Notre Dame Journal of Formal Logic
Volume 39, Number 3, Summer 1998

A Kripkean Approach to Unknowability and Truth

LEON HORSTEN

Abstract We consider a language containing partial predicates for subjective knowability and truth. For this language, inductive hierarchy rules are proposed which build up the extension and anti-extension of these partial predicates in stages. The logical interaction between the extension of the truth predicate and the anti-extension of the knowability predicate is investigated.

1 Introduction Kripke [10] develops a context-insensitive approach to the semantic paradoxes, in which the truth predicate is treated as a *partial* predicate. Toward the end of his paper he briefly considers extending his approach to notions other than truth, such as necessity ([10], pp. 78–79). He considers treating necessity as a predicate and giving its semantics in terms of the necessity *operator* and the truth predicate. Furthermore, he says:

> We can even "kick away the ladder," and take Nec(x) [i.e., the necessity *predicate*] as a primitive, treating it in a possible-world scheme as *if* it were defined by an operator plus a truth predicate. Like remarks apply to the propositional attitudes, if we are willing to treat them, using possible worlds, like modal operators. (I myself [sic] think that such a treatment involves considerable philosophical difficulties.) It is possible that the present approach can be applied to the supposed predicates of sentences in question without using either intensional operators or possible worlds, but at present I have no idea how to do so. ([10], p. 75)

Until today, there have been no published attempts to work out the suggestion that Kripke makes in the (first part of the) last sentence of this quotation. In the present paper it will be attempted to do this for *knowledge predicates* rather than for necessity predicates.[1] Toward the end of the paper some remarks are made on extending the approach of this paper to languages containing necessity predicates.

In the present paper, formal languages are considered which have a predicate (**B**) expressing the intuitive notion of *subjective knowability*, or *knowability in principle* (by a fixed knowing agent).[2] Theories of subjective knowability are generated

Received May 28, 1997; revised March 5, 1999

by building up the extension and the anti-extension of **B** in stages that are indexed by ordinals. Theories of knowability in principle will be collections of sentences that are made true by models in which the partial interpretation of **B** is a fixed point of this inductive procedure: this is what makes the approach thoroughly Kripkean. In the first of our constructions, a truth predicate (**T**) is used as an auxiliary tool. It provides a convenient way to show that an inductive rule for building up the extension of **B** has a consistent fixed point. Later it is shown that the truth predicate can also be given a more substantial role, namely, when an investigation is made of the *logical interaction* of the concept of knowability and the concept of truth in a Kripkean context.

The structure of this paper is as follows. In Section 2, two simple Kripkean hierarchy rules for knowability and truth are described, and some properties of these rules are established. It is shown that it is hard to see how the *extension* of the knowability predicate can be significantly enriched as the inductive rule goes through successor and limit stages. Nevertheless, the stronger hierarchy rules that are proposed in Section 3 describe ways of nontrivially enriching the *anti-extension* of the knowability predicate, which in turn leads to a nontrivial enrichment of the truth predicate. This makes the theory that is proposed here more about *un* knowability than about knowability (as one referee aptly put it). The final section compares the theory that is advanced here with some of the context-sensitive approaches to the epistemic paradoxes.

2 Simple Kripkean hierarchies of knowability and truth Let us look at some knowability hierarchies in the context of a three-valued valuation scheme. We will use, besides a knowability predicate, a Kripkean truth predicate as an auxiliary notion. So let $\mathcal{L}_{PA^{\mathbf{TB}}}$ be the language of first-order arithmetic extended with **T** ('truth') and **B** ('knowability') as our only partial predicates. In other words, some sentences are taken to be *definitely* knowable and some sentences are *definitely unknowable. About other sentences it is* indeterminate *whether they are knowable or not.*

In order to simplify notation, we will in the sequel write $\mathbf{T}A$ *(*$\mathbf{B}A$*) instead of* $\mathbf{T}\left(\underline{g\,(A)}\right)$ *(*$\mathbf{B}\left(\underline{g\,(A)}\right)$*), where* $\underline{g\,(A)}$ *is an expression denoting the Gödel number of A. Strictly speaking our notation is of course not even well formed, but it will do for our purposes. Let* $CL_{PA^{\mathbf{TB}}}(S)$ *be the closure of a set S under Peano Arithmetic (where* **T** *and* **B** *are allowed in instances of the induction scheme) and first-order logic. Let* $I \models_t \Delta$ *abbreviate that the interpretation I* Tarski-satisfies *the set* Δ *of sentences. And let* $I \models_{sv} \Delta$ *abbreviate that the interpretation I* supervaluation-satisfies *the set* Δ *of sentences.*

2.1 The naïve hierarchy rule $\mathcal{R}_0$

Definition 2.1 $\langle I_n(\mathbf{B})_\alpha, I_n(\mathbf{T})_\alpha\rangle$ denotes the partial structure, with the standard model $\mathbb{N}$ as the underlying arithmetical structure, defined by stage α of the inductive rule $\mathcal{R}_n$.

Definition 2.2 $I_n(\mathbf{B})^+_\alpha$ denotes the extension of **B** at stage α of the rule $\mathcal{R}_n$; $I_n(\mathbf{B})^-_\alpha$ denotes the anti-extension of **B** at stage α of the rule $\mathcal{R}_n$; $I_n(\mathbf{T})^+_\alpha$ denotes the extension of **T** at stage α of the rule $\mathcal{R}_n$; $I_n(\mathbf{T})^-_\alpha$ denotes the anti-extension of **T** at stage α of the rule $\mathcal{R}_n$.

Then the following is a naïve way of defining a hierarchy rule $\mathcal{R}_0$ for truth and knowability:

1. $I_0(\mathbf{B})^+_0 = I_0(\mathbf{B})^-_0 = I_0(\mathbf{T})^+_0 = I_0(\mathbf{T})^-_0 = \varnothing$;
2. $I_0(\mathbf{B})^+_{\alpha+1} = CL_{PA^{\mathrm{TB}}}\left[I_0(\mathbf{B})^+_\alpha \cup \left\{\mathbf{T}C \mid C \in I_0(\mathbf{B})^+_\alpha\right\} \cup \left\{\neg\mathbf{T}C \mid \neg C \in I_0(\mathbf{B})^+_\alpha\right\}\right]$;
3. $I_0(\mathbf{T})^+_{\alpha+1} = \{A \mid \langle I_0(\mathbf{B})_\alpha, I_0(\mathbf{T})_\alpha\rangle \models_{sv} A\}$;
4. $I_0(\mathbf{T})^-_{\alpha+1} = \{A \mid \langle I_0(\mathbf{B})_\alpha, I_0(\mathbf{T})_\alpha\rangle \models_{sv} \neg A\}$;
5. $I_0(\mathbf{B})^-_{\alpha+1} = I_0(\mathbf{T})^-_{\alpha+1}$;
6. $I_0(\mathbf{T})^+_\gamma = \bigcup I_0(\mathbf{T})^+_{<\gamma}$ for γ a limit ordinal;
7. $I_0(\mathbf{T})^-_\gamma = \bigcup I_0(\mathbf{T})^-_{<\gamma}$ for γ a limit ordinal;
8. $I_0(\mathbf{B})^+_\gamma = \left(\bigcup I_0(\mathbf{B})^+_{<\gamma}\right) \cup \left\{\forall x A(x) \mid \text{for all } n,\ A(\underline{n}) \in I_0(\mathbf{B})^+_\beta \text{ for some } \beta < \gamma\right\}$ for γ a limit ordinal;
9. $I_0(\mathbf{B})^-_\gamma = \bigcup I_0(\mathbf{T})^-_{<\gamma}$ for γ a limit ordinal.

The clauses of $\mathcal{R}_0$ deserve some comment. We come to know facts about the natural numbers (and about the knowability and truth-value of such facts) by proving them. In the present approach, we take these proofs to be carried out in a classical, two-valued language. Then it is natural to work with a supervaluation scheme, as is done in $\mathcal{R}_0$ (clauses 3 and 4). It is also possible to work with a Kleene scheme. But then the clauses for defining the extension of **B** would have to be different: we would have to take the knower to prove things in *partial* Peano Arithmetic.[3]

The successor clause for $I_1(\mathbf{B})^+$ says that what the idealized knower knows at the successor stage is the closure of:

1. all sentences which she already knew at the previous stage;
2. *that* what she knew at the previous stage is true;
3. that the negations of these sentences are not true.

This seems to embody reasonable reflection properties for knowability. At limit stages, the knower reflects on the fact that she has proved $A(\underline{n})$ for all natural numbers n and correctly concludes from this that $\forall x A(x)$. In the hierarchy determined by $\mathcal{R}_0$, the knower "learns" about arithmetic and the notion of truth. We could strengthen $\mathcal{R}_0$ a bit so that the knower also learns about *knowledge*. For this purpose, we would add to the extension of **B** at successor stages $\alpha+1$ also the sets $\left\{\mathbf{B}A \mid A \in I_1(\mathbf{B})^+_\alpha\right\}$ and $\left\{\neg\mathbf{B}A \mid \neg A \in I_1(\mathbf{B})^+_\alpha\right\}$. Analogues of the propositions that we will prove for $\mathcal{R}_0$ also hold for the resulting rule.

2.2 The inductive rule $\mathcal{R}_1$ The inductive rule $\mathcal{R}_0$ is in certain respects too strong. The limit-clause for the extension of **B** implies that the extension of **B** at the least fixed point is closed under the infinitary ω-rule. But systems that are closed under the ω-rule prove all first-order arithmetical truths (see e.g., Hinman [7], pp. 121–22). So the least fixed point model of $\mathcal{R}_0$ takes all arithmetical truths to be subjectively knowable. Even though great minds like Gödel and Hilbert have expressed sympathy

for this consequence, many philosophers and mathematicians have taken issue with it. So philosophical caution seems to demand that at limit stages we let the extension of **B** be simply the union of the extensions of **B** at earlier stages.[4] But if we weaken $\mathcal{R}_0$ in this way, then the extension of **B** at the least fixed point becomes uninteresting. It is easy to see that it becomes the theory that is obtained by closing Peano Arithmetic under the inference rules:

From A, infer $\mathbf{T}A$.

From $\neg A$, infer $\neg\mathbf{T}A$.

Considerations such as these point to the fact that it is very difficult to see how inductive hierarchy rules can give an interesting account of how the extension of **B** "grows" in stages. Because of this, we will let the extension of **B** be *constant* throughout the stages of the hierarchy rules that will be formulated in the sequel.

The extension of **B** should then be identified with a sufficiently strong axiomatic theory of truth and knowability. Reinhardt [12] argues in the context of a Kleene valuation scheme that the collection of all sentences A such that $\mathbf{T}(A)$ is provable in the Kripke-Feferman system constitutes a rich theory of truth.[5] But we are working in the context of a supervaluation scheme. Therefore it is more suitable to concentrate on Cantini's axiomatic theory VF.[6] Let VFT be the theory that consists of all sentences A such that $\mathbf{T}(A)$ is provable in VF. Then we will set the extension of **B** equal to VFT at all stages of the inductive rules that will henceforth be considered.

We have seen that the knower also ought to be capable of proving nontrivial propositions about knowability (and its relation to truth). This should lead one to consider extensions of VF by plausible axioms governing the notion of knowability.[7] The resulting proof systems would be interesting in their own right.[8] Nevertheless, such a proof-theoretic investigation would detract from the model-theoretic investigation that we are currently engaged in. Therefore, this line will not be pursued further here.

Since we have decided to let the extension of **B** coincide with VFT already at stage 0, we have to ensure that VFT is also included in the extension of **T** at stage 0 of our inductive rules.[9] This can be done in the following way. Consider the classical supervaluation rule $\mathcal{S}$. Identify the extension (anti-extension) of **T** at stage 0 with the extension (anti-extension) of **T** at the least fixed point F of $\mathcal{S}$ (where **B** is treated as a partial predicate of which the extension and anti-extension remain empty at all stages). It is easy to see that the least fixed point F of $\mathcal{S}$ makes all axioms of VF true ([5], p. 250).

Here then is a simple but intuitively sound hierarchy rule $\mathcal{R}_1$:

1. $I_1(\mathbf{B})^+_0 = VFT$,
2. $I_1(\mathbf{B})^-_0 = \varnothing$,
3. $I_1(\mathbf{T})^+_0 = \{A \mid F \models_{sv} A\}$,
4. $I_1(\mathbf{T})^-_0 = \{A \mid F \models_{sv} \neg A\}$,
5. $I_1(\mathbf{T})^+_{\alpha+1} = \{A \mid \langle I_1(\mathbf{B})_\alpha, I_1(\mathbf{T})_\alpha\rangle \models_{sv} A\}$,
6. $I_1(\mathbf{T})^-_{\alpha+1} = \{A \mid \langle I_1(\mathbf{B})_\alpha, I_1(\mathbf{T})_\alpha\rangle \models_{sv} \neg A\}$,
7. $I_1(\mathbf{B})^+_{\alpha+1} = I_1(\mathbf{B})^+_\alpha$,

8. $I_1(\mathbf{B})^-_{\alpha+1} = I_1(\mathbf{T})^-_{\alpha+1} \bigcup I_1(\mathbf{B})^-_0$,
9. $I_1(\mathbf{T})^+_\gamma = \bigcup I_1(\mathbf{T})^+_{<\gamma}$ for γ a limit ordinal,
10. $I_1(\mathbf{B})^+_\gamma = \bigcup I_1(\mathbf{B})^+_{<\gamma}$ for γ a limit ordinal,
11. $I_1(\mathbf{T})^-_\gamma = \bigcup I_1(\mathbf{T})^-_{<\gamma}$ for γ a limit ordinal,
12. $I_1(\mathbf{B})^-_\gamma = \bigcup I_1(\mathbf{B})^-_{<\gamma}$ for γ a limit ordinal.

Note that at successor stages, $\mathcal{R}_1$ includes $I_1(\mathbf{B})^-_0$ in the anti-extension of **B**. This inclusion is doing no work in $\mathcal{R}_1$, since $I_1(\mathbf{B})^-_0 = \varnothing$. But in Section 3, inductive rules will be defined in which $\mathcal{R}_1$ is used as an *auxiliary* rule. In these situations, $I_1(\mathbf{B})^-_0$ will not usually be set equal to $\varnothing$, and $I_1(\mathbf{B})^-_0$ will have to be included in $I_1(\mathbf{B})^-_{\alpha+1}$ in order to ensure monotonicity.

We go on to prove some simple propositions about $\mathcal{R}_1$ which will be shared by the stronger hierarchy rules that will be proposed in subsequent sections.

2.3 Simple properties of $\mathcal{R}_1$

Proposition 2.3 *$\mathcal{R}_1$ is monotone in* **B** *and* **T**.

Proof: It suffices to note that $\models_{sv}$ is monotone. □

Proposition 2.4 *For all A and α: if $A \in I_1(\mathbf{B})^+_\alpha$, then $A \in I_1(\mathbf{T})^+_\alpha$.*

Proof: By transfinite induction. Since $I_1(\mathbf{B})^+$ is constant, the only nontrivial case is when $\alpha = 0$. But this follows from the fact that F makes all theorems of VF true. □

Proposition 2.5 *For all α, $I_1(\mathbf{T})^+_\alpha \cap I_1(\mathbf{T})^-_\alpha = \varnothing$.*

Proof: By a *reductio*. If there is at least β for which the property fails, then it must be a successor ordinal. Let $\beta = \alpha + 1$ be the least such ordinal. This means that we have an $A \in I_1(\mathbf{T})^+_{\alpha+1} \cap I_1(\mathbf{T})^-_{\alpha+1}$, and it is given that $I_1(\mathbf{T})^+_\alpha \cap I_1(\mathbf{T})^-_\alpha = \varnothing$. Since $I_1(\mathbf{B})^+_\alpha \subseteq I_1(\mathbf{T})^+_\alpha$, and $I_1(\mathbf{B})^-_\alpha = I_1(\mathbf{T})^-_\alpha$, we also have $I_1(\mathbf{B})^+_\alpha \cap I_1(\mathbf{B})^-_\alpha = \varnothing$. But then for all A, if $\langle I_1(\mathbf{B})_\alpha, I_1(\mathbf{T})_\alpha\rangle \models_{sv} A$, it cannot be the case that $\langle I_1(\mathbf{B})_\alpha, I_1(\mathbf{T})_\alpha\rangle \models_{sv} \neg A$. Contradiction. □

Given the previous propositions it follows that $\mathcal{R}_1$ reaches a consistent least fixed point in both **T** and **B**. We now show that the extension of **B** at the least fixed point of $\mathcal{R}_1$ is *sound*, in the following sense.

Proposition 2.6 *Consider the least fixed point $\langle I_1(\mathbf{B})_{\mathcal{R}_1}, I_1(\mathbf{T})_{\mathcal{R}_1}\rangle$ of $\mathcal{R}_1$. For all $A \in I_1(\mathbf{B})^+_{\mathcal{R}_1}$, $\langle I_1(\mathbf{B})^+_{\mathcal{R}_1}, I_1(\mathbf{T})^+_{\mathcal{R}_1}\rangle \models_t A$.*

Proof: Suppose $A \in I_1(\mathbf{B})^+_{\mathcal{R}_1}$. Then $A \in I_1(\mathbf{T})^+_{\mathcal{R}_1}$, whereby

$$\langle I_1(\mathbf{B})_{\mathcal{R}_1}, I_1(\mathbf{T})_{\mathcal{R}_1}\rangle \models_{sv} A.$$

So by closing off (which we can by the previous proposition), we see that

$$\langle I_1(\mathbf{B})^+_{\mathcal{R}_1}, I_1(\mathbf{T})^+_{\mathcal{R}_1}\rangle \models_t A$$

□

2.4 Some easy fixed point calculations First we consider the *absolute Gödel sentence*, which says of itself that it is unknowable in principle.

Proposition 2.7 *Let G be such that $\vdash_{PA^{\mathrm{TB}}} G \longleftrightarrow \neg\mathbf{B}(G)$. Then $G, \neg G \notin I_1(\mathbf{B})^+_{\mathcal{R}_1}$.*

Proof:

Case 1: $\neg G \in I_1(\mathbf{B})^+_{\mathcal{R}_1} \Longrightarrow \mathbf{B}(G) \in I_1(\mathbf{B})^+_{\mathcal{R}_1} \Longrightarrow \mathbf{B}(G) \in I_1(\mathbf{T})^+_{\mathcal{R}_1} \Longrightarrow G \in I_1(\mathbf{B})^+_{\mathcal{R}_1}$. Contradiction.

Case 2: $G \in I_1(\mathbf{B})^+_{\mathcal{R}_1} \Longrightarrow \neg\mathbf{B}(G) \in I_1(\mathbf{B})^+_{\mathcal{R}_1} \Longrightarrow \neg\mathbf{B}(G) \in I_1(\mathbf{T})^+_{\mathcal{R}_1} \Longrightarrow \mathbf{B}(G) \in I_1(\mathbf{T})^-_{\mathcal{R}_1} \Longrightarrow G \in I_1(\mathbf{B})^-_{\mathcal{R}_1}$. Contradiction. □

In other words, according to $\mathcal{R}_1$, the absolute Gödel sentence is absolutely undecidable by the knowing agent. Note that the reasoning of the above proposition holds for *any* consistent fixed point of $\mathcal{R}_1$.

Proposition 2.8 *Let G be as in the previous proposition. Then $G, \neg G \notin I_1(\mathbf{T})^+_{\mathcal{R}_1}$.*

Proof:

Case 1: $\neg G \in I_1(\mathbf{T})^+_{\mathcal{R}_1} \Longrightarrow \mathbf{B}(G) \in I_1(\mathbf{T})^+_{\mathcal{R}_1} \Longrightarrow G \in I_1(\mathbf{B})^+_{\mathcal{R}_1} \Longrightarrow G \in I_1(\mathbf{T})^+_{\mathcal{R}_1}$. Contradiction.

Case 2: $G \in I_1(\mathbf{T})^+_{\mathcal{R}_1} \Longrightarrow \neg\mathbf{B}(G) \in I_1(\mathbf{T})^+_{\mathcal{R}_1} \Longrightarrow \mathbf{B}(G) \in I_1(\mathbf{T})^-_{\mathcal{R}_1} \Longrightarrow G \in I_1(\mathbf{B})^-_{\mathcal{R}_1} \Longrightarrow G \in I_1(\mathbf{T})^-_{\mathcal{R}_1}$. Contradiction. □

Only the last step of this last proof will not go through for the stronger inductive rules that will be considered in the next section. And intuitively it is not clear that it *should* go through: there may be sentences which are definitely unprovable but which are nevertheless definitely true. Actually, there is a familiar Gödelian argument to show that G should really be in the extension of the *truth* predicate. For suppose that the absolute Gödel sentence is knowable in principle. If it is knowable, then it is true. But it says of itself that it is unknowable! So we reach a contradiction. Therefore G must be unknowable. But since this is exactly what it says of itself, it must be true.[10] On the other hand, this argument certainly has the flavor of *strengthened liar*-type reasoning. And at least in the case of truth, we know that strengthened liar arguments lead to serious trouble. In the Gödelian argument under consideration, the suspicious step is the move from the inconsistency of the assumption that G is knowable to the conclusion that G is (definitely!?) unknowable. One might seriously wonder whether such a principle will not, in the present setting, necessarily lead to contradictions. This question will be taken up in Section 3.

Next we consider the *knower sentence*,[11] which says of its own negation that it is subjectively knowable.

Proposition 2.9 *Let K be such that $\vdash_{PA^{\mathrm{TB}}} K \longleftrightarrow \mathbf{B}(\neg K)$. Then $K, \neg K \notin I_1(\mathbf{B})^+_{\mathcal{R}_1}$.*

Proof:

Case 1: $\neg K \in I_1(\mathbf{B})^+_{\mathcal{R}_1} \Longrightarrow \neg\mathbf{B}(\neg K) \in I_1(\mathbf{B})^+_{\mathcal{R}_1} \Longrightarrow \neg\mathbf{B}(\neg K) \in I_1(\mathbf{T})^+_{\mathcal{R}_1} \Longrightarrow \mathbf{B}(\neg K) \in I_1(\mathbf{T})^-_{\mathcal{R}_1} \Longrightarrow \neg K \in I_1(\mathbf{B})^-_{\mathcal{R}_1}$. Contradiction.

Case 2: $K \in I_1(\mathbf{B})^+_{\mathcal{R}_1} \Longrightarrow \mathbf{B}(\neg K) \in I_1(\mathbf{B})^+_{\mathcal{R}_1} \Longrightarrow \mathbf{B}(\neg K) \in I_1(\mathbf{T})^+_{\mathcal{R}_1} \Longrightarrow \neg K \in I_1(\mathbf{B})^+_{\mathcal{R}_1}$. Contradiction. □

So according to $\mathcal{R}_1$ the knower sentence is also absolutely undecidable by the knowing agent.

Proposition 2.10 *Let K be as in the previous proposition. Then* $K, \neg K \notin I_1(\mathbf{T})^+_{\mathcal{R}_1}$.

Proof:

Case 1: $K \in I_1(\mathbf{T})^+_{\mathcal{R}_1} \Longrightarrow \mathbf{B}(\neg K) \in I_1(\mathbf{T})^+_{\mathcal{R}_1} \Longrightarrow \neg K \in I_1(\mathbf{B})^+_{\mathcal{R}_1} \Longrightarrow \neg K \in I_1(\mathbf{T})^+_{\mathcal{R}_1}$. Contradiction.

Case 2: $\neg K \in I_1(\mathbf{T})^+_{\mathcal{R}_1} \Longrightarrow \neg\mathbf{B}(\neg K) \in I_1(\mathbf{T})^+_{\mathcal{R}_1} \Longrightarrow \mathbf{B}(\neg K) \in I_1(\mathbf{T})^-_{\mathcal{R}_1} \Longrightarrow \neg K \in I_1(\mathbf{B})^-_{\mathcal{R}_1} \Longrightarrow \neg K \in I_1(\mathbf{T})^-_{\mathcal{R}_1}$. Contradiction. □

Here again only the last step of the last proof will not go through for the stronger inductive rules that will be considered in the next section. And by a parallel argument to that for the truth of *G*, one might ask whether *K* should not really be in the anti-extension of the *truth* predicate. For suppose that the negation of *K* is knowable. Then the negation of *K* must be true. But since *K* says of itself that its negation is knowable, we obtain that its negation must be unknowable. But this contradicts the assumption. So the negation of *K* is unknowable. But since *K* says of itself that its negation *is* knowable, it must be false.[12] By arguments similar to those for *G* and *K* one can convince oneself that neither the liar sentence (L) nor its negation belong to the extension or the anti-extension of **B** or **T** (this is left to the reader). Also, the question can be raised whether L and $\neg L$ should not really be in the anti-extension of **B**. For the familiar *liar argument* shows that from the assumption that L is true, a contradiction can easily be derived: if L is true, then $\neg L$, whereby it is not the case that L is true. Since it is therefore inconsistent to assume that L is true, L must be determinately unknowable. Here the suspicious principle is the move that allows one to conclude from the inconsistency of assuming that a sentence is true to the conclusion that it is definitely unknowable. But note that this principle is *weaker* than the suspicious principle involved in the Gödelian argument for the truth of *G*.

In sum, the previous propositions show that the knower sentence and the absolute Gödel sentence behave as one would at first blush expect on a Kripkean picture, but one wonders whether it is possible to construct arguably sound inductive rules for which the least fixed point verifies *G*, falsifies *K*, and takes both L and $\neg L$ to be definitely unknowable.

3 More Kripkean hierarchies

3.1 The rule $\mathcal{R}_2$ We will now introduce a new inductive rule $\mathcal{R}_2$, which is just like $\mathcal{R}_1$, except for the successor clause for $I(\mathbf{B})^-_{\alpha+1}$. To describe this rule we first introduce some terminology.

Definition 3.1 If U, V are partial structures, then we say that U is a *substructure* of V (abbreviated: $U \subseteq V$) if and only if $U^{\mathbf{B}+} \subseteq V^{\mathbf{B}+}$, $U^{\mathbf{B}-} \subseteq V^{\mathbf{B}-}$, $U^{\mathbf{T}+} \subseteq V^{\mathbf{T}+}$, and $U^{\mathbf{T}-} \subseteq V^{\mathbf{T}-}$.

Definition 3.2 A partial structure $U = \langle U^{\mathbf{B}+}, U^{\mathbf{B}-}, U^{\mathbf{T}+}, U^{\mathbf{T}-} \rangle$ is said to be *inclusive* if and only if

$$U^{\mathbf{T}+} \subseteq \{A \mid \langle U^{\mathbf{B}+}, U^{\mathbf{B}-}, U^{\mathbf{T}+}, U^{\mathbf{T}-} \rangle \models_{sv} A\}$$

and

$$U^{\mathbf{T}-} \subseteq \{A \mid \langle U^{\mathbf{B}+}, U^{\mathbf{B}-}, U^{\mathbf{T}+}, U^{\mathbf{T}-} \rangle \models_{sv} \neg A\}$$

Definition 3.3 For every partial structure $U = \langle U^{\mathbf{B}+}, U^{\mathbf{B}-}, U^{\mathbf{T}+}, U^{\mathbf{T}-} \rangle$, we say that U is *normal* if and only if $U^{\mathbf{B}+} \cap U^{\mathbf{B}-} = U^{\mathbf{T}+} \cap U^{\mathbf{T}-} = \varnothing$, U is inclusive, and $U^{\mathbf{B}+} \subseteq U^{\mathbf{T}+}$.

The idea behind this definition is that in order to be a putative candidate for being a suitable partial interpretation for $\mathcal{L}_{PA^{\mathbf{TB}}}$, a structure must at least be inclusive and have the property that whatever it takes to be knowable is true.

For partial structures U extending the initial stage $\langle I_1(\mathbf{B})_0, I_1(\mathbf{T})_0 \rangle$ of the rule $\mathcal{R}_1$, we denote as $I_i(\mathbf{B}, U)^+_\alpha$ the extension of $\mathbf{B}$ at stage α of the rule which is just like $\mathcal{R}_i$ except that the initial partial structure is U. Similar conventions hold for $I_i(\mathbf{B}, U)^-_\alpha$, $I_i(\mathbf{T}, U)^+_\alpha$, $I_i(\mathbf{T}, U)^-_\alpha$.

We also introduce notation to refer to the structure that has been defined by the αth stage of a rule $\mathcal{R}_i$.

Definition 3.4 $\mathcal{S}^i_\alpha \equiv \langle I_i(\mathbf{B})^+_\alpha, I_i(\mathbf{B})^-_\alpha, I_i(\mathbf{T})^+_\alpha, I_i(\mathbf{T})^-_\alpha \rangle$

When it is clear from the context which rule we are referring to, we will sometimes omit the superscript from $\mathcal{S}^i_\alpha$.

Now we define $\mathcal{R}_2$ to be the inductive rule which is just like $\mathcal{R}_1$, except for the fact that the successor clause for $I(\mathbf{B})^-_{\alpha+1}$ now reads:

$$I_2(\mathbf{B})^-_{\alpha+1} = I_2(\mathbf{T})^-_{\alpha+1} \cup \left\{ \begin{array}{c} A \mid \text{for all normal structures } U \supseteq S^2_\alpha\text{: if } U \models_{sv} A, \\ \text{then there is a } \beta \text{ such that } I_1(\mathbf{T}, U)^+_\beta \text{ is inconsistent} \end{array} \right\}$$

In other words, at successor stages $\alpha + 1$ we add to the anti-extension of $\mathbf{B}$ those sentences φ which are such that *if* φ were assumed to be true and the hierarchy were continued in accordance with the rule $\mathcal{R}_1$ from stage α onward, *then* $I(\mathbf{T})^+_\beta$ would become inconsistent for some β. The underlying idea is that if assuming φ to be true would lead, according to some correct rule, to an inconsistency at some stage, then that sentence φ is definitely unknowable. Or, shorter still, if it is inconsistent for a given sentence to be true, then it is definitely unknowable.

As with $\mathcal{R}_1$, it will be shown that $\mathcal{R}_2$ has a consistent least fixed point. In addition it will be shown that the least fixed point of $\mathcal{R}_2$ makes certain sentences which $\mathcal{R}_1$ classifies as neither determinately knowable nor determinately unknowable, come out definitely unknowable. As we will see, one such sentence is the knower sentence of Section 2.4.

All this would be of little value if we did not have strong reason to believe that $\mathcal{R}_2$ intuitively is a *sound* rule, that is, that it classifies intuitively true sentences as true, intuitively false ones as false, intuitively knowable sentences as knowable, and intuitively unknowable sentences as unknowable. Here is a philosophical argument for the soundness of $\mathcal{R}_2$. We already know that the basis of $\mathcal{R}_2$ (stage 0) is sound.

Since $\mathcal{R}_2$'s clause for limit stages is unobjectionable, it only remains to be verified that successor stages preserve soundness. And here, evidently, we have to take a close look at the clause for the anti-extension of **B**: if it is inconsistent to assume that a sentence is made true, then this sentence is definitely unknowable. Now suppose that for a given sentence A it is inconsistent that it is made true. Then there are two possibilities. Either A is determinately false or A has no determinate truth-value. In the former case A is, as a matter of course, definitely unknowable. But even in the latter case, the only reasonable thing to say is that A is definitely unknowable, for *a sentence that has no determinate truth-value can never be known*. So in either case A is definitely unknowable. This appears to be a compelling argument for the soundness of $\mathcal{R}_2$.

3.2 Properties of $\mathcal{R}_2$ The extension of **B** remains constant at all stages of the rule $\mathcal{R}_2$. Therefore we easily obtain a generalization of Proposition 2.4.

Proposition 3.5 *For all normal U extending S^1_0, and for all β: $I_1(\mathbf{B}, U)^+_\beta \subseteq I_1(\mathbf{T}, U)^+_\beta$, and $I_2(\mathbf{B}, U)^+_\beta \subseteq I_2(\mathbf{T}, U)^+_\beta$.*

Proof: The proof is the same as for Proposition 2.4. □

Note that we retrieve Proposition 2.4 from this more general proposition by taking $U = \langle I_1(\mathbf{B})_0, I_1(\mathbf{T})_0 \rangle$.

Proposition 3.6 *$\mathcal{R}_2$ is monotone in* **T** *and* **B**.

Proof: By transfinite induction: first we note that $\mathcal{R}_1$ always preserves monotonicity when it is used as an auxiliary rule. For this we are using the fact that for all α, $I_1(\mathbf{B})^-_0 \subseteq I_1(\mathbf{B})^-_{\alpha+1}$. It then suffices to check that $I_2(\mathbf{B})^-_\alpha \subseteq I_2(\mathbf{B})^-_{\alpha+1}$ for all α.

Case 1: $\alpha = 0$. Obvious.

Case 2: $\alpha = \beta + 1$. Let $A \in I_2(\mathbf{B})^-_\alpha$. Then either $A \in I_2(\mathbf{T})^-_\alpha$ or for all normal $U \supseteq S^2_\beta$: if $U \models_{sv} A$, then there is a γ such that $I_1(\mathbf{T}, U)^+_\gamma$ is inconsistent. In the former disjunct, we have $A \in I_2(\mathbf{B})^-_{\alpha+1}$, since $I_2(\mathbf{T})^-_\alpha \subseteq I_2(\mathbf{T})^-_{\alpha+1} \subseteq I_2(\mathbf{B})^-_{\alpha+1}$. But the latter disjunct is also acceptable. For take any normal $U^* \supseteq S_{\beta+1}$ such that $U^* \models_{sv} A$. Since $S_{\beta+1} \supseteq S_\beta$, we have $U^* \supseteq S_\beta$. So, by the inductive hypothesis, $I_1(\mathbf{T}, U^*)^+_\gamma$ is inconsistent for some γ. So $A \in I_2(\mathbf{B})^-_{\alpha+1}$.

Case 3: α is a limit ordinal γ. Then there is a smallest $\theta < \gamma$ such that $A \in I_2(\mathbf{B})^-_\beta$ for all $\beta \geq \theta$. Then either $A \in I_2(\mathbf{T})^-_\theta$, whereby $A \in I_2(\mathbf{T})^-_\gamma \subseteq I_2(\mathbf{T})^-_{\gamma+1} \subseteq I_2(\mathbf{B})^-_{\gamma+1}$, or for all normal $U \supseteq S_\theta$: if $U \models_{sv} A$, then $I_1(\mathbf{T}, U)^+_\gamma$ is inconsistent for some γ. Now take any normal $U^* \supseteq S_\gamma$ such that $U^* \models_{sv} A$. Since $S_\gamma \supseteq S_\theta$, we have $U^* \supseteq S_\theta$. So by the inductive hypothesis, $I_1(\mathbf{T}, U^*)^+_\gamma$ must be inconsistent for some γ. So for either disjunct we have $A \in I_2(\mathbf{B})^-_{\gamma+1}$. □

Proposition 3.7 *For all α:*

(a) $I_2(\mathbf{T})^+_\alpha \cap I_2(\mathbf{T})^-_\alpha = I_2(\mathbf{B})^+_\alpha \cap I_2(\mathbf{B})^-_\alpha = \varnothing$.

(b) For all β: $I_1\left(\mathbf{T}, S^2_\alpha\right)^+_\beta$ is consistent.

Proof: By a double induction. Suppose that there is a least α for which the property fails. Then it must be a successor ordinal. Let $\beta = \alpha + 1$ be the least such ordinal.

Case 1: Suppose that part (a) of this property fails. Then either there is an $A \in I_2(\mathbf{T})^+_{\alpha+1} \cap I_2(\mathbf{T})^-_{\alpha+1}$ or an $A \in I_2(\mathbf{B})^+_{\alpha+1} \cap I_2(\mathbf{B})^-_{\alpha+1}$. The first of these possibilities can be dismissed by the inductive hypothesis. So take any $A \in I_2(\mathbf{B})^+_{\alpha+1}$, and suppose, for a *reductio*, that $A \in I_2(\mathbf{B})^-_{\alpha+1}$. There are two possibilities.

Subcase 1: $A \in I_2(\mathbf{T})^-_{\alpha+1}$. But since $A \in I_2(\mathbf{B})^+_{\alpha+1}$, we have $A \in I_2(\mathbf{T})^+_{\alpha+1}$ (by Proposition 3.5). So we are contradicting the fact that $I_2(\mathbf{T})^+_{\alpha+1} \cap I_2(\mathbf{T})^-_{\alpha+1} = \varnothing$.

Subcase 2: We have for all normal $U \supseteq S_\alpha$: if $U \models_{sv} A$, then $I_1(\mathbf{T}, U)^+_\beta$ is inconsistent for some β. But since $A \in I_2(\mathbf{B})^+_{\alpha+1}$, we have $A \in I_2(\mathbf{T})^+_{\alpha+1}$. That means that $\langle I_2(\mathbf{B})_\alpha, I_2(\mathbf{T})_\alpha\rangle \models_{sv} A$, that is, $S_\alpha \models_{sv} A$. So there must be a β such that $I_1(\mathbf{T}, S_\alpha)^+_\beta$ is inconsistent, contradicting part (b) of the inductive hypothesis.

Case 2: Suppose that part (b) of this property fails. We show that

$$I_1\left(\mathbf{T}, S_{\alpha+1}\right)^+_\gamma \cap I_1\left(\mathbf{T}, S_{\alpha+1}\right)^-_\gamma = I_1\left(\mathbf{B}, S_{\alpha+1}\right)^+_\gamma \cap I_1\left(\mathbf{B}, S_{\alpha+1}\right)^-_\gamma = \varnothing$$

for all γ. We proceed by an induction on γ (so this is the induction inside the main induction). It suffices to look at successor ordinals, so suppose there is a least ordinal $\gamma = \delta + 1$ for which this property fails. By the inductive hypothesis we cannot have an $A \in I_1\left(\mathbf{T}, S_{\alpha+1}\right)^+_\gamma \cap I_1\left(\mathbf{T}, S_{\alpha+1}\right)^-_\gamma$, so

Subcase 1: Suppose there is an

$$A \in I_1\left(\mathbf{B}, S_{\alpha+1}\right)^+_{\delta+1} \cap I_1\left(\mathbf{B}, S_{\alpha+1}\right)^-_{\delta+1}.$$

If $A \in I_1\left(\mathbf{B}, S_{\alpha+1}\right)^+_{\delta+1}$, then $A \in I_1\left(\mathbf{T}, S_{\alpha+1}\right)^+_{\delta+1}$. So if $A \in I_1\left(\mathbf{B}, S_{\alpha+1}\right)^-_{\delta+1}$, then if $A \in I_1\left(\mathbf{T}, S_{\alpha+1}\right)^-_{\delta+1}$, we would contradict the fact that $I_1\left(\mathbf{T}, S_{\alpha+1}\right)^+_\gamma \cap I_1\left(\mathbf{T}, S_{\alpha+1}\right)^-_\gamma = \varnothing$.

Subcase 2: It only remains to consider the possibility that $A \in I_1\left(\mathbf{B}, S_{\alpha+1}\right)^-_0$. But this possibility is easily dismissed, since $I_1\left(\mathbf{B}, S_{\alpha+1}\right)^+_0 \cap I_1\left(\mathbf{B}, S_{\alpha+1}\right)^-_0 = \varnothing$, and $I_1\left(\mathbf{B}, S_{\alpha+1}\right)^+_0 = I_1\left(\mathbf{B}, S_{\alpha+1}\right)^+_\beta$ for all β. □

It follows from these propositions that $\mathcal{R}_2$ has a consistent least fixed point which has a model in the natural numbers.

Note that in $\mathcal{R}_2$ there is real *logical interaction* between the knowability predicate and the truth predicate: the anti-extension of **B** at α depends on putative later extensions of **T**, and the anti-extension of **B** at α, of course, codetermines the extension of **T** at stage $\alpha + 1$. The extension of the truth predicate is thereby enriched by the anti-extension of the knowability predicate in a way that cannot be obtained in a similar way in a language which has only *one* partial predicate (for truth). All this is made possible by the conceptual relation between knowability and truth: knowability entails truth, so not being definitely true entails being definitely unprovable (but does *not* in general entail being definitely false!). Since this conceptual relation also holds between necessity and truth, something similar can be done for a language with partial predicates for truth and necessity.

Proposition 3.8 *$L, \neg L \in I_2(\mathbf{B})^-_{\mathcal{R}_2}$, where $I_2(\mathbf{B})^-_{\mathcal{R}_2}$ is the anti-extension of* **B** *at the least fixed point of $\mathcal{R}_2$.*

Proof: Take any (normal) $U \supseteq S_0$ such that $U \models_{sv} L$. Extend U according to $\mathcal{R}_1$ until you reach a least fixed point U_f. By monotonicity, $U_f \models_{sv} L$. But we also have $U_f \models_{sv} L \longleftrightarrow \neg\mathbf{T}(L)$. So we have $U_f \models_{sv} \neg\mathbf{T}(L)$. Since U_f is a fixed point, we have $U_f \models_{sv} \neg L$, which gives us a contradiction. So $L \in I_2(\mathbf{B})^-_1$, whence by monotonicity we have $L \in I_2(\mathbf{B})^-_{\mathcal{R}_2}$. A similar argument yields that $\neg L \in I_2(\mathbf{B})^-_{\mathcal{R}_2}$. □

Note that we did not make use of the normality of U in this proof.

Proposition 3.9 $\neg G \in I_2(\mathbf{B})^-_{\mathcal{R}_2}$, $K \in I_2(\mathbf{B})^-_{\mathcal{R}_2}$.

Proof: This is similar to the proof of the previous proposition. □

The absolute Gödel sentence does not belong to the anti-extension of **B** at the least fixed point of $\mathcal{R}_2$.

Proposition 3.10 $G \notin I_2(\mathbf{B})^-_{\mathcal{R}_2}$.

Proof: We must show that for each S_α, there is a normal $U \supseteq S_\alpha$ such that $U \models_{sv} G$ and $I_1(\mathbf{T}, U)^+_\beta$ is consistent for each β. Consider an arbitrary S_α. We form a partial structure U by adding G to the anti-extension of **B** of S_α. Then $U \models_{sv} \neg\mathbf{B}(G)$, whereby $U \models_{sv} G$. Moreover, U is normal.

$U^{\mathbf{B}+} \cap U^{\mathbf{B}-} = U^{\mathbf{T}+} \cap U^{\mathbf{T}-} = \varnothing$. For otherwise G would belong to the extension of **B** of S_α, whereby G would belong to the anti-extension of **T** of S_α, contradicting the normality of S_α. Moreover, it is easy to see that the extension and the anti-extension of **B** and **T** cannot become overlapping after an application of the successor clause of $\mathcal{R}_1$. Thus $I_1(\mathbf{T}, U)^+_\beta$ is consistent for each β. □

This, of course, implies that G does not come out true at the least fixed point of $\mathcal{R}_2$. In order to make the absolute Gödel sentence come out true at the least fixed point we need to consider an inductive rule that is stronger than $\mathcal{R}_2$.

3.3 The rule $\mathcal{R}_3$ The rule $\mathcal{R}_3$ is defined exactly like $\mathcal{R}_2$ except that the successor clause for $I(\mathbf{B})^-_{\alpha+1}$ now reads:

$$I_3(\mathbf{B})^-_{\alpha+1} = I_3(\mathbf{T})^-_{\alpha+1} \cup \left\{ A \mid \begin{array}{c} \text{for all normal structures } U \supseteq S^3_\alpha\text{: if } U \models_{sv} \mathbf{B}(A), \\ \text{then there is a } \beta \text{ such that } I_1(\mathbf{T}, U)^+_\beta \text{ is inconsistent} \end{array} \right\}$$

The motivating idea behind $\mathcal{R}_3$ is that if it is inconsistent to assume that a sentence φ is *knowable*, then φ is definitely unknowable. This rule will allow us to classify the absolute Gödel sentence as definitely true.

3.4 Properties of $\mathcal{R}_3$

Proposition 3.11 *$\mathcal{R}_3$ is monotone in* **T** *and* **B**.

Proof: The proof is the same as for Proposition 3.6. □

Proposition 3.12 *If U is normal, then for all β: $I_3(\mathbf{B}, U)^+_\beta \subseteq I_3(\mathbf{T}, U)^+_\beta$.*

Proof: The proof is the same as for Proposition 3.5. □

Proposition 3.13 *For all α:*

(a) $I_3(\mathbf{T})^+_\alpha \cap I_3(\mathbf{T})^-_\alpha = I_3(\mathbf{B})^+_\alpha \cap I_3(\mathbf{B})^-_\alpha = \varnothing$.

(b) For all β: $I_1\left(\mathbf{T}, S^3_\alpha\right)^+_\beta$ is consistent.

Proof: As with $\mathcal{R}_2$ (Proposition 3.7), except for case 1, subcase 2, which now goes as follows. We have for all normal $U \supseteq S_\alpha$: if $U \models_{sv} \mathbf{B}(A)$, then there is a β such that $I_1(\mathbf{T}, U)^+_\beta$ is inconsistent. We have $A \in I_3(\mathbf{B})^+_{\alpha+1}$. But since the extension of $\mathbf{B}$ does not grow at successor stages of $\mathcal{R}_3$, we have $A \in I_3(\mathbf{B})^+_\alpha$. This implies that $S_\alpha = \langle I_3(\mathbf{B})_\alpha, I_3(\mathbf{T})_\alpha \rangle \models_{sv} \mathbf{B}(A)$. Therefore there must be a β such that $I_1\left(\mathbf{T}, S^3_\alpha\right)^+_\beta$ is inconsistent, contradicting part (b) of the inductive hypothesis. □

It is easy to see that $\mathcal{R}_3$ is at least as strong as $\mathcal{R}_2$.

Proposition 3.14 *For all A, if* $A \in I_2(\mathbf{T})^+_{\mathcal{R}_2}$, *then* $A \in I_3(\mathbf{T})^+_{\mathcal{R}_3}$.

Proof: It suffices to note that, by normality, if for all normal $U \supseteq S$ such that $U \models_{sv} A$ there is a β such that $I_1(\mathbf{T}, U)^+_\beta$ is inconsistent, then if $S \subseteq S^*$, it must be the case that for all $U^* \supseteq S^*$ such that $U^* \models_{sv} \mathbf{B}(A)$ there is a β such that $I_1(\mathbf{T}, U^*)^+_\beta$ is inconsistent. The result then follows by monotonicity. □

In fact, $\mathcal{R}_3$ is stronger: it classifies the absolute Gödel sentence as true.

Proposition 3.15 $G \in I_3(\mathbf{T})^+_{\mathcal{R}_3}$.

Proof: Take any normal $U \supseteq S_0$ such that $U \models_{sv} \mathbf{B}(G)$. Extend U according to $\mathcal{R}_1$ until you reach a least fixed point U_f. By monotonicity we have $U_f \models_{sv} \mathbf{B}(G)$. And since U is normal, U_f is also normal. So $G \in U_f^{\mathbf{T}+}$. And since U_f is a fixed point, we have $U_f \models_{sv} G$. But we also have $U_f \models_{sv} G \longleftrightarrow \neg\mathbf{B}(G)$. So we have $U_f \models_{sv} \neg\mathbf{B}(G)$, which leaves us with a contradiction.

So $G \in I_3(\mathbf{B})^-_1$. If we extend the process to a least fixed point, we still have $G \in I_3(\mathbf{B})^-_{\mathcal{R}_3}$, that is, $\left\langle I_3(\mathbf{B})_{\mathcal{R}_3}, I_3(\mathbf{T})_{\mathcal{R}_3} \right\rangle \models_{sv} \neg\mathbf{B}(G)$. But since $\left\langle I_3(\mathbf{B})_{\mathcal{R}_3}, I_3(\mathbf{T})_{\mathcal{R}_3} \right\rangle \models_{sv} G \longleftrightarrow \neg\mathbf{B}(G)$, we have $\left\langle I_3(\mathbf{B})_{\mathcal{R}_3}, I_3(\mathbf{T})_{\mathcal{R}_3} \right\rangle \models_{sv} G$. And since we are at a fixed point, we have $G \in I_3(\mathbf{T})^+_{\mathcal{R}_3}$. □

G is an example of a sentence which is in $I_3(\mathbf{B})^-_1$ but not in $I_1(\mathbf{B})^-_\alpha$ for any α. By using suitable coding techniques, for *each* successor stage $\alpha+1$ which is smaller than the closure ordinal of $\mathcal{R}_3$, a new sentence can be found which is in $I_3(\mathbf{B})^-_{\alpha+1}$ but which does not belong to the anti-extension of $\mathbf{B}$ at any stage of the inductive rule which is just like $\mathcal{R}_3$ until stage α and like $\mathcal{R}_1$ afterward.[13] So there is a *strong* sense in which $\mathcal{R}_3$ is an extension of $\mathcal{R}_1$. For instance, take the sentence G' such that

$$\vdash_{PA^{\mathbf{TB}}} G' \longleftrightarrow \left(\neg\mathbf{B}(G) \rightarrow \neg\mathbf{B}\left(G'\right)\right).$$

It is not hard to see that G' first enters the anti-extension of $\mathbf{B}$ of $\mathcal{R}_3$ at stage 2. G' never enters the anti-extension of $\mathbf{B}$ of the rule which is like $\mathcal{R}_3$ until stage 1 and like $\mathcal{R}_1$ afterward.

Proposition 3.16 $\neg K \in I_2(\mathbf{T})^+_{\mathcal{R}_3}$.

Proof: Take any normal $U \supseteq S_0$ such that $U \models_{sv} \mathbf{B}(\neg K)$. Extend U according to $\mathcal{R}_1$ until you reach a least fixed point U_f. By monotonicity, $U_f \models_{sv} \mathbf{B}(\neg K)$. Since U is normal, U_f is also normal. So $\neg K \in U_f^{\mathbf{T}+}$. And since U_f is a fixed point, we have $U_f \models_{sv} \neg K$. But since $U_f \models_{sv} \mathbf{B}(\neg K)$ and since we have

$$U_f \models_{sv} K \longleftrightarrow \mathbf{B}(\neg K),$$

we also have $U_f \models_{sv} K$. Contradiction.

So $\neg K \in I_3(\mathbf{B})^-_1$, whereby $\neg K$ must be an element of $I_3(\mathbf{B})^-_{\mathcal{R}_\ni}$. So

$$\langle I_3(\mathbf{B})_{\mathcal{R}_3}, I_3(\mathbf{T})_{\mathcal{R}_3}\rangle \models_{sv} \neg\mathbf{B}(\neg K).$$

Since also

$$\langle I_3(\mathbf{B})_{\mathcal{R}_3}, I_3(\mathbf{T})_{\mathcal{R}_3}\rangle \models_{sv} K \longleftrightarrow \mathbf{B}(\neg K),$$

we have $\langle I_3(\mathbf{B})_{\mathcal{R}_3}, I_3(\mathbf{T})_{\mathcal{R}_3}\rangle \models_{sv} \neg K$. And since $\langle I_3(\mathbf{B})_{\mathcal{R}_3}, I_3(\mathbf{T})_{\mathcal{R}_3}\rangle$ is a fixed point, we have $\neg K \in I_3(\mathbf{T})^+_{\mathcal{R}_3}$. □

So the least fixed point of $\mathcal{R}_3$ makes the absolute knower sentence come out false, which is in line with our intuitions about this sentence.

Sentences involving *iterated* knowability predicates behave in essentially the same way. For instance, take a sentence K' such that it is provable in $PA^{\mathbf{TB}}$ that $K' \longleftrightarrow \mathbf{BB}(\neg K')$. An argument much like the previous one shows that $\neg K'$ is also in the extension of the truth predicate at the least fixed point of $\mathcal{R}_3$. Again, this is in line with our intuitions about such sentences.

3.5 On the philosophical motivation of $\mathcal{R}_3$ The philosophical argument that was given in Section 3.1 for the intuitive soundness of $\mathcal{R}_2$ does not carry over to the stronger rule $\mathcal{R}_3$. Prima facie the fact that it is inconsistent for a sentence A to be knowable does not by itself ensure that it is *definitely* unknowable. For how can we be sure that in such cases there always is a fact of the matter whether A is knowable?

I will argue against this that $\mathcal{R}_3$'s successor clause for the extension of **B** does have a considerable degree of plausibility. Nevertheless, the support that I am able to give for $\mathcal{R}_3$ is admittedly significantly weaker than the philosophical support that was adduced for $\mathcal{R}_2$.

$\mathcal{R}_3$ can be motivated by means of a comparison with the kind of reasoning that is involved in our evaluation of the liar sentence:

L Sentence **L** is not true.

Using the left-to-right direction ($\mathbf{T}(L) \rightarrow L$) of the naïve Tarskian truth scheme, we see that it is inconsistent to assume that **L** is true. In normal circumstances this would be taken as ample reason to conclude that **L** is false. After all, this would just be a simple instance of a *reduction ad absurdum* inference. But in the present case there are overriding reasons against drawing this inference. For we can use the *right-to-left* direction ($L \rightarrow \mathbf{T}(L)$) of the Tarskian truth scheme to show that it is equally inconsistent to hold that **L** is false. Hence, we conclude in a Kripkean spirit that **L** has no determinate truth value.

Compare this with an evaluation of the absolute Gödel sentence:

G Sentence **G** is subjectively unknowable.

We have seen in Section 2.4 how an instance of the reflexivity principle $\mathbf{B}(A) \rightarrow A$ can be used to show that it is inconsistent to assume that **G** is subjectively knowable. Thus we are again strongly tempted to conclude from our *reductio* that **G** should be classified as subjectively unknowable. And this time there is no overriding reason to resist this temptation. For to argue, by analogy with our argument concerning the liar sentence, that it is also inconsistent to hold that **G** is unknowable, one would need to

appeal to the *converse* $A \rightarrow \mathbf{B}(A)$ of the reflection principle for subjective knowability. But *this* is a principle that has very little intuitive appeal.

In this way one obtains the impression that there can be no overriding reasons against concluding from the inconsistency of the knowability of a sentence to its determinate unknowability (unlike the parallel situation for the notion of truth). And in the absence of such overriding considerations, one ought not to resist the intuitive pull of *reductio ad absurdum*-type of inference patterns. In sum, since we have evidence for it (its intuitive plausibility) and no threat of overriding evidence against it (inconsistencies), we have good reasons to believe the clause for the anti-extension of **B** to be intuitively sound.

So we have good reasons for believing that $\mathcal{R}_3$ is indeed a coherent inductive rule and that it classifies certain paradoxical sentences in what appears to be the "right" way. Of course, it does not follow from this that, perhaps in the context of some additional justifiable constraints, $\mathcal{R}_3$ does not classify some other paradoxical sentences in the "wrong" way (although I have been unable to come up with such examples). As was said before, there is no getting away from the fact that $\mathcal{R}_2$ is substantially more secure than $\mathcal{R}_3$.

3.6 Further issues and open problems It would seem that the rules $\mathcal{R}_2$ and $\mathcal{R}_3$ give rise to a *hierarchy* of inductive rules. In $\mathcal{R}_2$, for instance, $\mathcal{R}_1$ is used as an auxiliary rule. But since we now know $\mathcal{R}_2$ to be an unobjectionable inductive rule, the rule $\mathcal{R}_{2'}$, which is just like $\mathcal{R}_2$ except that *its* auxiliary rule is $\mathcal{R}_2$ instead of $\mathcal{R}_1$, is also unobjectionable, and so on. It is clear that the resulting hierarchy of inductive rules must have a least fixed point. But I am at present unable to see whether there are sentences in the extension of the truth predicate according to $\mathcal{R}_{2'}$ which are not also in the extension of the truth predicate according to $\mathcal{R}_2$. In other words, for all I know this hierarchy of inductive rules may reach a fixed point *very* quickly.

Of course there are complexity questions even for $\mathcal{R}_2$ and $\mathcal{R}_3$. On the one hand, their extensions are at least as complex as the extension of **T** in the least fixed model of the classical Kripkean construction with supervaluations, so they are at least Π^1_1 (see Burgess [4]). But they are also at most Π^1_1, because any inductively defined set (over the standard model of arithmetic) is Π^1_1, even if inductive definitions are iterated (see e.g., [7], pp. 89–90). So their complexity is exactly Π^1_1.[14]

In the inductive rules that we have considered, the extension of **B** was kept constant at all stages. If we would concentrate on *necessity* instead of knowability, things would be somewhat different. First, it would seem to be more natural to work with a Kleene scheme than with supervaluations. But second, we could let the extension of the necessity predicate coincide with the extension of the truth predicate at all stages. For if a sentence φ of $\mathcal{L}_{PA^{\mathbf{TB}}}$ is definitely true, then it is so in virtue of facts about the natural numbers and facts about the logical properties of truth and necessity. But since these facts are all necessary, φ must be necessary also. So the extension of the necessity predicate would be as complex as the extension of the truth predicate.

4 Comparison with context-sensitive approaches Context-sensitive theories of the paradoxes have been proposed in order to validate certain types of intuitively correct strengthened-liar-type reasoning concerning semantic and epistemic notions.

The way in which they are able to deal with the *epistemic* paradoxes is generally considered one of the strengths of these approaches.[15] Within the context-sensitive tradition there are several proposals as to how the epistemic paradoxes should be handled. In Burge's and in Gaifman's versions of the theory, truth and falsehood accrues to *tokens* of sentences. On such an account the sentence token,

G Sentence **G** is subjectively unknowable.

has no truth-value, but by that very fact a distinct token of the sentence type will be true ([9], p. 125). On Barwise and Etchemendy's Austinian account, truth and falsity accrues to *propositions* which have an implicit situation index built in. On this account, the proposition expressed by **G** asserts that its own knowability is not contained in the situation *s* to which **G** is restricted. This proposition is considered true, but the proof witnessing that it is true does not belong to situation *s* ([9], p. 128). The fact witnessing the truth of **G** does, however, belong to a more comprehensive situation s'. A similar analysis is given of other intuitively paradoxical epistemic sentences such as the knower sentence.

It would appear at first sight that—short of moving up to the metalanguage—the context-insensitive theories would find it difficult to recognize a sense in which sentences such as the absolute Gödel sentence and the knower sentence have a definite truth-value. For instance in [11], which purports to give a Kripkean, context-insensitive theory of knowledge, the absolute Gödel sentence and the knower sentence are left without a truth-value.

Nevertheless, the context-insensitive theory that was sketched in the present paper yields evaluations that are more in consonance with the strengthened-liar-type evaluations of the context-sensitive approaches. On this account the absolute Gödel sentence is definitely true and the knower sentence is definitely false. But the proofs of these facts are inaccessible to the knowing agent: the truth of the absolute Gödel sentence and the falsity of the knower sentence cannot be established "from the inside."

Acknowledgments I am indebted to Tony Anderson, John Burgess, Igor Douven, Herman Roelants, Albert Visser, and two anonymous referees for valuable comments on earlier versions of this paper. The research for this paper was supported by a postdoctoral fellowship of the Fund for Scientific Research (Flanders), which is gratefully acknowledged.

NOTES

1. Morgenstern [11] also explores a Kripkean approach to *knowability*. However, her approach differs substaintially from the theory that is developed here (cf. Section 4).
2. For a discussion of the notion of subjective knowability, see Koons [9], p. 46 ff.
3. Thanks to Albert Visser for pointing this out.
4. A referee pointed out that due to the transfinite character of $\mathcal{R}_0$ it is not *immediately* clear that even the successor clause of $\mathcal{R}_0$ for **B** is not too strong. This becomes obvious only when we see that at transfinite stages no new sentences are added to the extension of **B**.
5. He calls the resulting theory KFT.
6. VF stands for a 'van Fraassen'. For a detailed description and a proof-theoretic investigation of VF, see Cantini [5].

7. Koons [9] explores an extension of the Kripke-Feferman system with axioms that govern the notion of subjective knowability ([9], pp. 124–26).
8. For instance, one question concerning such systems that would arise is the following: what is the proof-theoretic strength of the intuitionistic theory that is entailed by such a system under Gödel's modal translation from intuitionistic to (epistemic) classical theories?
9. Otherwise our inductive rules would be badly synchronized, in the sense that there would be stages of the rules at which not everything that is definitely knowable would also be definitely true.
10. Note the similarity between this argument and the standard argument that establishes the truth of the Gödel sentence for, say, Peano Arithmetic.
11. Kaplan and Montague [8] have used this sentence to generate an epistemic paradox: the so-called Paradox of the Knower.
12. Note the similarity between this argument and the standard argument that establishes the falsity of the so-called Jeroslow sentence for, say, Peano Arithmetic.
13. This holds also with $\mathcal{R}_2$ substituted for $\mathcal{R}_3$.
14. Thanks to an anonymous referee for pointing this out.
15. See Anderson [1], Burge [3], and Gaifman [6]. Koons [9] is an excellent survey of applications of context-sensitive approaches to epistemic and doxastic paradoxes.

REFERENCES

[1] Anderson, C. A., "The paradox of the knower," *The Journal of Philosophy*, vol. 80 (1983), pp. 338–55.

[2] Barwise, J., and J. Etchemendy, *The Liar*, Oxford University Press, Oxford, 1987.

[3] Burge, T., "Epistemic paradox," *The Journal of Philosophy*, vol. 81 (1984), pp. 5–29.

[4] Burgess, J., "The truth is never simple," *The Journal of Symbolic Logic*, vol. 51 (1986), pp. 663–81.

[5] Cantini, A., "A theory of formal truth arithmetically equivalent to ID_1," *The Journal of Symbolic Logic*, vol. 55 (1990), pp. 244–59.

[6] Gaifman, H., "Pointers to truth," *The Journal of Philosophy*, vol. 89 (1992), pp. 223–61.

[7] Hinman, P., *Recursion–Theoretic Hierarchies*, Springer–Verlag, New York, 1978.

[8] Kaplan, D., and R. Montague, "A paradox regained," *Notre Dame Journal of Formal Logic*, vol. 1 (1960), pp. 79–90.

[9] Koons, R. C., *Paradoxes of Belief and Strategic Rationality*, Cambridge, Cambridge University Press, 1992.

[10] Kripke, S., "Outline of a theory of truth," pp. 53–81 in *Recent Essays on Truth and the Liar Paradox*, edited by R. Martin, Oxford, Oxford University Press, 1984.

[11] Morgenstern, L., "A first-order theory of planning, knowledge and action," pp. 99–114 in *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the 1986 Conference*, edited by J. Halpern, Morgan Kaufman, Los Altos, 1986.

[12] Reinhardt, W., "Some remarks on extending and interpreting theories with a partial predicate for truth," *Journal of Philosophical Logic*, vol. 15 (1986), pp. 219–51.

Department of Philosophy
University of Leuven
Kardinaal Mercierplein 2
B–3000 Leuven
BELGIUM
email: Leon.Horsten@hiw.kuleuven.ac.be