (2018). A rational way of playing: Revision theory for strategic interaction. *Journal of Philosophical Logic*, **47**(3), 419–448. https://doi.org/10.1007/s10992-017-9433-2.

[3] Cantini, A. (1990). A theory of formal truth arithmetically equivalent to ID_1. *Journal of Symbolic Logic*, **55**(1), 244–259. https://doi.org/10.2307/2274965.

[4] Chapuis, A. (2000). Rationality and circularity. In Chapuis, A., & Gupta, A., editors. *Circularity, Definition, and Truth*. New Delhi: Munshiram Manoharlal Publishers, pp. 123–154.

[5] ———. (2003). An application of circular definitions: Rational decision. In Löwe, B., Räsch, T., & Malzkorn, W., editors. *Foundations of the Formal Sciences II*. Kluwer Academic, pp. 47–54.

[6] Feferman, S. (1991). Reflecting on incompleteness. *Journal of Symbolic Logic*, **56**(1), 1–49. https://doi.org/10.2307/2274902.

[7] Field, H. (2008). *Saving Truth from Paradox*. Oxford: Oxford University Press.

[8] Friedman, H., & Sheard, M. (1987). An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic*, **33**(1), 1–21. https://doi.org/10.1016/0168-0072(87)90073-x.

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

[10] ———. (2000). On circular concepts. In Chapuis, A., & Gupta, A., editors. *Circularity, Definition, and Truth*. New Delhi: Munshiram Manoharlal Publishers, pp. 49–78.

[11] ———. (2006). Finite circular definitions. In Bolander, T., Hendricks, V. F., & Pedersen, S. A., editors. *Self-Reference* CSLI Lecture Notes, no. 178, Stanford: CSLI Publications, pp. 79–93.

[12] ———. (2018). In praise of a logic of definitions that tolerates ω-inconsistency. *Philosophical Issues*, **28**(1), 176–195. https://doi.org/10.1111/phis.12121.

[13] ———. (2026). *The Concept of Truth: The Israel Lectures*. New York: Oxford University Press.

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

[15] Gupta, A., & Standefer, S. (2018). Intersubstitutivity principles and the generalization function of truth. *Synthese*, **195**(3), 1065–1075. https://doi.org/10.1007/s11229-017-1318-y.

[16] Halbach, V. (2014). *Axiomatic Theories of Truth* (second edition). Cambridge: Cambridge University Press.

[17] Herzberger, H. G. (1982). Naive semantics and the liar paradox. *Journal of Philosophy*, **79**(9), 479–497. https://doi.org/10.2307/2026380.

[18] Horwich, P. (1998). *Truth* (second edition). Oxford: Oxford University Press.

[19] Kremer, P. (1993). The Gupta-Belnap systems $\mathbf{S}^{\#}$ and $\mathbf{S}^{*}$ are not axiomatisable. *Notre Dame Journal of Formal Logic*, **34**(4), 583–596. https://doi.org/10.1305/ndjfl/1093633907.

[20] McGee, V. (1985). How truthlike can a predicate be? A negative result. *Journal of Philosophical Logic*, **14**(4), 399–410. https://doi.org/10.1007/bf00649483.

[21] Orilia, F. (2000). Property theory and the revision theory of definitions. *Journal of Symbolic Logic*, **65**(1), 212–246. https://doi.org/10.2307/2586533.

[22] Quine, W. V. (1986). *Philosophy of Logic* (second edition). Cambridge: Harvard University Press.

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press

[23] Schindler, J. (2026). Can the revision theory tolerate ω-inconsistency?
[24] ———. (2026). The varieties of revision.
[25] Varzi, A. C. (2007). Supervaluationism and its logics. *Mind*, **116**(463), 633–676. https://doi.org/10.1093/mind/fzm633.
[26] Williamson, T. (1994). *Vagueness*. London: Routledge.
[27] ———. (2018). Alternative logics and applied mathematics. *Philosophical Issues*, **28**(1), 399–424. https://doi.org/10.1111/phis.12131.
[28] ———. (2018). Supervaluationism and good reasoning. *Theoria: Revista de Teoría, Historia y Fundamentos de la Ciencia*, **33**(3), 521–537. https://doi.org/10.1387/theoria.18490.
[29] Yaqub, A. M. (1993). *The Liar Speaks the Truth: A Defense of the Revision Theory of Truth*. New York: Oxford University Press.

DEPARTMENT OF PHILOSOPHY
UNIVERSITY OF PITTSBURGH
PITTSBURGH, USA
E-mail: jschindler@pitt.edu

https://doi.org/10.1017/S1755020326101142 Published online by Cambridge University Press