The Review of Symbolic Logic
Volume 0, Number 0, Month 2014

Solovay-type theorems for circular definitions

SHAWN STANDEFER
University of Pittsburgh

Abstract. We present an extension of the basic revision theory of circular definitions with a unary operator, $\Box$. We present a Fitch-style proof system that is sound and complete with respect to the extended semantics. The logic of the box gives rise to a simple modal logic, and we relate provability in the extended proof system to this modal logic via a completeness theorem, using interpretations over circular definitions, analogous to Solovay's completeness theorem for GL using arithmetical interpretations. We adapt our proof to a special class of circular definitions as well as to the first-order case.

§1 Introduction One of the important discoveries in provability logic is the connection between Peano arithmetic (PA) and the modal logic GL, first demonstrated by Robert Solovay.[1] Solovay showed that GL is complete with respect to provability in PA under all so-called arithmetical interpretations. These interpretations connect necessity in the modal language to provability predicates in the arithmetical language.

Revision theory is a general theory of circular definitions. It was originally developed as a theory of truth by Anil Gupta and, independently, Hans Herzberger, with important early contributions by Nuel Belnap.[2] The theory was generalized to a theory of circular definitions in (Gupta 1988–89), which was further elaborated in (Gupta & Belnap, 1993).[3]

In this paper, we will show that there is a connection, similar to the one between PA and GL, between the circular definitions of revision theory and a particular modal logic, which we call "RT," for revision theory.[4] We will prove that RT is complete with respect to validity under all $\mathscr{D}$-interpretations for all sets of circular definitions $\mathscr{D}$.

The modal logic RT arises naturally from an extension of revision theory that we will present below. (Gupta & Standefer, 2014) presents an extension of revision theory that uses different primitives, which have independent philosophical interest. The extension presented here adds a unary connective, $\Box$, to basic revision theory. This connective can be glossed as saying, roughly, "according to the current hypothesis," or, in the context of a revision sequence, "at the previous stage." The modal logic RT is the logic one obtains from viewing the box simply as a modal operator.

[1] See (Solovay, 1976) or (Boolos, 1993).
[2] See (Gupta, 1982), (Herzberger, 1982), and (Belnap, 1982), respectively.
[3] Many logicians have done fruitful work on revision theory. A partial list of contributions includes (Kremer, 1993), (Yaqūb, 1993), (Antonelli, 1994), (Chapuis, 1996), (Orilia, 2000), (Löwe & Welch, 2001), (Welch, 2001), (Kühnberger et al., 2005), (Horsten et al., 2012), (Asmus, 2013), and (Bruni, 2013).
[4] An anonymous referee has pointed out that this logic is also known in the literature as "$KD!$".

doi:10.1017/S1755020300000000

The addition of $\Box$ to revision theory increases the expressive power of the theory. The most striking demonstration of the increase in expressive power is that the box permits the definition of an object language correlate of the definitional clauses of a set of circular definitions. Whenever a circular definition, say,

$$Gx =_{Df} A(x, G),$$

is used, then the object language sentence

$$\forall x(Gx \equiv \Box A(x, G))$$

will be valid. In basic revision theory, it is not generally true that a circular definitional clause will be reflected in an object language validity.

We will begin by motivating and presenting the fundamental definitions for extended revision theory (§2). Once the basics are in place, we will state an important regularity theorem for the extended theory, the proof of which is left to an appendix (§4). We briefly present the rules to add to the Fitch-style proof system for basic revision theory to obtain a complete proof system for extended revision theory. We then prove three Solovay-type completeness theorems, two for *RT* (§3.1, §3.2) and one for its first-order form *RTQ* (§3.3). The proof of §3.2 uses a notion from basic revision theory, that of finite definition, and in §5, we explain how to generalize this notion to the extended theory.

§2 An extension of revision theory In this section, we will present an extension of revision theory.[5] Revision theory provides a semantic treatment of circularly defined predicates.[6] These predicates may have circular and interdependent definitions of the following form.[7]

$$\begin{array}{c} G_1(\overline{x}_1) =_{Df} A_1(\overline{x}_1, G_1, \ldots, G_k) \\ G_2(\overline{x}_2) =_{Df} A_2(\overline{x}_2, G_1, \ldots, G_k) \\ \vdots \\ G_k(\overline{x}_k) =_{Df} A_k(\overline{x}_k, G_1, \ldots, G_k) \end{array}$$

Any of the *definienda*, G_i, may appear in any of the *definientia*, A_j. In any A_i, only the variables $\overline{x}_i$ may occur freely. We will adopt the convention of using G for circularly defined predicates.

We will work with languages containing constants and variables but no function symbols. We start with a ground model $M(= \langle D, I\rangle)$ that interprets a base language $\mathscr{L}$, which is extended to a new language, $\mathscr{L}^+$, that contains the circularly defined predicates and $\Box$. A set of circular definitions $\mathscr{D}$ provides a *revision operator*, $\Delta_{\mathscr{D},M}$, which is used to revise *hypotheses* about what satisfies the G_i.[8] Given a

[5] For a full exposition of basic revision theory, see (Gupta & Belnap, 1993).

[6] Revision theory can handle other types of expressions, but we will focus on predicates here.

[7] Sets of definitions can be infinite.

[8] We will use 'Δ' for revision operators in the extended theory and, when needed, 'δ' for revision operators in the basic theory.

hypothesis h about what satisfies G_i and information from the ground model, the revision operator generates a new hypothesis, $\Delta_{\mathscr{D},M}(h)$. The revision operator can be iterated, and these possibly transfinite sequences of hypotheses form *revision sequences*, which are central to the semantics of revision theory.

We will now explain this sketch in more detail. We begin by motivating the new definition of hypothesis and we define some important related concepts, including similarity, correspondence and falling under a hypothesis (§2.1). These concepts are used to define revision of hypotheses and to state the semantic clauses for the box and for circularly defined predicates (§2.2). These two sections cover the basics of extended revision theory. In §2.3, we state the definitions of revision sequences and of validity, which are basically unchanged from basic revision theory. Finally, we explain how to modify a Fitch-style proof system for basic revision theory so that it is sound and complete with respect to semantic consequence in extended revision theory (§2.4).

2.1 Hypotheses and related definitions In extended revision theory, the hypotheses are used to interpret not only the circularly defined predicates, but also all *boxed formulas*, formulas whose main connective is $\Box$. Thus, hypotheses must be extended from guesses about satisfaction for the *definienda* to guesses about satisfaction for the whole language.[9] This presents some immediate difficulties, for we would like formulas that are not identical but intuitively "say the same thing," such as $\Box\exists xRxy$ and $\Box\exists yRyz$, to be satisfied by precisely the same objects. To overcome this difficulty, we define hypotheses so that they are certain equivalence classes of pairs of formulas and assignments to variables.

To define hypotheses, we must introduce the auxiliary concept of *similarity*, for which we need a few definitions. Let $\mathscr{F}$ be the set of formulas of $\mathscr{L}^+$ containing no names. Let $\mathscr{F}_1$ be the set of formulas from $\mathscr{F}$ each of which contains at most one free occurrence of each variable. Let $\mathscr{V}_M$ be the set of assignments of values to variables relative to a model M.

Informally, similarity is a relation between pairs of formulas and assignments. Two pairs are similar when their formulas are the same up to relettering of bound variables and their assignments agree on the free variables "in the same positions." Similarity matters because hypotheses will be defined as subsets of $\mathscr{F} \times \mathscr{V}_M$ closed under similarity. The sense of relettering is made precise by the notion of an *alphabetic variant*. We will say that A is a one-step variant of B iff A has an occurrence a subformula $\forall xC(x)$ where B has $\forall yC(y)$, where y does not occur freely in C and y is free for x in C. A is an alphabetic variant of B just in case there is a sequence, $A = D_0, \ldots, D_n = B$, such that each D_i is a one-step variant of D_{i+1}, with the possibility that $n = 0$.[10]

Definition 1 (Similarity)

Let A and B be formulas in $\mathscr{F}$ and v and v' assignments to variables. Then, $\langle A, v\rangle$ is similar *to $\langle B, v'\rangle$ iff A and B each have exactly n occurrences of free variables and there is a formula $C(x_1, \ldots, x_n) \in \mathscr{F}_1$, whose free variables are all and only $x_1, \ldots, x_n$, none of which occur in A or B, such that*

[9] One can define hypotheses differently so that they make guesses about a proper subset of the language. We find the definition given below easier to use.

[10] This definition is based on (Hughes & Cresswell, 1996, 240).

1. *for some variables* $y_1, \ldots, y_n$, A *is an alphabetic variant of* $C(y_1, \ldots, y_n)$,
2. *for some variables* $z_1, \ldots, z_n$, B *is an alphabetic variant of* $C(z_1, \ldots, z_n)$, *and*
3. *for all* $i \leq n$, $v(y_i) = v'(z_i)$.

An example of similar pairs is $\langle \Box \exists x Rxy, v \rangle$ and $\langle \Box \exists y Ryz, u \rangle$, where $v(y) = u(z)$. Another example, is $\langle Rxx, v \rangle$ and $\langle Ryz, u \rangle$, where $v(x) = u(y) = u(z)$. We note that similarity is an equivalence relation.

We can now define hypotheses as sets of pairs of formulas and assignments that are closed under similarity.

Definition 2 (Hypothesis)
A hypothesis h *is a subset of* $\mathscr{F} \times \mathscr{V}_M$ *such that for all similar pairs* $\langle A, v \rangle$ *and* $\langle B, u \rangle$, $\langle A, v \rangle \in h$ *iff* $\langle B, u \rangle \in h$

Since hypotheses are closed under similarity, they respect alphabetic variants. This fact is used in showing that the classical quantifier rules are sound for arbitrary hypotheses.[11]

Before proceeding to the semantics for the box, we must deal with the restriction of the hypotheses to $\mathscr{F}$. Consider the sentence $\Box Fb$. We will want to say that $\Box Fb$ is true relative to a hypothesis h just in case h contains the pair $\langle Fb, v \rangle$, for some v, but no hypothesis will contain the pair $\langle Fb, v \rangle$, as $Fb \notin \mathscr{F}$. To overcome this hurdle, we allow pairs to *correspond* to pairs that are in some hypotheses.

Definition 3 (Corresponds)
A pair $\langle A, v \rangle$ corresponds in M *to a pair* $\langle B, v' \rangle$ *iff* $B \in \mathscr{F}$, *and there are sequences* $\langle x_1, \ldots, x_n \rangle, \langle y_1, \ldots, y_m \rangle$ *and* $\langle c_1, \ldots, c_m \rangle$ *such that* B *has exactly the variables* $x_1, \ldots, x_n, y_1, \ldots, y_m$ *free, no* x_i *is* y_j, *the* c_i *are all distinct names, the* y_i *are all distinct variables, and*

1. $A = B(x_1, \ldots, x_n, c_1, \ldots, c_m)$,
2. *for all* $i, 1 \leq i \leq n, v'(x_i) = v(x_i)$,
3. *for all* $i, 1 \leq i \leq m, v'(y_i) = I(c_i)$.[12]

When a pair $\langle A, v \rangle$ corresponds to a pair in the hypothesis h, we say that the pair $\langle A, v \rangle$ falls under h.

Definition 4 (Falling under)
Let C *be a formula let* v *be an assignment to variables. Then* $\langle C, v \rangle$ falls under h relative to M, *in symbols* $\langle C, v \rangle \in_M h$, *iff there is a pair* $\langle C', v' \rangle$ *such that* $\langle C, v \rangle$ *corresponds in* M *to* $\langle C', v' \rangle$ *and* $\langle C', v' \rangle \in h$.

2.2 Revision semantics With the definitions of hypotheses and related notions in hand, we can give the definition of revision and the semantics for the box. First, a bit of notation. If h is a hypothesis, then $M + h$ is the model just like M, except that h is used to interpret circularly defined predicates and boxed formulas.

Revision operators are defined as follows.

[11] The formula in $\mathscr{F}_1$ in the definition of similarity plays a part in showing that identity elimination is sound for arbitrary hypotheses. We believe that both the $\mathscr{F}_1$ clause and the use of alphabetic variants can be dropped, in which case the revision process would take care of the equivalences; the soundness proof for the resulting semantics would be slightly more complex.

[12] We will drop "in M" when talking about correspondence when the model is clear.

Definition 5 (Revision operator)

The revision operator $\Delta_{\mathscr{D},M}$ is an operation from hypotheses to hypotheses that satisfies the following condition for all $A \in \mathscr{F}$ and $v \in \mathscr{V}_M$.

$$M + h, v \models A \Leftrightarrow \langle A, v\rangle \in \Delta_{\mathscr{D},M}(h)$$

Before revision, hypotheses may disagree about base language formulas. This is not a problem, since hypotheses are only used to determine satisfaction for boxed formulas and for defined predicates. For formulas entirely in the base language, the ground model is used to determine satisfaction.

The truth conditions for the box are simple: $\Box A$ is true in $M + h$ just in case A is true according to h. More generally,

$$M + h, v \models \Box A \Leftrightarrow \langle A, v\rangle \in_M h.$$

In terms of revised hypotheses,

$$M + \Delta_{\mathscr{D},M}(h), v \models \Box A \Leftrightarrow M + h, v \models A.$$

The box is, in a sense, a *cross-stage* connective. Whether $\Box A$ is true at a stage depends on the truth value of A at another stage. This differs from all the classical connectives, which are *same-stage* connectives. Whether, say, $A \supset B$ is true at a stage depends only on the truth values of A and B at that same stage.

Since the definition of hypothesis differs from that of basic revision theory, we will define the interpretation of the defined predicates. In basic revision theory, the hypotheses directly assign extensions to defined predicates. In the extended theory, extensions are not needed, as satisfaction for defined predicates is defined as follows.[13]

$$M + h, v \models G\bar{t} \Leftrightarrow \langle A(\bar{t}, G), v\rangle \in_M h$$

One important philosophical aspect of revision theory is the claim that revision yields better hypotheses.[14] It will be worthwhile to elaborate on that claim with respect to the box. Revision improves hypotheses with respect to the box in a straightforward way: As revision proceeds, iterations of the box become increasingly compositional. The result is a simple logic for the box.

The improvement brought about by revision can be captured in a regularity theorem. As long as hypotheses agree on the formulas occurring in the set $\mathscr{D}$ of circular definitions, in both *definientia* and *definienda*, then the hypotheses will increasingly agree after revision. This motivates the following definitions.

Definition 6 ($sub(\mathscr{D})$, $\equiv_{\mathscr{D}}$)

Let $sub(\mathscr{D})$ be the set of subformulas of the definienda *and the* definientia *in $\mathscr{D}$. If $B \in sub(\mathscr{D})$, we will say that B is a subformula of $\mathscr{D}$.*

Let h and h' be hypotheses. Define $h \equiv_{\mathscr{D}} h'$ iff $\forall B \in sub(\mathscr{D}), \forall v \in \mathscr{V}_M$,

$$\langle B, v\rangle \in_M h \Leftrightarrow \langle B, v\rangle \in_M h'.$$

---

[13] One can recover extensions for defined predicates in a straightforward way. Note that extensions are needed with some alternative definitions of similarity.

[14] This point is made by (Gupta & Belnap, 1993, 121). For discussion, see (Shapiro, 2006) and (Gupta, 2011, 160-161).

The relation $\equiv_{\mathscr{D}}$ is an equivalence relation.

Define $\Delta^n_{\mathscr{D},M}$ by recursion as $\Delta^0_{\mathscr{D},M}(h) = h$ and $\Delta^{n+1}_{\mathscr{D},M}(h) = \Delta_{\mathscr{D},M}(\Delta^n_{\mathscr{D},M}(h))$. Let the *modal depth* of A, $d(A)$, be the greatest number of nested boxes occurring in A. We can now state the Regularity Theorem.

Theorem 1 (Regularity Theorem) *Suppose $h \equiv_{\mathscr{D}} h'$. If $d(A) \leq n$, then for all $m \geq n$,*

$$M + \Delta^m_{\mathscr{D},M}(h), v \models A \Leftrightarrow M + \Delta^m_{\mathscr{D},M}(h'), v \models A.$$

We leave the proof to the appendix (§4) . The Regularity Theorem has the following corollary, which highlights another sense in which revision improves hypotheses with respect to the box.

Corollary 2.1. *For all formulas A that contain no* definienda*, for all n, if $d(A) \leq n$ then for all $m \geq n$,*

$$M + \Delta^m_{\mathscr{D},M}(h), v \models A \Leftrightarrow M + \Delta^m_{\mathscr{D},M}(h'), v \models A.$$

2.3 Revision sequences and semantic consequence We will now give the formal definitions of revision sequences and validity. A note on the notation: we use $\mathscr{S}$ for sequences of hypotheses and $\mathscr{S}_\alpha$ to denote the αth element of $\mathscr{S}$.

Definition 7 (Stably in/out, coherence, revision sequence)

Let On be the class of all ordinals.

Let λ be a limit ordinal and $A \in \mathscr{F}$. $\langle A, v\rangle$ is stably in [stably out of] *$\mathscr{S}$ at λ iff for all $\langle A, v\rangle \in h$*

$$\exists\alpha < \lambda \forall\beta(\alpha \leq \beta < \lambda \supset \langle A, v\rangle \in [\not\in]\mathscr{S}_\beta)$$

A hypothesis h coheres *with $\mathscr{S}$ at λ iff*

1. *if $\langle A, v\rangle$ is stably in $\mathscr{S}$ at λ, then $\langle A, v\rangle \in h$, and*
2. *if $\langle A, v\rangle$ is stably out of $\mathscr{S}$ at λ, then $\langle A, v\rangle \notin h$.*

$\mathscr{S}$ is a revision sequence *for $\mathscr{D}$ in M iff $\mathscr{S}$ is an On-long sequence of hypotheses and for all ordinals α and β,*

1. *if $\alpha = \beta + 1$, then $\mathscr{S}_\alpha = \Delta_{\mathscr{D},M}(\mathscr{S}_\beta)$, and*
2. *if α is a limit ordinal, then $\mathscr{S}_\alpha$ coheres with $\mathscr{S}$ at α.*

We will, following Gupta and Belnap, define two concepts of validity: S_0 and $S^\#$. For the latter, we need the following definitions.

Definition 8 (Cofinal hypothesis, recurring hypothesis) *A hypothesis h is* cofinal *in a revision sequence $\mathscr{S}$ for $\Delta_{\mathscr{D},M}$ iff $\forall\alpha\exists\beta \geq \alpha(\mathscr{S}_\beta = h)$.*

A hypothesis h is recurring *for $\Delta_{\mathscr{D},M}$ iff h is cofinal in some revision sequence for $\Delta_{\mathscr{D},M}$.*

One can establish that all definitions have cofinal and recurring hypotheses.[15]

We are now in a position to define the two notions of validity.

[15] See Theorem 5C.7 of (Gupta & Belnap, 1993). The proofs here proceed in the same way.

Definition 9 (S_0 validity)
Given a set of definitions $\mathscr{D}$, a sentence A is valid in M on $\mathscr{D}$ *in S_0, in symbols $M \models_0^{\mathscr{D}} A$, iff there is a natural number n, such that, for all hypotheses h, A is true in $M + \Delta^n_{\mathscr{D},M}(h)$. A sentence A is valid on $\mathscr{D}$ in S_0 iff A is valid in M on $\mathscr{D}$ in S_0 for all ground language models M, or in symbols $\models_0^{\mathscr{D}} A$.*

Definition 10 ($S^\#$ validity)
Given a set of definitions $\mathscr{D}$, a sentence A is valid in M on $\mathscr{D}$ *in $S^\#$, in symbols $M \models_\#^{\mathscr{D}} A$, iff for all recurring hypotheses h, there is a natural number n, such that for all $m \geq n$, A is true in $M + \Delta^m_{\mathscr{D},M}(h)$. A sentence A is valid in $S^\#$ on $\mathscr{D}$ iff A is valid in M in $S^\#$ for all models M of the ground language, or in symbols $\models_\#^{\mathscr{D}} A$.*

2.4 Proof system There is a Fitch-style proof system for basic revision theory: the system C_0.[16] The system uses indexed formulas, A^i, where the index can be any integer. The indices track the relative stage of revision for a given formula. The rules for the logical connectives require their premises and conclusions to have the same index, as in the following.

$$\left|\begin{array}{ll} A^i & \\ (A \vee B)^i & \vee\text{I} \end{array}\right. \qquad \left|\begin{array}{ll} (A \,\&\, B)^i & \\ B^i & \&\text{E} \end{array}\right.$$

The only rules that change the indices in C_0 are the *index shift rule* and the *definition rules*. Index shift permits one to go from A^i to A^k, provided that A contains no defined predicates. For each definition $G\overline{x} =_{Df} A_G(\overline{x})$ in $\mathscr{D}$, there is a pair of definition rules.[17]

$$\left|\begin{array}{ll} A_G(\overline{t})^i & \\ G(\overline{t})^{i+1} & \text{DefI} \end{array}\right. \qquad \left|\begin{array}{ll} G(\overline{t})^{i+1} & \\ A_G(\overline{t})^i & \text{DefE} \end{array}\right.$$

To accommodate the new connective, $\Box$, we add the following rules to C_0 to get the system $C_0^\Box$.

$$\left|\begin{array}{ll} A^i & \\ (\Box A)^{i+1} & \Box\text{I} \end{array}\right. \qquad \left|\begin{array}{ll} (\Box A)^{i+1} & \\ A^i & \Box\text{E} \end{array}\right.$$

We will use the turnstile, $\vdash_0^{\mathscr{D}}$, for provability in $C_0^\Box$ and $\vdash_0^{\mathscr{D}} A$ to indicate that A^0 is provable in $C_0^\Box$. We then have the following.

Theorem 2 (Soundness and Completeness) *Let A be a sentence. Then,*

$$\vdash_0^{\mathscr{D}} A \Leftrightarrow \models_0^{\mathscr{D}} A.$$

The proof of soundness is a standard induction on the construction of a proof, so we omit it. The proof of completeness uses the modified Henkin construction from (Gupta & Belnap, 1993). Some care must be taken in defining the hypotheses, but it is otherwise straightforward; we omit the proof.

[16] (Gupta & Belnap, 1993, 157-160)
[17] The terms $\overline{t}$ must be free for $\overline{x}$ in $A_G(\overline{x})$.

Soundness holds for $S^{\#}$, but completeness, in general, fails. In basic revision theory, the completeness result holds for a restricted class of definitions, the *finite definitions*.[18] Finite definitions are ones for which the revision process come to an end, in the sense of not generating new hypotheses, after a finite number of steps. Once the notion of finite definition is suitably generalized, the completeness result holds for extended revision theory as well. Finite definitions will play a role in one of the theorems proved below (§3.2). Let us now turn to the Solovay-type theorems.

§3 Solovay-type completeness theorems In this section we will prove some Solovay-type completeness theorems linking extended revision theory and the modal logics RT and RTQ. We will begin with some definitions to establish terminology. We will use standard definitions of Kripke models.[19] A Kripke model M is a triple (W, R, V), where W is a non-empty set of worlds, $R \subseteq W \times W$, and V is an interpretation. For the propositional case, we will use the notation, $M, w \Vdash p$, for $V(p, w) = \mathbf{t}$. The '$\Vdash$' relation is extended to complex sentences in the standard way. For the first-order case, the '$\Vdash$' notation will be adapted in the obvious way. We will use only first-order Kripke models with constant domains, which we explain below.

The modal logic we will investigate, RT, is obtained by adding to the modal logic K the axiom

$$\sim\Box A \equiv \Box\sim A.^{20}$$

The logic is sound and complete for models in which every world has exactly one R-successor.

The first-order logic RTQ is obtained by adding to RT axioms for quantifiers, the rule of generalization, and both directions of the Barcan formula.

$$\forall x \Box Ax \equiv \Box \forall x Ax$$

Adding both directions of the Barcan formula will restrict us to constant domain Kripke models. The reason for the restriction is that in a revision sequence, the domain of the ground model does not change from one stage to the next. Consequently, the box of revision theory will obey both directions of the Barcan formula, and so studying the first-order modal logic of the box is done most naturally with the Barcan formulas.

From a certain perspective, RT models look like revision sequences. There is something to this idea, but there is an important difference. In revision sequences, the box and the "accessibility relation" of the revision sequence go in the same direction. If p is true at stage k, $\Box p$ is true at stage $k + 1$. In the Kripke models, the box looks down the accessibility relation. If p is true at w, then for all u such that uRw, $\Box p$ is true at u. This difference does not undermine the noted analogy, as we will see.

There is a connection between the modal logic RT, provability in $C_0^{\Box}$, and validity for circular definitions. This connection is similar to Solovay's arithmetical

[18] See (Martinez, 2001) and (Gupta, 2006) for more on finite definitions.

[19] See (Hughes & Cresswell, 1996), for example.

[20] A version of the axiom using both modalities is $\Diamond A \equiv \Box A$. We will officially treat '$\Diamond$' as defined in terms of box and negation.

completeness theorem relating the provability logic GL and provability in PA.[21] The connection between GL and PA is made via *arithmetical interpretations*. An arithmetical interpretation * for a propositional modal language is a function that maps sentences of the modal language to sentences of PA, defined as follows.

- $(p)^* \in Sent_{\mathscr{L}_{PA}}$
- $(\bot)^* = \bot$
- $(\sim A)^* = \sim(A^*)$
- $(A \supset B)^* = (A^*) \supset (B^*)$
- $(\Box A)^* = Pr(\ulcorner A^* \urcorner)$[22]

The conceptually most important clause is the last: the box is interpreted as a provability predicate. Solovay's completeness theorem makes the following important connection between GL and PA.

Theorem 3 (Solovay's Theorem) $GL \vdash A \Leftrightarrow$ *for all arithmetical interpretations* *, $PA \vdash A^*$.

For a proof see (Boolos, 1993, 125-131). Different completeness theorems can be proved by restricting to different sets of sentences for the interpretation of atoms. First-order interpretations can be defined as well. Here atomic formulas are mapped to formulas of PA with the same free variables and the interpretations commute with the quantifiers. The analogous first-order completeness theorem does not hold.[23] In this section we will prove similar completeness theorems for the logics RT and RTQ, using interpretations based on languages with circular definitions rather than arithmetic sentences. We will begin by proving the theorem for the simplest case, propositional logic (§3.1). This will permit us to clearly demonstrate the proof technique. We will use the technique to prove another theorem for the propositional case (§3.2) and a completeness theorem for the first-order case (§3.3).

3.1 A propositional Solovay-type theorem The first task is to define the relevant notion of interpretation. Given a base language $\mathscr{L}$, expand it to $\mathscr{L}^+$ with the addition of $\Box$ and a set of definitions $\mathscr{D}$. Let a $\mathscr{D}$-interpretation * be a function from sentences of RT to sentences of $\mathscr{L}^+$ satisfying the following.

- $(p)^* \in Sentence_{\mathscr{L}^+}$
- $(\bot)^* = \bot$
- $(\sim A)^* = \sim(A^*)$
- $(A \circ B)^* = (A^*) \circ (B^*)$, for $\circ \in \{\&, \vee, \supset\}$
- $(\Box A)^* = \Box(A^*)$

Note that in the final clause, the box on the left is that of RT while the box on the right is that of extended revision theory.

We are now ready to state the Solovay-like theorem relating RT and $C_0^{\Box}$.

Theorem 4 *Let $\mathscr{L}$ be a ground language with at least one name.*

$$RT \vdash A \Leftrightarrow \forall \mathscr{D}, \forall \mathscr{D}\text{-interpretations } ^*, \vdash_0^{\mathscr{D}} (A^*)$$

[21] Solovay had other results in this area. We state only one here. For more, see (Boolos, 1993), especially chapter 5.

[22] The notation, $\ulcorner A \urcorner$, stands for the Gödel number of A. The notation is specified only relative to a particular Gödel numbering.

[23] See (Boolos, 1993, Ch. 17-18) for details.

In §3.2, we will prove a version of the theorem for a restricted class of box-free definitions. The proofs of the two theorems have some interesting conceptual differences. In the proof of the theorem for all definitions, the work primarily goes into the definition of the hypotheses. The circular definitions and the interpretations are relatively straightforward. By contrast, in the proof of the theorem for the box-free definitions, the work goes into specifying the set of circular definitions and the appropriate interpretations, and the starting hypotheses do not matter.

We will briefly outline the proof strategy before providing the details. The left-to-right direction is proved by a straightforward induction, showing that the $\mathscr{D}$-interpretations of each line of an RT proof is derivable in $C_0^{\Box}$.

For the right-to-left direction, we argue contrapositively. Let the *box normal form*, A_{bnf}, of A be the formula obtained by distributing all boxes past all other connectives. In RT, A is equivalent to A_{bnf}, as noted below in corollary 3.2.. Assume that $RT \nvdash A$, so $RT \nvdash A_{bnf}$. We use the completeness of RT to obtain a canonical model falsifying A_{bnf}. This gives a finite set of worlds and a distribution of truth values falsifying A_{bnf}. This is translated to a finite number of steps of a revision sequence, which allows us to determine the definition needed to produce that pattern of truth values. Finally, we show that the transformations used to obtain the box normal form are provably equivalent in $C_0^{\Box}$.

Figure 1 summarizes the proof strategy for the right-to-left direction. The double

$$\begin{array}{ccc} RT \nvdash A & \rightsquigarrow & \nvdash_0^{\mathscr{D}} A^* \\ \Downarrow & & \Uparrow \\ RT \nvDash A_{bnf} & \Rightarrow & \nvDash_0^{\mathscr{D}} (A_{bnf})^* \end{array}$$

Fig. 1. Proof strategy for the Solovay-type completeness theorems

line arrows indicate the steps taken in the proof. The wavy arrow is the desired conclusion. The two steps represented by the vertical arrows on the left and right are accounted for by completeness and soundness results, respectively. The primary contribution of this section is the step represented by the arrow along the bottom, namely showing how to determine an appropriate $\mathscr{D}$ and an invalidating model.

We begin by highlighting some useful features of RT.

Lemma 1 *RT has the following theorems.*

1. $\Box{\sim}A \equiv {\sim}\Box A$
2. $\Box(A \supset B) \equiv (\Box A \supset \Box B)$
3. $\Box(A \vee B) \equiv (\Box A \vee \Box B)$
4. $\Box A \vee \Box{\sim}A$
5. ${\sim}\Box\bot$
6. $\bot \equiv \Box^n \bot$, *for all n*
7. $\Box^n(A \supset B) \equiv (\Box^n A \supset \Box^n B)$, *for all n.*

The lemma suffices to establish the following.

COROLLARY 3.2. *For every A, $RT \vdash A \equiv A_{bnf}$.*

Theorem 5 (RT completeness) *RT is sound and complete with respect to the class of Kripke models in which every world has exactly one R successor.*

Proof. This is proved using the methods of (Hughes & Cresswell, 1996). □ We

will make use of a theorem concerning k-restricted models, which we now define. Let M be an RT model, w a world of M, and k a natural number. We will say that $M'(= \langle W', R', V' \rangle)$ is the *k-restriction of M at w* iff W' is the set of all worlds $w' \in W$ that can be reached in at most k-many R-steps from w, that is, that $w_0 R w_1 R, \ldots, R w_k$, with $w = w_0$ and R' and V' are the appropriate restrictions of R and V.

Theorem 6 *Let M be an RT model, w a world of M, and k a natural number. Let $M'(= \langle W', R', V' \rangle)$ be the k-restriction of M at w. Then, for all $j \leq k$, for all A such that $d(A) \leq (k - j)$, $M, w_j \Vdash A$ iff $M', w_j \Vdash A$.*[24]

Let us note the first-order version of this holds for the constant domain models of RTQ.

Let us say that a formula A is a *boxed atom* iff for some $n \geq 0$, it has the form $\Box^n p$, where p is an atom. A *maximal boxed atom occurrence* is an occurrence of a boxed atom that does not occur as a subformula of another boxed atom. As an example, in the sentence $\Box^4 p \,\&\, \Box^2 p \,\&\, p$ there are three maximal boxed atom occurrences, $\Box^4 p$, $\Box^2 p$, and p. We will generally restrict attention to maximal boxed atom occurrences. Let us begin the proof.

Assume that $RT \nvdash A$, so $RT \nvdash A_{bnf}$. Then by the completeness theorem, there is a model M and a world w_0 such that $M, w_0 \nVdash A_{bnf}$. Let $d(A)$ be k. Then there are $k + 1$ worlds, such that $w_0 R w_1 \ldots R w_k$. For each maximal boxed atom occurrence, $\Box^n p$, of A, $M, w_0 \Vdash \Box^n p$ just in case $M, w_n \Vdash p$. We associate with each atom q of A, a $k+1$-long sequence of truth values, $\mathbf{q}$, such that $\mathbf{q}_i = \mathbf{t}$ just in case $M, w_i \Vdash q$, otherwise $\mathbf{q}_i = \mathbf{f}$. We can, by theorem 6, restrict our attention to just the pattern of values of the $\mathbf{q}$'s. We will use the set of $\mathbf{q}$'s to define the required model and hypothesis.

Next, we define a set of circular definitions central to our proof.

Definition 11 *Let $\mathscr{G}$ be the set of definitions containing, for each n, m,*

$$G_n^m x =_{Df} \Box^n G_n^m x.$$

In constructing the invalidating interpretation, we will assign to each atom of A_{bnf} a defined predicate from $\mathscr{G}$. It is helpful to note that A_{bnf} is a truth-functional combination of maximal boxed atom occurrences, so the invalidating interpretation will have the same structure. Only finitely many distinct atoms can appear in A_{bnf}, so a finite $\mathscr{D} \subseteq \mathscr{G}$ will suffice for the rest of the proof. In fact, the subset we will use is even simpler, since all the G's in the set of definitions will have the same subscript. It is straightforward to see that for any finite $\mathscr{D} \subseteq \mathscr{G}$, the extensions assigned to $\mathscr{D}$ will cycle in a pattern with a finite period.[25]

[24] See (Blackburn et al., 2002, 76) for a more general version of this theorem. Ours is specific to RT.

[25] Here and below, we will talk about the extension of a predicate according to a hypothesis. It can be eliminated at the cost of making it more cumbersome to state certain portions of the proof. The extension of a formula according to a hypothesis can be defined from the basic definition of a hypothesis. The extension of an n-ary formula $\Box B(\overline{x})$ according to h is the set of tuples $\overline{d}$ such that for some v, $v(\overline{x}) = \overline{d}$

At this point, it may be helpful to see an example of a cycle of definitions from $\mathscr{G}$. Take $Gx =_{Df} \Box^3 Gx$ and some model M. The revisions of a given hypothesis h cycle in a simple pattern, as illustrated by table 1.[26] For the definition of G^m_n,

Table 1. *Pattern of truth values across revisions*

	h	$\Delta_{\mathscr{D},M}(h)$	$\Delta^2_{\mathscr{D},M}(h)$	$\Delta^3_{\mathscr{D},M}(h)$	$\Delta^4_{\mathscr{D},M}(h)$	$\Delta^5_{\mathscr{D},M}(h)$	...
Ga	**t**	**t**	**f**	**f**	**t**	**t**	
$\Box Ga$	**f**	**t**	**t**	**f**	**f**	**t**	
$\Box^2 Ga$	**f**	**f**	**t**	**t**	**f**	**f**	
$\Box^3 Ga$	**t**	**f**	**f**	**t**	**t**	**f**	

the initial hypothesis will return to its values over the subformulas of the definition every $n+1$ revisions, with the pattern of satisfaction lagging one stage behind, as illustrated by table 2.

Table 2. *Pattern of satisfaction across revisions*

$\models$	h	$\Delta_{\mathscr{D},M}(h)$	$\Delta^2_{\mathscr{D},M}(h)$	$\Delta^3_{\mathscr{D},M}(h)$	$\Delta^4_{\mathscr{D},M}(h)$	$\Delta^5_{\mathscr{D},M}(h)$	...
Ga	**t**	**f**	**f**	**t**	**t**	**f**	
$\Box Ga$	**t**	**t**	**f**	**f**	**t**	**t**	
$\Box^2 Ga$	**f**	**t**	**t**	**f**	**f**	**t**	
$\Box^3 Ga$	**f**	**f**	**t**	**t**	**f**	**f**	

The RT countermodel N provides a $k+1$-long sequence of R-successor worlds, starting with w_0. This provides a sequence of truth values, $\mathbf{q}$, for each atom q of A_{bnf}. The desired set of definitions is the set $G^1_k, \ldots, G^j_k$, where j is the number of distinct atoms in A. We can use any $\mathscr{D}$-interpretation such that sets $(q_i)^* = G^i_k a$, for each atom q_i of A. The interpretation of all other atoms can be arbitrary.

Next, we must construct the hypotheses to be used to falsify the interpretation of A_{bnf}. The model M can be arbitrary. For the hypotheses, we use the sequences $\mathbf{q}$ to determine the values for the G^i_k at each stage of revision. Define the sequence $\langle h_m \rangle_{m \in \omega}$ as follows.

- $\langle \Box^n G^i_n a, v \rangle \in_M h_0$ iff $N, w_n \Vdash q_i$, for each atom q_i.
- $h_{m+1} = \Delta_{\mathscr{D},M}(h_m)$.

Next, we need to show that $M \not\models^{\mathscr{D}}_0 (A_{bnf})^*$. We use the sequence of hypotheses $\langle h_m \rangle_{m \in \omega}$. After k revisions, $M + h_k \not\models (A_{bnf})^*$. For all natural numbers n,

$$h_k \equiv_{\mathscr{D}} h_{n \cdot (k+1)+k},$$

and $\langle B, v \rangle \in_M h$. Similarly, the extension of $G\overline{x}$ is the set of $\overline{d}$ such that for some v, $v(\overline{x}) = \overline{d}$ and $\langle A(\overline{x}, G), v \rangle \in_M h$.

[26] The table uses truth values rather than pairs from hypotheses. The correspondence is straightforward.

so, by theorem 1, for all n,

$$M + h_{n\cdot(k+1)+k} \not\models (A_{bnf})^*.$$

Therefore, $M \not\models_0^{\mathscr{D}} (A_{bnf})^*$.

COROLLARY 3.3.

1. $\not\models_0^{\mathscr{D}} (A_{bnf})^*$
2. $\not\vdash_0^{\mathscr{D}} (A_{bnf})^*$

To finish the proof, we need the following lemma.

Lemma 2 *For all $\mathscr{D}$ and all $\mathscr{D}$-interpretations *, $\vdash_0^{\mathscr{D}} (A \equiv A_{bnf})^*$.*

The desired lemma is a corollary of the following.

Lemma 3 *All instances of each of the following are theorems of $C_0^{\Box}$.*

1. $((\Box{\sim}A) \equiv ({\sim}\Box A))^*$
2. $(\Box(A \,\&\, B) \equiv (\Box A \,\&\, \Box B))^*$
3. $((\Box(A \vee B) \equiv (\Box A \vee \Box B))^*$
4. $((\Box(A \supset B)) \equiv (\Box A \supset \Box B))^*$

Proof. $C_0^{\Box}$ has all instances of each equivalence scheme, without the interpretations, as theorems, so it has every interpretation of each instance as a theorem. □

3.2 A box-free variant Next we will prove a variant of the propositional Solovay-type theorem for a special set of definitions that do not use the box. We restrict attention to the set of definitions that are finite definitions in basic revision theory. In the basic theory, a set of definitions $\mathscr{D}$ is a *finite definition* just in case, for all ground models, there is a natural number n such that for all hypotheses h, $\delta^n_{\mathscr{D},M}(h)$ has a finite period, i.e. there is a p such that $\delta^{n+p}_{\mathscr{D},M}(h) = \delta^n_{\mathscr{D},M}(h)$. For finite definitions, the revision process is over in a finite number of steps, in the sense that after a finite number of revisions, only a distinguished subset of hypotheses recur over successor stages. Let us denote the set of definitions that are finite in basic revision theory $\mathcal{FIN}$. Note that no definition in $\mathcal{FIN}$ contains the box.

We will work with extended revision theory rather than basic revision theory, although all sets of $\mathscr{D}$ will be drawn from basic revision theory, so they will not use the box. We will prove the following.

Theorem 7 *Let $\mathscr{L}$ be a language containing a binary relation symbol, '$<$'.*

$$RT \vdash A \;\Leftrightarrow\; \forall \mathscr{D} \in \mathcal{FIN}, \forall \mathscr{D}\text{-interpretations } ^*,\ \vdash_0^{\mathscr{D}} A^*$$

The proof proceeds in much the same way as the proof in the previous section, so, rather than step through the proof in detail, we will highlight the changes that need to be made.

Let $LINORD(n)$ be a sentence saying that '$<$' is a discrete linear order with a least element, 0, and there are at least n distinct objects in the ordering. If there are not names for these n objects, enrich the language with them. Let $RESET(m,n)$ be a sentence saying that there are no more than n objects satisfying H_n^m and no object outside the ordering satisfies H_n^m. Let $\mathscr{H}$ be the set of the following definitions, for each $m, n \in \omega$.

$$H_n^m x =_{Df} LINORD(n) \,\&\, RESET(m,n) \,\&\, [(\forall y(0 \leq y < x \supset H_n^m y)]$$

Table 3. *Pattern of truth values across revisions*

	h	$\Delta_{\mathscr{D},M}(h)$	$\Delta^2_{\mathscr{D},M}(h)$	$\Delta^3_{\mathscr{D},M}(h)$	$\Delta^4_{\mathscr{D},M}(h)$	$\Delta^5_{\mathscr{D},M}(h)$	...
H_30	**f**	**t**	**t**	**t**	**f**	**t**	
H_31	**f**	**f**	**t**	**t**	**f**	**f**	
H_32	**f**	**f**	**f**	**t**	**f**	**f**	

In all models in which either '<' is not interpreted as a discrete linear order with a least element or there are not at least n elements, the revision sequence for H^m_n will settle at the empty set for the extension of H^m_n. Let us suppose that M is a model in which $LINORD(n)$ is satisfied. Then there are n objects. We may as well call them 0,1,2,..., $n-1$. As an example, the revision sequence for H_3x is as follows.[27] We will assume that the initial extension of H_3 is empty. After one revision, the extension of H_3 resets either to the empty set or to an initial <-segment containing no more than 3 objects. After at most 3 revisions, it resets to the empty set, and then falls into the pattern illustrated in table 3, until it reaches 3 elements, at which point it resets to empty after the next revision. With this observation, we can prove the following.

Lemma 4 *Each finite subset $Y \subseteq \mathscr{H}$ is a finite definition (in basic revision theory).*

For any H^m_n, we can construct a table of the pattern of truth values it takes over the n elements it applies to. This has an eventual period of $n+1$. For $0 \leq k < n+1$, let $Col(H^m_n)_k$ be a sentence that is true whenever the kth column of the table of patterns of truth values matches the current stage of revision. For H^m_n, there are $n+1$ columns and $Col(H^m_n)_k$ is defined as

$$\pm_0 H^m_n(0) \,\&\, \pm_1 H^m_n(1) \,\&\, \ldots \,\&\, \pm_n H^m_n(n),$$

where $\pm_j$ is nothing if $j < k$ and $\pm_j$ is '$\sim$' otherwise. For the example above, $Col(H_3)_0$ is

$$\sim H_3(0) \,\&\, \sim H_3(1) \,\&\, \sim H_3(2),$$

and $Col(H_3)_2$ is

$$H_3(0) \,\&\, H_3(1) \,\&\, \sim H_3(2).$$

Suppose h occurs in a revision sequence after a stage at which $RESET(m,n)$ is false. It is clear from the definition that if $Col(H^m_n)_k$ is true at $M+h$, then, for $j \neq k$, $Col(H^m_n)_j$ will not be true at $M+h$. Combinations of the $Col(H^m_n)$ sentences are what will interpret the boxed atoms for the countermodel. They will be used to obtain the desired pattern of truth values.

We will assume that the modal depth of A is at least 1, otherwise we are just dealing with classical validity. If A contains m distinct atoms and A has modal depth n, then the definition $\mathscr{D}$ will be the subset of $\mathscr{H}$ containing the definitions for $H^0_n, \ldots, H^{m-1}_n$. To define * we need some more notation. Let

$$[p_i] = \{(n+1) - k : k \leq n+1 \,\&\, M, w_k \Vdash p_i\}.$$

[27] We will drop the superscript and focus on the named 3 elements.

Let

$$(p_i)^* = \bigvee_{k \in [p_i]} Col(H_n^i)_k,$$

for each atom p_i in A and for all other atoms, $(q)^* = \bot$. If $[p_i] = \emptyset$, then $\bigvee_{k \in [p'_i]} Col(H_n^i)_k = \bot$.

Let the countermodel M' contain at least n elements. As before, we construct a sequence of hypotheses, $\langle h_i \rangle_{i \in \omega}$.

- $h_0 = \emptyset$
- $h_{i+1} = \Delta_{\mathscr{D}, M'}(h_i)$

After $n+1$ revisions, the H_n^ms will cycle back to empty extensions. If the H_n^m are empty at stage k, then at stage $k+n$, $M' + h_{k+n} \models A^*_{bnf}$ iff $M, w_0 \Vdash A_{bnf}$. This is because h_{k+n} agrees with w_0 on the evaluation of all boxed atoms. For each j, there are k and r such that $h_{j+k} \equiv_{\mathscr{D}} h_{(r \cdot (n+1))+n}$, and $M' + h_{j+k} \not\models_0^{\mathscr{D}} A^*_{bnf}$. Therefore, $M' \not\models_0^{\mathscr{D}} A^*_{bnf}$, as desired.

The use of a binary predicate, '$<$', appears to be necessary to prove the theorem in basic revision theory. Now we will turn to the first-order Solovay-type theorem.

3.3 A first-order Solovay-type theorem We can obtain a first-order version of the theorem using the technique of §3.1. We must slightly alter the definition of a $\mathscr{D}$-interpretation. Instead of a propositional modal language, we start with a first-order modal language with no names or function symbols. The atomic clause in the definition of a $\mathscr{D}$-interpretation should be as follows.

- $F(x_1, \ldots, x_n)^* = B(x_1, \ldots, x_n)$, where B is a formula in the language $\mathscr{L}^+_{\mathscr{D}}$ such that $x_1, \ldots, x_n$ are all and only the free variables in B.

The other clauses of the definition of $\mathscr{D}$-interpretation remain the same.

The theorem we wish to prove is the following, where our modal language contains no names.

Theorem 8 *RTQ* $\vdash A$ *iff* $\forall \mathscr{D}, \forall \mathscr{D}$*-interpretations* $^* \models_0^{\mathscr{D}} A^*$

As in the propositional case, one can ask about the analogous proposition for definitions that do not contain the box. We will return to that case briefly after proving theorem 8.

As in the propositional case, the soundness direction, the left-to-right direction, is immediate. The converse direction will take some more work.

We note the following.

Theorem 9 (*RTQ* completeness) *RTQ is sound and complete with respect to constant domain Kripke models in which every world has exactly one successor.*

Proof. This is proved using the methods of (Hughes & Cresswell, 1996). □

To prove the right-to-left direction of theorem 8, we will argue for the contrapositive, so we will assume that A is a sentence such that $RTQ \not\vdash A$. It follows that $RTQ \not\vdash A_{bnf}$. By the completeness of RTQ, there is then a model M and a world w_0 such that $M, w_0 \not\Vdash A_{bnf}$. We assume that A, and so A_{bnf}, contains only the atomic predicates $F_1, \ldots, F_m$ and, for notational simplicity, that all the F_i have the same arity.

The model M assigns a pattern of extensions to the predicates that falsify A_{bnf} at w_0. Since M is an RT model, each world in M has unique R successor. Let $\alpha = d(A) + 1$ and let $\beta = d(A)$. We need only look at the sequence of α worlds, starting from w_0. The extensions in those worlds is shown in table 4. The predicate

Table 4. *Pattern of extensions for predicates across worlds*

	w_0	w_1	w_2	...	w_β
F_1	X_0^1	X_1^1	X_2^1		X_β^1
F_2	X_0^2	X_1^2	X_2^2		X_β^2
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
F_m	X_0^m	X_1^m	X_2^m		X_β^m

F_j is assigned the extension X_n^j in w_n, for each $n \leq \alpha$. We will use this pattern of extensions to define the refuting hypotheses.

We will modify the set $\mathscr{G}$ from the propositional case to obtain the desired set of definitions. Abusing notation slightly, let $\mathscr{G}$ be the set of circular definitions, $G_n^m(\overline{x}) =_{Df} \Box^n G_n^m(\overline{x})$, for each n and m. As before, the extensions assigned to G_n^m repeat in a sequence over $(n+1)$-many stages. Each finite subset of $\mathscr{G}$ is itself a finite definition.

As in the propositional case, it may be helpful to have an example of a cycle of extensions for a definition from $\mathscr{G}$. Take $G_3x =_{Df} \Box^3 G_3x$ and some model M. The hypotheses cycle in the pattern found in table 5, presented in terms of extensions rather than pairs from hypotheses.[28]

Table 5. *Pattern of extensions assigned by hypotheses*

	h	$\Delta_{\mathscr{D},M}(h)$	$\Delta^2_{\mathscr{D},M}(h)$	$\Delta^3_{\mathscr{D},M}(h)$	$\Delta^4_{\mathscr{D},M}(h)$	$\Delta^5_{\mathscr{D},M}(h)$	...
Gx	X_0	X_3	X_2	X_1	X_0	X_3	
$\Box Gx$	X_1	X_0	X_3	X_2	X_1	X_0	
$\Box^2 Gx$	X_2	X_1	X_0	X_3	X_2	X_1	
$\Box^3 Gx$	X_3	X_2	X_1	X_0	X_3	X_2	

For the defined predicate G_n^m, the initial hypothesis will return to its values over $sub(\mathscr{D})$ every $n+1$ revisions. Table 6 lists the sets of elements satisfying the formulas, $Gx, \Box Gx, \Box^2 Gx$, and $\Box^3 Gx$, at different stages of revision.

We now define the desired set $\mathscr{D}$ of definitions. We let $\mathscr{D}$ be the set of definitions for $G_\beta^1, \ldots,$ and G_β^m from $\mathscr{G}$, where m is the number of distinct atomic predicates in A_{bnf}. The desired $\mathscr{D}$-interpretation * is the one that assigns to each F_i the predicate G_β^i, with the appropriate variables. For the model, we take the domain of D and assign all predicates not occurring in $(A_{bnf})^*$ an empty extension. Call this model M'. Define a sequence of hypotheses $\langle h_n \rangle_{n \in \omega}$ as follows.

[28] We drop the subscript on the predicate and corresponding superscript on the extension.

Table 6. *Pattern of sets satisfied by given hypotheses*

$\models$	h	$\Delta_{\mathscr{D},M}(h)$	$\Delta^2_{\mathscr{D},M}(h)$	$\Delta^3_{\mathscr{D},M}(h)$	$\Delta^4_{\mathscr{D},M}(h)$	$\Delta^5_{\mathscr{D},M}(h)$	...
Gx	X_3	X_2	X_1	X_0	X_3	X_2	
$\Box Gx$	X_0	X_3	X_2	X_1	X_0	X_3	
$\Box^2 Gx$	X_1	X_0	X_3	X_2	X_1	X_0	
$\Box^3 Gx$	X_2	X_1	X_0	X_3	X_2	X_1	

- $\langle \Box^k G^j_\beta \overline{x}, v\rangle \in h_0$ iff $v(\overline{x}) \in X^j_k$, where X^j_k is the extension of F_j in w_k.
- $h_{n+1} = \Delta_{\mathscr{D},M'}(h_n)$.

The desired hypotheses to falsify $(A_{bnf})^*$ are all hypotheses that agree with M in the following sense. For each j, k such that $1 \leq j \leq m$ and $k \leq \beta$, X^j_k is the set of tuples satisfying $\Box^k G^j_\beta \overline{x}$ in $M' + h$ just in case M assigns X^j_k to F_j in w_k. Suppose h_n is such a hypothesis. For all p, $h_n \equiv_{\mathscr{D}} h_{p\cdot\alpha+n}$.

It remains to see that $\not\models^{\mathscr{D}}_0 (A_{bnf})^*$. For this to be true, we need a model N and for each n, a hypothesis h such that $N + \Delta^n_{\mathscr{D},N}(h) \not\models (A_{bnf})^*$. For the model we take M'. For each n, we want to pick a hypothesis that will yield one of the falsifying hypotheses after n revisions.

Now we can finally show that $M' + \Delta^n_{\mathscr{D},M'}(h) \not\models (A_{bnf})^*$. The falsifying hypothesis $\Delta^n_{\mathscr{D},M'}(h)$ agrees with the model M in the following sense: for each $k \leq \beta$,

$$M' + \Delta^n_{\mathscr{D},M'}(h), v \models \Box^k G^j_\beta(\overline{x}) \;\Leftrightarrow\; M, v, w_k \Vdash F_j(\overline{x}).$$

The latter holds just in case $M, v, w_0 \Vdash \Box^k F_j(\overline{x})$. Since $M, w_0 \not\models A_{bnf}$, we have

$$M' + \Delta^n_{\mathscr{D},M'}(h) \not\models (A_{bnf})^*.$$

To close, let us turn to the box-free variant for the first-order case. It is an open question whether that holds in general or for the restricted class of finite definitions. We do not see how to adapt our proof of the first-order theorem to a definition that lacks boxes. In the propositional case, the box-free definition that we gave cycled through truth values, in other words, extensions with respect to a single element. We do not see how to replicate these cycles with the potentially arbitrary extensions needed to invalidate the relevant sentences.

§4 Appendix: Regularity Theorem In this appendix we will prove a few facts about the concepts introduced in §2. We begin by showing that revision preserves similarity. First, we note that correspondence and similarity both preserve satisfaction. We will omit the proofs, as both are by simple inductions.

Lemma 5 *If $\langle A, v\rangle$ corresponds in M to $\langle B, u\rangle$, then*

$$M + h, v \models A \;\Leftrightarrow\; M + h, u \models B.$$

Lemma 6 *If $\langle A, v\rangle$ and $\langle B, u\rangle$ are similar, then*

$$M + h, v \models A \;\Leftrightarrow\; M + h, u \models B.$$

These lemmas are sufficient to obtain the desired corollary.

COROLLARY 4.4. *For all hypotheses h, $\Delta_{\mathscr{D},M}(h)$ is a hypothesis.*

Lemma 6 has the following corollary, which, although obvious, is heuristically suggestive.

COROLLARY 4.5. *Suppose h is a hypothesis. Let*

$$h' = \{\langle A, v\rangle \in \mathscr{F} \times \mathscr{V} : M + h, v \models A\}.$$

Then h' is a hypothesis, and $h' = \Delta_{\mathscr{D},M}(h)$.

The next lemma extends revision to all formulas of the language, not just those in $\mathscr{F}$.

Lemma 7 *For all formulas A and assignments v,*

$$\langle A, v\rangle \in_M \Delta_{\mathscr{D},M}(h) \Leftrightarrow M + h, v \models A.$$

Proof. Assume $\langle A, v\rangle \in_M \Delta_{\mathscr{D},M}(h)$. So, there is a pair $\langle B, u\rangle \in \Delta_{\mathscr{D},M}(h)$ to which $\langle A, v\rangle$ corresponds. By the definition of revision, $\langle B, u\rangle \in \Delta_{\mathscr{D},M}(h)$ iff $M + h, u \models B$. By lemma 5, this is true iff $M + h, v \models A$.

Assume $M + h, v \models A$. Suppose $\langle A, v\rangle$ corresponds to $\langle B, u\rangle$. By lemma 5, $M + h, u \models B$, so $\langle B, u\rangle \in \Delta_{\mathscr{D},M}(h)$. Therefore $\langle A, v\rangle \in_M \Delta_{\mathscr{D},M}(h)$.

□

Now we will sketch the proof of the Regularity Theorem. We begin by noting some useful lemmas.

Lemma 8 *If $h \equiv_{\mathscr{D}} h'$, then for all $A \in sub(\mathscr{D})$, then*

$$M + h, v \models A \Leftrightarrow M + h', v \models A.$$

Proof. The proof is by induction on the complexity of A.

□

Lemma 9 *If $h \equiv_{\mathscr{D}} h'$ and $A \in sub(\mathscr{D})$, then*

$$\langle A, v\rangle \in_M \Delta_{\mathscr{D},M}(h) \Leftrightarrow \langle A, v\rangle \in_M \Delta_{\mathscr{D},M}(h').$$

Proof. The proof is by induction on the complexity of A. We will present only the box case.

Case: A is $\Box B$

$$\begin{aligned}
\langle \Box B, v\rangle \in_M \Delta_{\mathscr{D},M}(h) &\Leftrightarrow \quad M + h, v \models \Box B \\
&\Leftrightarrow \quad \langle B, v\rangle \in_M h \\
&\Leftrightarrow \quad \langle B, v\rangle \in_M h', \text{ by IH} \\
&\Leftrightarrow \quad M + h', v \models \Box B \\
&\Leftrightarrow \langle \Box B, v\rangle \in_M \Delta_{\mathscr{D},M}(h')
\end{aligned}$$

□

The preceding lemmas are sufficient for the following corollary.

COROLLARY 4.6. *If $h \equiv_{\mathscr{D}} h'$, then $\Delta_{\mathscr{D},M}(h) \equiv_{\mathscr{D}} \Delta_{\mathscr{D},M}(h')$.*

For the proof of the Regularity Theorem, we need the following auxiliary concept, another kind of equivalence between hypotheses. This equivalence is, however, significantly different from $\equiv_{\mathscr{D}}$.

Definition 12 ($\equiv_n$) *Let h and h' be hypotheses. $h \equiv_n h'$ iff and for all $v \in \mathscr{V}_M$ and for all B such that $d(B) \leq n$,*

$$M + h, v \models B \Leftrightarrow M + h', v \models B.$$

We note two obvious facts about the $\equiv_n$ relations.

Lemma 10 *The $\equiv_n$ relations are equivalence relations.*
For $k \leq n$, if $h \equiv_n h'$, then $h \equiv_k h'$.

The two relations, $\equiv_n$ and $\equiv_{\mathscr{D}}$, combine to give an informative and useful relation on hypotheses, as indicated by the following lemma.

Lemma 11 *For $n \geq 0$, if $h \equiv_{\mathscr{D}} h'$ and $h \equiv_n h'$, then $\Delta_{\mathscr{D},M}(h) \equiv_{n+1} \Delta_{\mathscr{D},M}(h')$.*

Proof. Assume $h \equiv_n h'$ and $h \equiv_{\mathscr{D}} h'$. We want to show that for all formulas A such that $d(A) \leq n + 1$, for all assignments v,

$$M + \Delta_{\mathscr{D},M}(h), v \models A \Leftrightarrow M + \Delta_{\mathscr{D},M}(h'), v \models A.$$

We proceed by induction on the complexity of formulas A. We here present only the box case.

Case: $\Box B$. Since $d(\Box B) \leq n + 1$, $d(B) \leq n$.

$$M + \Delta_{\mathscr{D},M}(h), v \models \Box B \Leftrightarrow \langle B, v\rangle \in_M \Delta_{\mathscr{D},M}(h) \Leftrightarrow M + h, v \models B.$$

By the assumption that $h \equiv_n h'$, this is equivalent to $M + h', v \models B$, which by the definition of revision is equivalent to $\langle B, v\rangle \in_M \Delta_{\mathscr{D},M}(h')$. This holds just in case $M + \Delta_{\mathscr{D},M}(h'), v \models \Box B$, as desired. □

This is sufficient to establish the following.

Lemma 12 *For all n, if $h \equiv_{\mathscr{D}} h'$ and $h \equiv_0 h'$, then $\Delta^n_{\mathscr{D},M}(h) \equiv_n \Delta^n_{\mathscr{D},M}(h')$.*

Proof. The proof is by induction on n. The base case and induction step, respectively, are taken care of by the two preceding lemmas. □

Lemma 13 *If $h \equiv_{\mathscr{D}} h'$, then $h \equiv_0 h'$.*

Proof. Assume $h \equiv_{\mathscr{D}} h'$. We will show that for all A such that $d(A) = 0$, $M + h, v \models A$ iff $M + h', v \models A$. The proof is by induction on the complexity of A. The cases are all trivial except for when A is a defined predicate, in which case the case is taken care of by the assumption that $h \equiv_{\mathscr{D}} h'$.

□

The last two lemmas suffice to establish the Regularity Theorem, which we restate here.

Theorem 10 (Regularity Theorem) *Suppose $h \equiv_{\mathscr{D}} h'$. If $d(A) \leq n$, then for all $m \geq n$,*

$$M + \Delta^m_{\mathscr{D},M}(h), v \models A \Leftrightarrow M + \Delta^m_{\mathscr{D},M}(h'), v \models A.$$

§5 Appendix: Finite definitions In §2, we said that in basic revision theory, S_0 and $S^{\#}$ coincide for finite definitions. We can maintain this equivalence in extended revision theory by generalizing the notion of finite definition.

In the context of the extended theory, the original sense of finite definition will not work, because we have broadened hypotheses to cover all formulas, not just defined predicates. To see the problem, consider the following sequence.

$$\top, \Box\top, \Box^2\top, \ldots, \Box^n\top, \ldots$$

Take a hypothesis that makes $\Box^n\top$ false for each $n \geq 1$. After m revisions, $\Box^m\top$ will be true, and so be added to the next revision of the hypothesis. Since for every n, $\Delta^{n+1}_{\mathscr{D},M}(h)$ disagrees with h on at least one formula, $\Box^n\top$, there is no n for which $\Delta^n_{\mathscr{D},M}(h) = h$.

For any definition, even finite definitions in the original sense, some hypotheses may have strange evaluations of boxed formulas; these evaluations will be filtered out after revision, but no finite upper bound can be put on the number of revisions required. The problem can, however, be fixed by restricting attention to formulas in $sub(\mathscr{D})$. More precisely, in the extended theory, we will adopt the following definition of finiteness.

Definition 13 (n-reflexive, reflexive, finite)
A hypothesis h is n-reflexive *for $\mathscr{D}$ iff* $\forall B \in sub(\mathscr{D})\ \forall v \in \mathscr{V}_M$

$$\langle B, v\rangle \in_M h \Leftrightarrow \langle B, v\rangle \in_M \Delta^n_{\mathscr{D},M}(h).$$

A hypothesis h is reflexive *for $\mathscr{D}$ iff there is some $n > 0$ for which h is n-reflexive.*
A definition $\mathscr{D}$ is finite *iff $\forall M \exists n \forall h\ \Delta^n_{\mathscr{D},M}(h)$ is reflexive.*

The problematic sequence displayed above will not interfere with certain definitions being finite, because that sequence will not be in $sub(\mathscr{D})$.

If the finite definition $\mathscr{D}$ is *simple*, in the sense that it contains definitions for only finitely many predicates, then, we can maintain the equivalence between S_0 and $S^{\#}$ for $\mathscr{D}$.[29]

Theorem 11 *If $\mathscr{D}$ is a simple, finite definition, then*

$$\models^{\mathscr{D}}_0 A \Leftrightarrow \models^{\mathscr{D}}_{\#} A.$$

We omit the proof here, since it would be a lengthy addition. For the interested reader, we will note that the proof is an adaptation of the proof from (Gupta, 2006) for the analogous claim in basic revision theory. Our requirement that definitions be simple stems from the requirement that finiteness take into account all formulas in $sub(\mathscr{D})$. We found it easier to work with definitions put into box normal form, but we were unable to prove the equivalence between definitions in box normal form and those not in normal form when the definitions contain infinitely many predicates. To state the problem more precisely, for a finite definition $\mathscr{D}$, let $\mathscr{D}_{bnf}$ be the result of putting all the *definientia* in $\mathscr{D}$ into box normal form. If $\mathscr{D}$ is simple, then the set of sentences valid on $\mathscr{D}$ coincides with the set of sentences valid on $\mathscr{D}_{bnf}$. It is an open question whether the same holds when $\mathscr{D}$ is not simple.

§6 Acknowledgements This work is based on my dissertation, which written at the University of Pittsburgh. I am very grateful to Anil Gupta for many conversations about this material, including posing the initial problem of this paper. I have

[29] The following theorem was proved in the author's dissertation, (Standefer, 2013).

benefitted greatly from his comments and suggestions. I would like to thank Robert Brandom for the research support he provided me. I would also like to thank Nuel Belnap, Kohei Kishida, Rohan French, the members of the UConn Logic Group, and the two anonymous referees of this journal for their helpful feedback.

Bibliography

Antonelli, A. (1994). The complexity of revision. *Notre Dame Journal of Formal Logic* **35**(1), 67–72.

Asmus, C. M. (2013). Vagueness and revision sequences. *Synthese* **190**(6), 953–974.

Belnap, N. (1982). Gupta's rule of revision theory of truth. *Journal of Philosophical Logic* **11**(1), 103–116.

Blackburn, P., de Rijke, M., & Venema, Y. (2002). *Modal Logic*. Cambridge University Press.

Boolos, G. (1993). *The Logic of Provability*. Cambridge University Press.

Bruni, R. (2013). Analytic calculi for circular concepts by finite revision. *Studia Logica* **101**(5), 915–932.

Chapuis, A. (1996). Alternative revision theories of truth. *Journal of Philosophical Logic* **25**(4), 399–423.

Gupta, A. (1982). Truth and paradox. *Journal of Philosophical Logic* **11**(1). A revised version, with a brief postscript, is reprinted in Martin (1984).

Gupta, A. (1988–89). Remarks on definitions and the concept of truth. *Proceedings of the Aristotelian Society* **89**, 227–246. Reprinted in Gupta (2011).

Gupta, A. (2006). Finite circular definitions. In Bolander, T., Hendricks, V. F., & Andersen, S. A., editors, *Self-Reference*, pp. 79–93. CSLI Publications.

Gupta, A. (2011). *Truth, Meaning, Experience*. Oxford University Press.

Gupta, A., & Belnap, N. (1993). *The Revision Theory of Truth*. MIT Press.

Gupta, A., & Standefer, S. (2014). Conditionals in theories of truth. Manuscript.

Herzberger, H. G. (1982). Notes on naive semantics. *Journal of Philosophical Logic* **11**(1), 61–102.

Horsten, L., Leigh, G., Leitgeb, H., & Welch, P. D. (2012). Revision revisited. *Journal of Philosophical Logic* **5**(4), 642–664.

Hughes, G. E., & Cresswell, M. J. (1996). *A New Introduction to Modal Logic*. Routledge.

Kremer, P. (1993). The Gupta-Belnap systems $S^{\#}$ and S^{*} are not axiomatisable. *Notre Dame Journal of Formal Logic* **34**(4), 583–596.

Kühnberger, K.-U., Löwe, B., Möllerfeld, M., & Welch, P. (2005). Comparing inductive and circular definitions: Parameters, complexity and games. *Studia Logica* **81**(1), 79–98.

Löwe, B., & Welch, P. D. (2001). Set-theoretic absoluteness and the revision theory of truth. *Studia Logica* **68**(1), 21–41.

Martin, R. L., editor (1984). *Recent Essays on Truth and the Liar Paradox*. Oxford University Press.

Martinez, M. (2001). Some closure properties of finite definitions. *Studia Logica* **68**(1), 43–68.

Orilia, F. (2000). Property theory and the revision theory of definitions. *Journal of Symbolic Logic* **65**(1), 212–246.

Shapiro, L. (2006). The rationale behind revision-rule semantics. *Philosophical Studies* **129**(3), 477–515.

Solovay, R. M. (1976). Provability interpretations of modal logic. *Israel Journal of Mathematics* **25**(3-4), 287–304.

Standefer, S. (2013). *Truth, Semantic Closure, and Conditionals*. Ph. D. thesis, University of Pittsburgh.

Welch, P. D. (2001). On Gupta-Belnap revision theories of truth, Kripkean fixed points, and the next stable set. *Bulletin of Symbolic Logic* **7**(3), 345–360.

Yaqūb, A. M. (1993). *The Liar Speaks the Truth: A Defense of the Revision Theory of Truth*. Oxford University Press.

DEPARTMENT OF PHILOSOPHY
UNIVERSITY OF PITTSBURGH
USA
E-mail: standefer@gmail.com