The Logic of Provability

We want to apply the methods of modal logic to get a better picture of provability. **Until further notice, Γ will be a recursively axiomatized arithmetical theory that includes PA and that doesn't imply any false Σ sentences. "Provability" will mean provability in Γ, and "Bew" will abbreviate "Bew_Γ."** If we understand "$\Box\,\phi$" to mean "Bew($\lceil\phi\rceil$)," things like Löb's Theorem and the Löb conditions (L1) to (L3) will convert straightforwardly to principles of modal logic. The modal system we get won't be among the most common systems – the most common systems all include KT[1] – but the methods of modal logic can be fruitfully applied to it nonetheless.

An *arithmetical interpretation* of our language for the modal sentential calculus is a function *i* that associates an arithmetical formula with each modal formula, subject to the following constraints:

$$i\,(\phi \vee \psi) = (i(\phi) \vee i(\psi))$$

$$i\,(\phi \wedge \psi) = (i(\phi) \wedge i(\psi))$$

$$i\,(\phi \rightarrow \psi) = (i(\phi) \rightarrow i(\psi))$$

$$i\,(\phi \leftrightarrow \psi) = (i(\phi) \leftrightarrow i(\psi))$$

$$i(\sim \phi) = \sim i(\phi)$$

$$i(\Box\,\phi) = \text{Bew}(\lceil i(\phi)\rceil)$$

A modal formula ϕ is *always provable* iff, for each arithmetical interpretation i, $i(\phi)$ is provable. ϕ is *always true* iff, for each arithmetical interpretation i, $i(\phi)$ is true.

---

[1] The prominent exceptions are systems of deontic logic, in which "$\Box\phi$" is read "ϕ is morally obligatory," and "$\Diamond\phi$" is read "ϕ is morally permissible. We don't live in a morally perfect world, so not everything that is true is morally permissible.

Löb's condition (L1) tells us that the set of always-provable formulas is closed under Necessitation. It follows from this that the instances of schema (4) are always true. (L2) tells us that they are, in fact, always provable. (L3) tells us that the instances of schema (K) are always provable. Since the set of always-provable sentences is closed under (TC), we conclude that the set of always-provable sentences is a normal modal system that includes K4.

Löb's Theorem tells us that all instances of the following schema are always true:

(L) $(\Box(\Box\phi \rightarrow \phi) \rightarrow \Box\phi)$

The proof of Löb's Theorem can be formalized in Γ, with the consequence that the instances of schema (L) are always provable. If we let GL (for "Gödel-Löb") be the smallest normal modal system that includes both (4) and (L), we see that the set of always-provable sentences includes GL. Dick de Jongh has shown that including schema (4) is redundant, so that GL can alternatively be characterized as the smallest normal modal system that includes (L).

The main theorem in provability logic, which was obtained by Robert Solovay,[2] gives an exact characterization of the set of always-provable formulas: The class of always-provable formulas is GL.

Before undertaking to prove Solovay's theorem, we need a better characterization of GL. Let's say a triple <W,R,a> (where a is an element of W and R is a binary relation on W) is a finite tree if it meets the following conditions:

Finitude: W is finite.

Transitivity: Whenever Ruv and Rvw, we have Ruw.

---

[2] "Provability Interpretations of Modal Logic," *Israel Journal of Mathematics* 25 (1976): 287-304. The definitive exposition of provability logic is George Boolos, *The Logic of Provability* (Cambridge: Cambridge University Press, 1995).

Anti-reflection: We never have Rww.

a is the trunk: If w ∈ W, then either a = w or Raw.

Branch-connection: If Ruw and Rvw, then either Ruv or u=v or Rvw.

The paradigm case of a finite true is a nonempty, finite set of finite sequences, with the property that every initial segment of a member of the set is a member of the set. Ruv holds if and only if v extends u.

Given an interpretation <W,R,I,a>, with <W,R,a> a finite true, we know from the fact that R is transitive that the set of formulas true in every world in the model is a normal modal system that includes (4). We want to see that it also includes (L). Let w ∈ W, and suppose that $\Box\phi$ is false in w. Then there is a world accessible from w in which ϕ is false, and hence, because W is finite, there has to be a bottommost[3] world v accessible from w in which ϕ is false. $\Box\phi$ is true in v, and so $(\Box\phi \rightarrow \phi)$ is false in v, and $\Box(\Box\phi \rightarrow \phi)$ is false in w. Hence $(\Box(\Box\phi \rightarrow \phi) \rightarrow \Box\phi)$ is true in every world. We have thus proved the right-to-left direction of the following:

Theorem. A sentence is true in every model <W,R,I,a>, with <W,R,a> a finite tree, if and only if it is an element of GL.

Proof: Suppose that χ isn't in GL. We want to construct a model <W,R,I,a>, with <W,R,a> a finite tree, in which χ is false. The construction we've used in the past, with maximal consistent sets of sentences, won't give us a finite tree. To keep everything finite, we don't look at all the sentences, but only at the sentences that are either subsentences of χ or negations of

[3] Following the custom of mathematicians, very few of whom were raised on the farm, I speak of trees as growing downward, with the trunk at the top and the leaves at the bottom.

subsentences of χ.[4] Since χ isn't an element of GL, we can find a set of sentences a* with the following properties:

$\sim \chi$ is an element of a*.

a* is GL-consistent.

Every member of a* is either a subsentence of χ or the negation of a subsentence of χ.

For each subsentence of χ, either it or its negation is in a*.

To form a*, we go through the subsentences of χ. When we come to a sentence, we add either it or its negation to our set.

Let W* be the set of all maximal GL-consistent sets of subsentences of χ and negated subsentences of χ. That is, a set of sentences is in W* iff it meets the last three of the four conditions above. If w* is an element of W* and ϕ is an atomic sentence that occurs in χ, we'll set $I^*(\phi,w^*) = 1$ iff $\phi \in w^*$. The tricky part is defining the accessibility relation R*. Here's the definition: R* w* v* iff the following two conditions are met:

For any sentence $\Box\phi$ that's an element of w*, both $\Box\phi$ and ϕ are elements of v*.

There is a sentence θ such that $\Box\theta$ is in v*, but $\Box\theta$ isn't in w*.

The proof that, for any sentence ψ that's a subsentence of χ, ψ is true in w* in the model $\langle W^*,R^*,I^*,a^*\rangle$ iff ψ is an element of w* is routine, except for one part. We need to show that, if

---

[4] The construction we give here, where we hire a finite model to do the job of the canonical model, comes up routinely for modal logicians in giving proofs that modal systems are decidable. One shows, for example, that KT4 is decidable by showing that, is a sentence is not in KT4 then one can construct a finite, reflexive, transitive model in which it's false. An infinite model, which is what the canonical frame provides, does us no good. Our completeness proof for GL has a rabbit-out-of-the-hat quality only because we're presenting it in isolation from its native enviroment in the theory of modal logics.

$\Box\psi$ is a subsentence of χ that isn't in w*, then there is a v* with R* w* v* that doesn't contain χ. To do this, we need to show that $\{\sim\psi, \Box\psi\} \cup \{\phi: \Box\phi \in w^*\} \cup \{\Box\phi: \Box\phi \in w^*\}$ is GL-consistent. If we do this, we can take v* to be a member of W* that contains this set. If the set is GL-inconsistent, we can find $\phi_1, \phi_2,..., \phi_n$, with each $\Box\phi_i$ in w* such that the following sentence is in GL:

$$((\Box\phi_1 \wedge \phi_1) \rightarrow ((\Box\phi_2 \wedge\phi_2) \rightarrow...\rightarrow ((\Box\phi_n \wedge \phi_n) \rightarrow (\Box\psi \rightarrow \psi))...)).$$

Because GL is normal, the following sentence is in GL:

$$(\Box(\Box\phi_1 \wedge \phi_1) \rightarrow (\Box(\Box\phi_2 \wedge\phi_2) \rightarrow...\rightarrow (\Box(\Box\phi_n \wedge \phi_n) \rightarrow \Box(\Box\psi \rightarrow \psi))...)).$$

Because GL includes K4, $(\Box\phi_i \rightarrow \Box(\Box\phi_i \wedge \phi_i))$ is in GL, for each i, and also, because GL includes (L), $(\Box(\Box\psi \rightarrow \psi) \rightarrow \Box\psi)$ is in GL. Consequently, the following sentence is in GL:

$$(\Box\phi_1 \rightarrow (\Box\phi_2 \rightarrow...\rightarrow (\Box\phi_n \rightarrow \Box\psi)...)).$$

Because each of the $\Box\phi_i$s is in w*, $\Box\psi$ is in w*. Contradiction.

We're still not done. <W*,R*,a*> will be finite, transitive, and antireflexive, but there's no reason to suppose that it's branch connected or that every member of W* other than a* is accessible from a*. We have to tinker with the model to make it a tree. We're going to let our "worlds" be finite R*-chains that begin with a*. More precisely, the members of W are nonempty finite sequences w of elements of W* that meet these conditions:

$(w)_0 = a^*$.

If j+1 < the length of w, then R* $(w)_j$ $(w)_{j+1}$.

If w and v are in W, Rwv iff v is an extension of w. a is the sequence whose only element is a*. If j+1 is the length of w, $I(\phi,w) = I^*(\phi, (w)_j)$. Then <W,R,a> is a tree, and if w is an

element of W of length j+1 and ϕ is a modal formula, ϕ is true in w in the model <W,R,I,a> iff ϕ is true in $(w)_j$ in the model <W*,R*,I*,a*>. So χ is false in the finite tree model <W,R,I,a>.⊠

This theorem gives us a decision procedure for GL. If a sentence is in GL, we can derive it, whereas if a sentence is outside GL we can construct a finite tree model in which it's false.

Now we're ready for the big time. Given a sentence χ that's not in GL, we want to find an arithmetical interpretation i such that $i(\chi)$ isn't a consequence of Γ. We can find a finite tree model <W,R,I,a> in which χ is false. It will do no harm if we take W to consist of the numbers 1, 2,..., n, so arranged that $i < j$ whenever Rij. Thus a = 1. We expand the model by adding 0 as an extra world, stipulating that every other world is accessible from 0 and that $I(\phi,0) = I(\phi,1)$, for ϕ atomic. At the end of the day, when we get our arithmetical interpretation, world 0 will play the role of the actual world, that is, the standard model. The sentences true in world 1 might or might not be true in the standard model; we don't want to presume. When we turn to the logic of almost-truth, world 0 will play a starring role.

Our plan is looking for an arithmetical interpretation that reproduces the structure of the tree is reminiscent of the strategy we used in seeing how to find an SC sentence with a given truth table. What we did there was to find, for each line of a truth table, a sentence, the state description, that described that line, then to take our sentence to be the disjunction of the state descriptions of the lines at which the given truth table assigns the value "true." Pursuing the same plan here, we want to find, for each world j, a sentence σ_j that describes that world. Once we've done that, we can take our arithmetical interpretation to be the function that assigns to each atomic formula the disjunction of the world-descriptions of the worlds in which the formula is true. Specifically, we find, for each $j \leq n$, a sentence σ_j meeting these conditions:

(i) PA implies the disjunction of the σ_js.

(ii) PA $\vdash \sim (\sigma_j \wedge \sigma_k)$, for $j \neq k$.

(iii) PA $\vdash (\sigma_j \rightarrow \sim \text{Bew}([\ulcorner \sim \sigma_k \urcorner]))$, whenever Rjk.

(iv) PA $\vdash (\sigma_j \rightarrow \text{Bew}([\ulcorner$the disjunction of the σ_ks with Rjk$\urcorner]))$, for $1 \leq j \leq n$.[5]

(v) σ_0 is true.

Defining our arithmetical interpretation i by stipulating that, for ϕ atomic, $i(\phi)$ is the disjunction of the σ_js, for j a world in which ϕ is true, gives us the following:

Claim. For any j, $1 \leq j \leq n$, and any modal formula ϕ, if ϕ is true in j, then PA $\vdash (\sigma_j \rightarrow i(\phi))$.

Proof: We prove by induction on the complexity of formulas that, for each formula ϕ, if ϕ is true in j, then PA $\vdash (\sigma_j \rightarrow i(\phi))$, whereas is ϕ is false in j, PA $\vdash (\sigma_j \rightarrow \sim i(\phi))$. If ϕ is atomic, then if ϕ is true in j, σ_j is one of the disjuncts of $i(\phi)$, whereas, if ϕ is false in j, condition (ii) assures us that σ_j is provably incompatible with each of the disjuncts of σ_j. In case ϕ is built up from simpler formulas by means of the SC connectives, the proof is easy and I won't go through it here. Here let's worry instead about showing that the claim holds when ϕ has the form $\Box\psi$.

Let's say the worlds accessible from j are $k_1, k_2, ..., k_m$. If $\Box\psi$ is true in j, then by inductive hypothesis, for each h, $1 \leq h \leq m$, PA $\vdash (\sigma_{k_h} \rightarrow i(\psi))$. So PA $\vdash ((\sigma_{k_1} \vee \sigma_{k_2} \vee ... \vee \sigma_{k_m}) \rightarrow i(\psi))$.

---

[5] In case there aren't any worlds accessible from j, let me stipulate that I'll take the "disjunction" of the σ_js with Rjk to be the logically inconsistent sentence "$\sim 0 = 0$." So (iv) tells us that, if there aren't any world accessible form j, PA $\vdash (\sigma_j \rightarrow \sim \text{Con}(\Gamma))$.

By (L1) and (L3), PA $\vdash$(Bew([$\ulcorner(\sigma_{k_1} \vee \sigma_{k_2} \vee ... \vee \sigma_{k_m})\urcorner$]) $\rightarrow$ Bew([i(ψ)]). Since, by (iv),[6] PA $\vdash$(σj $\rightarrow$ Bew([$\ulcorner(\sigma_{k_1} \vee \sigma_{k_2} \vee ... \vee \sigma_{k_m})\urcorner$])), PA $\vdash$($\sigma_j \rightarrow$ i($\Box\psi$)).

If, on the other hand, $\Box\psi$ is false in j, then there is a world k accessible from j in which ψ is false. By inductive hypothesis, PA $\vdash$($\sigma_k \rightarrow \sim$ i(ψ)). It follows by (L1) and (L3) that $\Gamma \vdash$ (Bew([$\ulcorner$i(ψ)$\urcorner$]) $\rightarrow$ Bew([$\ulcorner\sim\sigma_k\urcorner$])), and so PA $\vdash$($\sim$ Bew([$\ulcorner\sim\sigma_k\urcorner$]) $\rightarrow \sim$i($\Box\psi$)). It follows by (iii) that PA $\vdash$($\sigma_j \rightarrow \sim$ i($\Box\psi$)).$\boxtimes$

Given the Claim, we know that PA $\vdash$($\sigma_1 \rightarrow \sim$i($\chi$)). It follows by (L1) and (L3) that PA $\vdash$ (Bew([$\ulcorner$i(χ)$\urcorner$]) $\rightarrow$ Bew([$\ulcorner\sim\sigma_1\urcorner$])), and so, by (iii), PA $\vdash$($\sigma_0 \rightarrow \sim$ Bew([$\ulcorner$i(χ)$\urcorner$])). Since, by (v), σ_0 is true, it follows that Bew([$\ulcorner$i(χ)$\urcorner$]) is false, so that i(χ) isn't a consequence of Γ.

It remains to find the σ_js. Figuring out what formulas to write down took a lot of ingenuity of Solovay's part, and I won't attempt to motivate the construction. I'll just write the formulas down and verify that they work. Define a formula f(x,y) as follows:

If z isn't the Gödel number of a formula whose only free variable is "x," f(y,z) = 0. Suppose that z is the Gödel number of a formula ψ(x) with "x" as its only free variable We define f(y,z) by induction on x:

f(0,z) = 0

If f(m,z) = j and m is a proof in Γ of ψ([k]) and Rjk, then f(m+1,z) = k

Otherwise, f(m+1,z) = f(m,z).

If z = $\ulcorner\psi$(x)$\urcorner$, then in calculating the value of f(y,z) for different values of y, we start at f(0,z) = 0 and make our way down the tree. If, at a certain point, we're at node j and we find a proof of

---

[6]This is where the proof gets stuck for j = 0, since (iv) only applies where $1 \leq j \leq n$. When we turn to the logic of always true formulas, we'll develop a restricted version of the Claim that applies to world 0.

⌜ψ[k]⌝, with Rjk, then we jump to node k. Because the tree in finite, the jumping will have to come to a halt eventually.

f is a recursive total function. So we can find encode the recursive definition of f as a Σ explicit definition, and having done so, we can prove the basic features of f in PA. For example, applying our skills at supplying numerical codes for finite sets, we can prove that, for each z, the function f(y,z), regarded as a function of y, is a nondecreasing total function whose range is a subset of {0,1,2,..., n}. The Self-Reference Lemma lets us find a formula σ(x) such that

PA $\vdash$ (∀z)(σ(z) ↔ the greatest element of {f(y,[⌜~σ(x)⌝]): y ∈ ℕ} is equal to z).

PA proves that, for each z, the greatest element of {f(y,z): y ∈ ℕ} is a number between 0 and n, inclusive. In particular, it proves that th greatest element of {f(y, ⌜~σ(x)⌝): y ∈ ℕ} is between 0 and n, inclusive. This gives us (i).

If j ≠ k, PA proves that j and k aren't both equal to the greatest element of {f(y,⌜~σ(x)⌝): y ∈ ℕ}; this gives us (ii).

Take a formula ψ(x) with "x" as its only free variable. If the greatest element of {f(y,⌜ψ(x)⌝: y ∈ ℕ} is equal to j, then there is an m such that If f(m,⌜ψ(x)⌝ is equal to j. If some ψ([k]) with Rjk were provable in Γ, then there would be a number p > m that proved ψ([k]). (Note that if a sentence is provable at all, then it has infinitely many proofs, because we can take a given proof and pad it out by adding pointless digressions.) So there must be a least p > m that proves a sentence ψ([k]), with Rjk. But then f(p+1,⌜ψ(x)⌝) would be equal to k, contrary to hypothesis that j is the largest element of {f(y,⌜ψ(x)⌝: y ∈ ℕ}. So if j is the greatest element of {f(y,⌜ψ(x)⌝): y ∈ ℕ}, then no ψ([k]) with Rjk is provable in Γ. Formalizing this argument in PA gives us:

PA $\vdash$the greatest element of $\{f(y,[\ulcorner\psi(x)\urcorner]): y \in n\} = [j] \rightarrow \sim Bew([\ulcorner\psi([k])\urcorner]))$,

whenever Rjk. Putting in $\sim\sigma(x)$ in place of $\psi(x)$ gives us (iii).

Again, take $\psi(x)$ to be a formula whose only free variable is "x."

(£) PA $\vdash [j]$ is the greatest element of $\{f(y,[\ulcorner\psi(x)\urcorner]): y \in \mathbb{N}\} \rightarrow (\exists y)f(y,[\ulcorner\psi(x)\urcorner]) = [j])$.

All sentences of the form $(\theta \rightarrow Bew([\ulcorner\theta\urcorner]))$, with $\theta\ \Sigma$, are provable in PA. In particular,

(¥) PA $\vdash((\exists y)f(y,[\ulcorner\psi(x)\urcorner]) = [j] \rightarrow Bew([\ulcorner(\exists y)f(y,[\ulcorner\psi(x)\urcorner]) = [j]\urcorner]))$.

Moreover,

PA $\vdash((\exists y)f(y,[\ulcorner\psi(x)\urcorner]) = [j] \rightarrow$ the greatest member of $\{f(y,[\ulcorner\psi(x)\urcorner]): y \in \mathbb{N}\}$

is equal either to $[j]$ or to $[k_1]$ or to $[k_2]$ or to ... or to $[k_m])$,

where $k_1, k_2,..., k_m$ are the worlds accessible form j. Applying (L1) and (L3), we get:

(€) PA $\vdash(Bew([\ulcorner(\exists y)f(y,[\ulcorner\psi(x)\urcorner]) = [j]\urcorner]) \rightarrow Bew([\ulcorner$the greatest member of $\{f(y,[\ulcorner\psi(x)\urcorner]):$

$y \in \mathbb{N}\}$is equal either to $[j]$ or to $[k_1]$ or to $[k_2]$ or to ... or to $[k_m])\urcorner]))$.

Putting (£), (¥), and (€) together, we get:

PA $\vdash[j]$ is the greatest element of $\{f(y,[\ulcorner\psi(x)\urcorner]): y \in \mathbb{N} \rightarrow$

$Bew([\ulcorner$the greatest member of $\{f(y,[\ulcorner\psi(x)\urcorner]):y \in \mathbb{N}\}$is equal either to $[j]$

or to $[k_1]$ or to $[k_2]$ or to ... or to $[k_m])\urcorner]))$.

Putting $\sim\sigma(x)$ in place of $\psi(x)$, we get:

($) PA $\vdash(\sigma([j]) \rightarrow Bew([\ulcorner(\sigma([j]) \vee \sigma([k_1]) \vee \sigma([k_2]) \vee ... \vee \sigma([k_m]))\urcorner])))$.

Where $1 \leq j \leq n$, we have:

PA $\vdash(\sigma([j]) \rightarrow [j]$ is the greatest element of $\{f(y,[\ulcorner\sim\sigma(x)\urcorner]: y \in \mathbb{N}\})$

PA $\vdash([j]$ is the greatest element of $\{f(y,[\ulcorner\sim\sigma(x)\urcorner]): y \in \mathbb{N}\} \rightarrow (\exists y)f(y,[\ulcorner\sim\sigma(x)\urcorner] = [j])$.

$$\text{PA} \vdash ((\exists y)f(y,[\ulcorner\sim\sigma(x)\urcorner) = [j] \rightarrow \text{Bew}([\ulcorner\sim\sigma([j])\urcorner])$$

(¢) $$\text{PA} \vdash (\sigma([j]) \rightarrow \text{Bew}([\ulcorner\sim\sigma([j])\urcorner]).$$

Putting ($) and (¢) together, using the fact that the set of provable sentences is closed under the form of inference:

$$(\alpha \vee \beta)$$

$$\sim\alpha$$

$$\therefore \beta$$

we get:

$$\text{PA} \vdash (\sigma([j]) \rightarrow \text{Bew}([\ulcorner(\sigma([k_1]) \vee \sigma([k_2]) \vee ... \vee \sigma([k_m]))\urcorner])).$$

which is (iv).

Finally, we want to prove (v),[7] that is, we want to show that the greatest element of $\{f(y,\ulcorner\sim\sigma(x)\urcorner): y \in \mathbb{N}\}$ is 0. Suppose, on the contrary, that the greatest element is $j > 0$, and let the worlds accessible from j be $k_1, k_2,..., k_m$. We have, from (iv):

$$\text{PA} \vdash \sigma([j]) \rightarrow \text{Bew}([\ulcorner(\sigma([k_1]) \vee \sigma([k_2] \vee ... \vee \sigma([k_m]))\urcorner]).$$

For each i, we have

$$\text{PA} \vdash (\sigma([k_i]) \rightarrow (\exists y)f(y,[\ulcorner\sim\sigma(x)\urcorner] = [k_i])$$

Therefore,

$$\text{PA} \vdash ((\sigma([k_1]) \vee \sigma([k_2]) \vee ... \vee \sigma([k_m])) \rightarrow ((\exists y)f(y,[\ulcorner\sigma(x)\urcorner]) = [k_1] \vee$$

$$(\exists y)f(y,[\ulcorner\sigma(x)\urcorner]) = [k_2] \vee ... \vee (\exists y)f(y,[\ulcorner\sigma(x)\urcorner]) = [k_m])).$$

---

[7] If we were assuming that Γ were true, instead of merely that it's Σ consequences are all true, (v) would be a piece of cake. For $j > 0$, $\sigma([j])$ asserts its own refutability, so that, if it were true, it would be a true refutable sentences. However, we are not assuming that Γ is true, so there may well be true sentences that are refutable in Γ. So we have more work to do.

Applying (L1) and (L3) yields:

$$\text{PA} \vdash (\text{Bew}([\ulcorner(\sigma([k_1]) \vee \sigma([k_2]) \vee \ldots \vee \sigma([k_m]))\urcorner]) \rightarrow \text{Bew}([\ulcorner((\exists y)f(y,[\ulcorner\sigma(x)\urcorner]) = [k_1] \vee (\exists y)f(y,[\ulcorner\sigma(x)\urcorner]) = [k_2] \vee \ldots \vee (\exists y)f(y,[\ulcorner\sigma(x)\urcorner]) = [k_m])\urcorner])),$$

and so

$$\text{PA} \vdash \sigma([j]) \rightarrow \text{Bew}([\ulcorner((\exists y)f(y,[\ulcorner\sigma(x)\urcorner]) = [k_1] \vee (\exists y)f(y,[\ulcorner\sigma(x)\urcorner]) = [k_2] \vee \ldots \vee (\exists y)f(y,[\ulcorner\sigma(x)\urcorner]) = [k_m])\urcorner]))$$

On the other hand, if j is the greatest element of $\{f(y,\ulcorner\sim\sigma(x)\urcorner): y \in \mathbb{N}\}$, then for none of the k_is does there exist a y with $f(y,\ulcorner\sim\sigma(x)\urcorner) = k_i$, for each of the k_is is greater than j. This observation can be formalized in PA, yielding:

$$\text{PA} \vdash (\sigma([j]) \rightarrow \sim((\exists y)f(y,[\ulcorner\sigma(x)\urcorner]) = [k_1] \vee (\exists y)f(y,[\ulcorner\sigma(x)\urcorner]) = [k_2] \vee \ldots \vee (\exists y)f(y,[\ulcorner\sigma(x)\urcorner]) = [k_m]))$$

Consequently, since $\sigma([j])$ is true, the disjunction $((\exists y)f(y,[\ulcorner\sigma(x)\urcorner]) = [k_1] \vee (\exists y)f(y,[\ulcorner\sigma(x)\urcorner]) = [k_2] \vee \ldots \vee (\exists y)f(y,[\ulcorner\sigma(x)\urcorner]) = [k_m]))$ is logically equivalent to a false Σ sentence provable in Γ, contrary to hypothesis.⊠

We now turn out attention to problem of determining which modal formulas are always true. Assuming that Γ is true, every always-provable formula will be always true, but not every always-true formula will be always provable, for all the instances of schema (T) will be always-true, but only those with always-provable consequents will be always provable. It turns out that these two observations, together with the recognition that the always-true formulas are closed under *modus ponens*, is enough to give us a complete inventory of always-true formula.

Further notice: From now on Γ will the a *true* recursively axiomatized theory that includes PA.

Let GLS (for "Gödel-Löb-Solovay") be the smallest collection of formulas that includes GL and all the instances of schema (T) and is closed under *modus ponens*. Since all the tautologies are in GL, we know that GLS is closed under (TC).

Theorem (Solovay). Given a modal formula χ, let the subformulas of χ that begin with "$\Box$" be $\Box\eta_1$, $\Box\eta_2$,..., $\Box\eta_m$. The following are equivalent:

① $\chi \in$ GLS.

② $(((\Box\eta_1 \rightarrow \eta_1) \wedge (\Box\eta_2 \rightarrow \eta_2) \wedge ... \wedge (\Box\eta_m \rightarrow \eta_m)) \rightarrow \chi) \in$ GL.

③ χ is always true.

Proof: That ② implies ① and that ① implies ③ are obvious, so all we need to show ③ implies ②. Actually, we'll show that the negation of ② implies the negation of ③. If the conditional $(((\Box\eta_1 \rightarrow \eta_1) \wedge (\Box\eta_2 \rightarrow \eta_2) \wedge ... \wedge (\Box\eta_m \rightarrow \eta_m)) \rightarrow \chi)$ isn't in GL, we follow the same procedure as before to find a model $<\{0,1,...,n\},R,I,0>$ in which the conditional is false at world 1. We want to show that a subformula of χ is true in world 0 if and only if it's true at world 1. For atomic formulas, this follows immediately from the way we, thinking ahead, stipulated truth values when extending the model to include world 0. For conjunctions, disjunctions, conditionals, biconditionals, and negations, the proof is easy. If $\Box\eta_j$ is true in world 0, then η_j is true in every world accessible from 0. Since every world accessible from world 1 is accessible from world 0, it follows that η_j is true in every world accessible from world 1, and so $\Box\eta_j$ is true in world 1. If, on the other hand, $\Box\eta_j$ is true in world 1, then η_j is true in every world accessible from world 1. The only world accessible from 0 that isn't accessible from 1 is 1 itself. Since $(\Box\eta_j \rightarrow \eta_j)$ is true in 1, η_j is true in true in 1, and thus true in every world accessible from 0, so that $\Box\eta j$ is true in 0.

In particular, since χ is false in 1, χ is false in 0.

We now want to show that, for each subsentence θ of χ, if θ is true in 0, PA $\vdash(\sigma_0 \rightarrow i(\theta))$, whereas if θ is false in 0, PA $\vdash(\sigma_0 \rightarrow \sim i(\theta))$. Since σ_0 is true, it will follow that $i(\chi)$ is false, as required.

The proof for θ atomic is the same as the proof we gave earlier for worlds 1, 2,..., n. The proof for θ a disjunction, conjunction, conditional, biconditional, or negation is routine.

Suppose that $\Box\eta_j$ is true in 0. For each $k > 0$, k is accessible from 0, and so η_j is true in k. We showed earlier that this shows that PA $\vdash(\sigma_k \rightarrow i(\eta_j))$. Since η_j is true in 1 and the same subsentences of χ are true in 0 and in 1, η_j is true in 0, and so, by inductive hypothesis, PA $\vdash(\sigma_0 \rightarrow i(\eta_j))$. It follows that PA $\vdash((\sigma_0 \vee \sigma_1 \vee ... \vee \sigma_n) \rightarrow i(\eta_j))$. Since PA $\vdash(\sigma_0 \vee \sigma_1 \vee ... \vee \sigma_n)$, we have PA $\vdash i(\eta_j)$. By (L1), PA $\vdash$ Bew($\lceil i(\eta_j) \rceil$), that is, PA $\vdash i(\Box\eta_j)$, and so PA $\vdash(\sigma_0 \rightarrow i(\Box\eta_j))$.

Now suppose instead that $\Box\eta_j$ is false in 0. Then there is a world $k > 0$ in which η_j is false. We showed earlier that this implies that PA $\vdash(\sigma_k \rightarrow \sim i(\eta_j))$. Applying contraposition, (L1), (L3), and contraposition again, we obtain PA $\vdash(\sim Bew(\lceil \sim\sigma_k \rceil) \rightarrow \sim i(\Box\eta_j))$. Since (iii) gives us PA $\vdash(\sigma_0 \rightarrow \sim Bew(\lceil \sim\sigma_k \rceil))$, PA $\vdash(\sigma_0 \rightarrow \sim i(\Box\eta_j))$ follows.⊠

From the equivalence of ② and ③ and the existence of a decision procedure for GL, we see that there is an algorithm of testing whether a modal formula is always true.