L'interprétation prouvabiliste de la logique modale: $\Box = \mathsf{Th}(\dots)$ *

Marcel Crabbé

1982

Le langage modal propositionnel

Alphabet :
— une liste dénombrable de symboles propositionnels ;
— les symboles logiques : $\perp$, $\rightarrow$ et $\Box$;
— les parenthèses.

Syntaxe :
— les énoncés modaux sont les éléments du plus petit ensemble qui contient les symboles propositionnels, le symbole $\perp$ et les expressions $\Box\varphi$ et $(\varphi \rightarrow \psi)$, lorsque φ et ψ sont dans l'ensemble.

Les modèles modaux

Un modèle modal est un triple $\langle I, R, M \rangle$ où

— I est un ensemble non vide, appelé l'ensemble des indices ;
— R est une relation sur I ;
— M est une fonction à deux arguments qui fait correspondre à chaque symbole propositionnel et à chaque indice une valeur de vérité (V ou F).

Soit $\mathcal{M} = \langle I, R, M \rangle$, un modèle modal, φ un énoncé modal et i un indice ($i \in I$). On définit $\mathcal{M} \models_i \varphi$ (φ est vrai dans $\mathcal{M}$ en i) par induction sur la complexité de φ, de la façon qui suit :

* Ceci est un commentaire de l'article de Solovay *Provability interpretations of modal logic,* Israël Journal of Mathematics, **25** (1976), pp. 287–304, présenté au séminaire interuniversitaire de logique.

— $\mathcal{M} \not\models_i \bot$;
— $\mathcal{M} \models_i \varphi$ ssi $M(\varphi, i) = V$, au cas où φ est un symbole propositionnel ;
— $\mathcal{M} \models_i (\varphi \to \psi)$ ssi $\mathcal{M} \not\models_i \varphi$ et/ou $\mathcal{M} \models_i \psi$;
— $\mathcal{M} \models_i \Box\, \varphi$ ssi $\mathcal{M} \models_j \varphi$, pour tout j tel que $i\ R\ j$.

Justifications

I représente l'ensemble des mondes possibles (ou des situations concevables).

R est la relation d'accessibilité : $i\ R\ j$ signifie que j est une situation concevable (ou possible) à partir du monde i.

$\Box\, \varphi$ signifie que φ est nécessaire, et φ est nécessaire veut dire que φ est vrai dans tous les mondes possibles.

Abréviations

On écrit $\neg\varphi$ pour $(\varphi \to \bot)$, $(\varphi \wedge \psi)$ pour $\neg(\varphi \to \neg\psi)$, $(\varphi \vee \psi)$ pour $(\neg\varphi \to \psi)$, $(\varphi \leftrightarrow \psi)$ pour $((\varphi \to \psi) \wedge (\psi \to \varphi))$ et $\Diamond\, \varphi$ pour $\neg\Box\neg\varphi$.

Remarques

On constate que :
— $\mathcal{M} \not\models_i \Box\, \varphi$ ssi il existe un $j \in I$ tel que $i\ R\ j$ et $\mathcal{M} \not\models_j \varphi$;
— $\mathcal{M} \models_i \Diamond\, \varphi$ ssi il existe un $j \in I$ tel que $i\ R\ j$ et $\mathcal{M} \models_j \varphi$;

$\Diamond\, \varphi$ signifie donc que φ est possible.

Th(...) **dans** PA

PA est l'arithmétique de Peano du premier ordre, formulée avec les variables v, v', v'', ... et les symboles 0, S, $+$ et $\cdot$

Si $n \in \mathbb{N}$, $\bar{n}$ est le numéral correspondant, à savoir $\underbrace{S \ldots S}_{n \text{ fois}} 0$.

À chaque expression (terme, formule ou suite finie de formules), e de PA, on fait correspondre un nombre de Gödel $\ulcorner e \urcorner$ et donc un numéral $\overline{\ulcorner e \urcorner}$, que l'on notera "$e$".

Chaque relation primitive récursive a une représentation *canonique* dans PA. Notamment, il y a une formule de PA, notée $\mathsf{Dem}(v', v)$, qui exprime la notion de démonstration. Si Σ est une démonstration de φ, alors $\mathsf{PA} \vdash \mathsf{Dem}(\text{“}\Sigma\text{”}, \text{“}\varphi\text{”})$. $\mathsf{Th}(v)$ est la formule $\exists v' \mathsf{Dem}(v', v)$. $\mathsf{Th}(v)$ dit dans PA que v est un théorème de PA.

On sait que :

1. si $\mathsf{PA} \vdash \varphi$, alors $\mathsf{PA} \vdash \mathsf{Th}(\text{“}\varphi\text{”})$;
2. $\mathsf{PA} \vdash \mathsf{Th}(\text{“}(\varphi \to \psi)\text{”}) \to (\mathsf{Th}(\text{“}\varphi\text{”}) \to \mathsf{Th}(\text{“}\psi\text{”}))$;
3. si φ est un énoncé Σ_1 — c'est-à-dire une formule close de la forme $\exists x_1 \dots x_n \, \psi(x_1 \dots x_n)$ où $\psi(x_1 \dots x_n)$ est la représentation canonique d'une relation primitive récursive — alors $\mathsf{PA} \vdash (\varphi \to \mathsf{Th}(\text{“}\varphi\text{”}))$;

3′. $\mathsf{PA} \vdash (\mathsf{Th}(\text{“}\varphi\text{”}) \to \mathsf{Th}(\text{“}\mathsf{Th}(\text{“}\varphi\text{”})\text{”}))$.

1, 2 et 4 sont les fameuses conditions de dérivabilité de Bernays.

$\Box = \mathsf{Th}(\dots)$

Une *valuation* du langage modal dans PA est une fonction qui associe à chaque symbole propositionnel un énoncé (formule close) de PA.

Soit v une telle valuation et φ un énoncé modal. On définit l'énoncé φ^v de PA comme suit :

— p^v est $v(p)$;
— $\bot^v$ est $\neg 0 = 0$;
— $(\varphi \to \psi)^v$ est $(\varphi^v \to \psi^v)$;
— $\Box \varphi^v$ est $\mathsf{Th}(\text{“}\varphi^v\text{”})$.

Remarque

Soit $\mathsf{Neg}(v, v')$ la formule de PA qui exprime que v' est la négation de la formule v. $\mathsf{Coh}(v)$ est la formule $\exists v' (\mathsf{Neg}(v, v') \wedge \neg \mathsf{Th}(v'))$. $\mathsf{Coh}(\text{“}\varphi\text{”})$ exprime donc que $\mathsf{PA} + \varphi$ est cohérent. En particulier, $\mathsf{Coh}(\text{“}0 = 0\text{”})$ est l'expression dans PA de la cohérence de PA. On voit que $\Diamond \varphi^v$ est un énoncé équivalent dans PA à $\mathsf{Coh}(\text{“}\varphi\text{”})$.

***

Nous abordons maintenant la partie délicate de l'article de Solovay, à savoir le plongement de certains modèles modaux dans PA.

Une relation R sur I est *bien fondée* si toute partie non vide X de I possède un élément i tel que $\{ j \mid j \, R \, i \} \cap X = \emptyset$.

Un *G-modèle* est un modèle modal dont la relation est transitive et dont la converse de la relation est bien fondée.

Un G-modèle fini (avec I fini) est caractérisé par le fait que sa relation est un ordre (partiel) strict fini, autrement dit, sa relation est finie, transitive et irréflexive.

Un ensemble d'énoncés Γ du langage modal a un modèle s'il existe un modèle modal $\mathcal{M}$ et un indice i de $\mathcal{M}$ tel que $\mathcal{M} \models_i \varphi$ pour tout φ de Γ.

Théorème. *Si Γ a un G-modèle fini, alors il y a une valuation v (du langage modal dans PA) telle que $\mathsf{PA} + \{ \varphi^v \mid \varphi \in \Gamma \}$ est cohérent.*

La démonstration comprendra deux étapes. Nous dérivons d'abord le théorème d'un lemme que nous démontrons ensuite.

Lemme. *Soit R une relation transitive et irréflexive sur $I = \{1, \ldots, n\}$. Il existe des énoncés $E_1, \ldots, E_n$ de PA qui vérifient les propriétés suivantes :*

1. *chaque E_i est cohérent avec PA ;*
2. *si $i \neq j$, alors $(E_i \wedge E_j)$ est incohérent avec PA ;*
3. $\mathsf{PA} \vdash (E_i \to \mathsf{Coh}(\text{“}E_j\text{”}))$*, si $i \mathrel{R} j$;*
4. $\mathsf{PA} \vdash (E_i \to \mathsf{Th}(\text{“}\bigvee\!\!\!\bigvee\{ E_j \mid i \mathrel{R} j \}\text{”}))$.

Pour montrer que ce lemme suffit à établir le théorème, nous considérons la valuation v définie par :

$$v(p) = \bigvee\!\!\!\bigvee \{ E_i \mid \mathcal{M} \models_i p \}$$

On prouve par induction que :

— $\mathcal{M} \models_i \varphi$ entraîne $\mathsf{PA} \vdash (E_i \to \varphi^v)$;
et
— $\mathcal{M} \not\models_i \varphi$ entraîne $\mathsf{PA} \vdash (E_i \to \neg\varphi^v)$.

- — si $\mathcal{M} \models_i p$, alors $(E_i \to \bigvee\!\!\!\bigvee\{ E_i \mid \mathcal{M} \models_i p \}$ est évidemment une tautologie ;
- — si $\mathcal{M} \not\models_i p$ et si $\mathcal{M} \models_j p$ alors, par le point 2 du lemme, $\mathsf{PA} \vdash (E_i \to \neg E_j)$. Donc, $\mathsf{PA} \vdash (E_i \to \neg \bigvee\!\!\!\bigvee\{ E_j \mid \mathcal{M} \models_j p \})$;
- — pour le cas où φ est le symbole $\perp$, on remarquera que $\mathsf{PA} \vdash (E_i \to \neg\neg 0 = 0)$;
- — si $\mathcal{M} \models_i (\psi \to \chi)$, alors, par hypothèse d'induction, $\mathsf{PA} \vdash (E_i \to \neg\psi^v)$ et/ou $\mathsf{PA} \vdash (E_i \to \chi^v)$. Donc, $\mathsf{PA} \vdash (E_i \to (\psi^v \to \chi^v))$.
- — si $\mathcal{M} \not\models_i (\psi \to \chi)$, alors, par hypothèse, $\mathsf{PA} \vdash (E_i \to \psi^v)$ et $\mathsf{PA} \vdash (E_i \to \neg\chi^v)$ et $\mathsf{PA} \vdash (E_i \to \neg(\psi^v \to \chi^v))$;

— si $\mathcal{M} \models_i \Box\psi$, alors, par hypothèse d'induction, pour tout j tel que $i\ R\ j$ on a que $\mathsf{PA} \vdash (E_j \to \psi^v)$. Donc, $\mathsf{PA} \vdash (\bigvee\!\!\!\bigvee\{\, E_j \mid i\ R\ j\,\} \to \psi^v)$ et de là $\mathsf{PA} \vdash (\mathsf{Th}(\text{“}\bigvee\!\!\!\bigvee\{\, E_j \mid i\ R\ j\,\}\text{”}) \to \mathsf{Th}(\text{“}\psi^v\text{”}))$, par les deux premières conditions de Bernays. Le point 4 du lemme permet finalement de conclure que $\mathsf{PA} \vdash (E_i \to \psi^v)$;

— si $\mathcal{M} \not\models_i \Box\psi$, alors, par hypothèse, il y a un j tel que $i\ R\ j$ et tel que $\mathsf{PA} \vdash (E_j \to \neg\psi^v)$. Les deux premières conditions de Bernays et le point 3 du lemme donnent successivement $\mathsf{PA} \vdash (\mathsf{Coh}(\text{“}E_j\text{”}) \to \mathsf{Coh}(\text{“}\neg\psi^v\text{”}))$ et $\mathsf{PA} \vdash (E_i \to \neg\mathsf{Th}(\text{“}\psi^v\text{”}))$.

Pour finir cette première étape, on remarquera que si $\varphi \in \Gamma$ et $\mathcal{M} \models_i \varphi$, alors $\mathsf{PA} \vdash (E_i \to \varphi^v)$. Le point 1 du lemme nous assure que $\mathsf{PA} \nvdash \neg E_i$. $\mathsf{PA} + \{\, \varphi^v \mid \varphi \in \Gamma\,\}$ est donc cohérent.

Démonstration du lemme

Commençons par construire les énoncés E_i. Concentrons-nous pour cela sur la *définition* de fonction suivante :

$g(n, 0) = 0$;
$g(n, m+1) = j$, si $g(n, m)\ R\ j$ ou $g(n, m) = 0$, ET
n est le nombre de Gödel d'une formule $\varphi(v, v')$ et $m+1$ est le nombre de Gödel d'une démonstration de la formule $\neg\exists v''\forall v(v'' \leq v \to \varphi(v, \bar{j}))$;
$g(n, m+1) = g(n, m)$, sinon.

g est une fonction primitive récursive à valeurs dans $\{0, \ldots, n\}$. Il y a donc une formule de PA qui représente canoniquement g. Soit $G(v'', v, v')$ cette formule. Par le théorème du point fixe, il existe une formule $H(v, v')$ telle que :

$$\mathsf{PA} \vdash (H(v, v') \leftrightarrow G(\text{“}H(v, v')\text{”}, v, v'))$$

Si $1 \leq i \leq n$, E_i est l'énoncé $\exists v''\forall v(v'' \leq v \to H(v, \bar{i}))$.

Quelques considérations préliminaires seront utiles avant d'aborder les quatre points du lemme.

$H(v, v')$ représente une fonction h qui a les propriétés suivantes :

$h(0) = 0$;
$h(m+1) = j$, si $h(m)\ R\ j$ ou $h(m) = 0$, ET $m+1$ est le nombre de Gödel d'une démonstration dans PA de $\neg E_j$;
$h(m+1) = h(m)$, sinon.

E_i exprime que i est la limite de h. h est croissante au sens suivant : si $m \leq n$, alors $h(m)\ R\ h(n)$ ou $h(m) = h(n)$. Comme la converse de R est bien fondée, h a une limite. Cette limite est non nulle si et seulement si h n'est pas constante. On peut formaliser ces remarques dans PA. Cela donne notamment :

a. $\mathsf{PA} \vdash (E_i \to \exists v\, H(v, \bar{i}))$;
b. $\mathsf{PA} \vdash (E_i \to \mathsf{Th}(\text{“}\neg E_i\text{”}))$;
c. $\mathsf{PA} \vdash (\forall v\, H(v, 0) \vee (E_1 \vee \ldots \vee E_n))$;
d. $\mathsf{PA} \vdash (\mathsf{Th}(\text{“}\neg E_i\text{”}) \to (E_1 \vee \cdots \vee E_n))$.

Pour obtenir d., on formalise l'argument suivant :
Sinon, h est la fonction constante 0, par c. Mais si $m+1$ « démontre » $\neg E_i$, on a à la fois $h(m) = 0$ et $h(m+1) = i = 0$.

e. $\mathsf{PA} \vdash (\exists v H(v, \bar{i}) \to (\neg E_i \to \bigvee\{\, E_j \mid i\ R\ j \,\}))$

Nous en arrivons enfin à la démonstration des quatre points du lemme.

1. Nous supposerons que les théorèmes de PA sont vrais dans le modèle standard $\mathcal{N}$.
Si $\mathcal{N} \models E_i$, alors $\mathcal{N} \models \mathsf{Th}(\text{“}\neg E_i\text{”})$ — par la remarque b.— et $\mathsf{PA} \vdash \neg E_i$, et donc $\mathcal{N} \not\models E_i$. Tous les E_i sont dès lors faux et h est « en vérité » la fonction constante 0.
Cela étant, si E_i n'est pas cohérent avec PA, alors $\mathsf{PA} \vdash \neg E_i$ et donc $\mathcal{N} \models \mathsf{Th}(\text{“}\neg E_i\text{”})$. Mais alors $\mathcal{N} \models (E_1 \vee \ldots \vee E_n)$, par la remarque d. Or on vient de voir que cela est impossible.
2. Ceci est trop évident pour qu'on s'y attarde.
3. On utilisera ici le fait que $\mathsf{PA} \vdash (\mathsf{Th}(x) \to \forall z \exists y (y > z \wedge \mathsf{Dem}(y, x)))$ (toute démonstration peut être allongée).
Supposons, par absurde, que E_i et $\mathsf{Th}(\text{“}\neg E_j\text{”})$ $(i\ R\ j)$ et raisonnons dans PA. Soit v'' tel que $\forall v\, (v \geq v'' \to H(v, \bar{i}))$. Choisissons un $y > v''$ tel que $\mathsf{Dem}(y, \text{“}\neg E_j\text{”})$. Alors on a V^* la fois $H(y, \bar{i})$ et $H(y, \bar{j})$. Cela est contradictoire, car H représente une fonction et R est irréflexive.
4. Les formules $\exists v\, H(v, \bar{i})$ sont Σ_1. Donc,

$$\mathsf{PA} \vdash (\exists v H(v, \bar{i}) \to \mathsf{Th}(\text{“}\exists v H(v, \bar{i})\text{”})$$

Par la remarque a., $\mathsf{PA} \vdash (E_i \to \mathsf{Th}(\text{“}\exists v H(v, \bar{i})\text{”})$ et, par la remarque e.,

$$\mathsf{PA} \vdash (\exists v H(v, \bar{i}) \to (E_i \to \bigvee\{\, E_j \mid i\ R\ j \,\}))$$

Donc, par les deux premières conditions de Bernays, $\mathsf{PA} \vdash (\mathsf{Th}(\text{“}\exists v H(v, \bar{i})\text{”}) \to (\mathsf{Th}(\text{“}\neg E_i\text{”}) \to \mathsf{Th}(\text{“}\bigvee\{\, E_j \mid i\ R\ j \,\}\text{”})))$. Combinant ceci avec la remarque b., on obtient

$$\mathsf{PA} \vdash (E_i \to \mathsf{Th}(\text{“}\bigvee\{\, E_i \mid i\ R\ j \,\}\text{”}))$$

■

Le système G, avec « G » comme Gödel

G est un système de logique modale introduit pour dégager la logique modale implicite de PA.

Les axiomes de G se répartissent en trois groupes.

1. les tautologies ;
2. les énoncés (modaux) de la forme :

$$\Box(\varphi \to \psi) \to (\Box\,\varphi \to \Box\psi) \qquad \text{(axiomes de distribution)}$$

3. les énoncés de la forme :

$$\Box(\Box\,\varphi \to \varphi) \to \Box\varphi \qquad \text{(axiomes de Löb)}$$

Les règles de G sont :

1. le *modus ponens* :

$$\text{si } G \vdash \varphi \to \psi \text{ et } G \vdash \varphi, \text{ alors } G \vdash \psi$$

2. la nécessitation :

$$\text{si } G \vdash \varphi, \text{ alors } G \vdash \Box\,\varphi$$

Proposition.

1. *Si* $G \vdash (\varphi \to \psi)$, *alors* $G \vdash (\Box\varphi \to \Box\psi)$ *et* $G \vdash (\Diamond\varphi \to \Diamond\psi)$ *;*
2. $G \vdash (\Box(\varphi_1 \wedge \ldots \wedge \varphi_n) \leftrightarrow (\Box\,\varphi_1 \wedge \ldots \wedge \Box\,\varphi_n))$ *;*
3. *Substitution des équivalents : si* $G \vdash (\varphi \to \psi)$, *alors* $G \vdash (\chi(\varphi) \to \chi(\psi))$ *;*
4. $G \vdash (\Box\,\varphi \to \Box\Box\,\varphi)$ *;*
5. *G est cohérent et* $G \nvdash (\Box p \to p)$.

Démonstration. Les points 1, 2 et 3 se démontrent aisément.

Les points 2 et 3 et le fait que toute tautologie est un axiome de G donnent :

$$G \vdash (\varphi \to (\Box(\varphi \wedge \Box\,\varphi) \to (\varphi \wedge \Box\,\varphi)))$$

De là, on déduit successivement :

$G \vdash (\Box\,\varphi \to \Box(\Box(\varphi \wedge \Box\varphi) \to (\varphi \wedge \Box\,\varphi)))$, par le point 1 ;
$G \vdash (\Box\,\varphi \to \Box(\varphi \wedge \Box\,\varphi))$, par l'axiome de Löb ;
$G \vdash (\varphi \to \Box\Box\,\varphi)$, par le point 2.

Ce qui prouve 4.

5. Remplaçons dans une démonstration de G les expressions de la forme $\Box\,\psi$ qui ne sont pas sous-formules d'une expression (différente) de la même forme, par $\neg\bot$. On montre facilement que l'on obtient ainsi une suite de tautologies. Si on applique ce procédé à une démonstration de $(\Box p \to p)$, on doit admettre que $(\neg\bot \to p)$ est une tautologie. Ce qui est absurde.

G dans PA

Les théorèmes de G sont valables dans PA, au sens précis suivant :
si v est une valuation du langage modal dans PA, alors pour tout énoncé modal φ :

$$G \vdash \varphi \text{ entraîne } \mathsf{PA} \vdash \varphi^v$$

Les deux premières conditions de Bernays rendent compte des axiomes de distribution et de la règle de nécessitation. Il reste à vérifier l'axiome de Löb.

Le théorème de Löb dit que si φ est un énoncé de PA, alors

$$\mathsf{PA} \vdash (\mathsf{Th}(\text{“}\varphi\text{”}) \to \varphi) \text{ ssi } \mathsf{PA} \vdash \varphi$$

Ce théorème n'est rien d'autre qu'une présentation du second théorème d'incomplétude de Gödel :

si $\mathsf{PA} + \varphi$ est cohérent, alors $\mathsf{PA} + \varphi$ ne démontre pas la cohérence de $\mathsf{PA} + \varphi$

Autrement dit :

si $\mathsf{PA} \nvdash \varphi$, alors $\mathsf{PA} + \neg\varphi \nvdash \neg\mathsf{Th}(\text{“}\varphi\text{”})$
donc $\mathsf{PA} \vdash (\mathsf{Th}(\text{“}\varphi\text{”}) \to \varphi)$ entraîne $\mathsf{PA} \vdash \varphi$

Cet argument est formalisable dans PA. Dès lors, pour toute valuation v et pour tout énoncé modal φ,

$$\mathsf{PA} \vdash (\Box(\Box\varphi \to \varphi) \to \Box\varphi)^v$$

Remarque

L'axiome de Löb acquiert une signification plus éclairante, si on le formule en termes de possibilité :

$$(\Diamond\varphi \to \Diamond(\varphi \wedge \Diamond\neg\varphi))$$

(si φ est possible (cohérent), alors il est possible (cohérent) de supposer à la fois φ et l'impossibilité de φ (la non-cohérence de φ)).

Théorème de complétude de Solovay

Pour toute valuation v, $\mathsf{PA} \vdash \varphi^v$ ssi $G \vdash \varphi$.

Nous venons de voir que $G \vdash \varphi$ entraîne $\mathsf{PA} \vdash \varphi^v$, pour toute valuation v. Nous savons également que si $\neg\varphi$ a un G-modèle fini, alors $\mathsf{PA} + \neg\varphi^v$ est cohérent, pour au moins une valuation v. Le théorème de complétude sera prouvé, si nous parvenons à montrer que tout énoncé modal cohérent avec G possède un G-modèle fini : car si $G \nvdash \varphi$, on disposera d'un G-modèle fini pour $\{\neg\varphi\}$ et d'une valuation v telle que $\mathsf{PA} + \neg\varphi^v$ est cohérent et donc $\mathsf{PA} \nvdash \varphi^v$.

Définition. Un ensemble Γ d'énoncés modaux est *cohérent avec G* s'il n'existe pas de partie finie Δ de Γ telle que $G \vdash \neg \bigwedge\!\!\!\bigwedge \Delta$.

Nous démontrons maintenant que si $\{\varphi\}$ est cohérent avec G, alors φ a un G-modèle fini.

Fixons $n,m \geq 0$, deux nombres naturels. Γ est l'ensemble des énoncés modaux dont les symboles propositionnels sont parmi les n premiers de la liste et qui mentionnent au plus m symboles $\Box$ et/ou $\rightarrow$. Si ϵ est une fonction de Γ dans $\{V, F\}$, Γ_ϵ désignera $\{\, \varphi_\epsilon \mid \varphi \in \Gamma \,\}$, où φ_ϵ est φ, si $\epsilon(\varphi) = V$ et φ_ϵ est $\neg\varphi$, si $\epsilon(\varphi) = F$.

Soit $I = \{\, \epsilon \mid \Gamma_\epsilon$ est cohérent avec $G \,\}$.

Il est clair que Γ, chaque Γ_ϵ et I sont des ensembles *finis* et non vides (par le point 5 de la proposition). En outre, si $\{\varphi\}$ est cohérent avec G il existe n, m et ϵ tel que $\varphi \in \Gamma_\epsilon$.

Si $\epsilon, \delta \in I$, on définit : $\epsilon \; R \; \delta$ ssi $\{\, \varphi \mid \Box\, \varphi \in \Gamma_\epsilon \,\} \cup \{\, \Box\, \varphi \mid \Box\, \varphi \in \Gamma_\epsilon \,\} \subseteq \Gamma_\delta$ et il existe un énoncé $\neg\Box\, \psi$ dans Γ_ϵ tel que $\{\neg\psi, \Box\psi\} \subseteq \Gamma_\delta$.

R est une relation *irréflexive* sur I. De plus R est *transitive* : supposons, en effet, que $\epsilon \; R \; \delta$ et $\delta \; R \; \gamma$ et vérifions qu'on a bien $\epsilon \; R \; \gamma$. En premier lieu, si $\Box\varphi \in \Gamma_\epsilon$, alors $\Box\, \varphi \in \Gamma_\delta$ (car $\epsilon R \delta$) et $\{\Box\, \varphi, \varphi\} \subseteq \Gamma_\gamma$ (car $\delta R \gamma$). Par ailleurs, il y a un énoncé de la forme $\neg\Box\, \psi$ dans Γ_δ tel que $\neg\psi$ et $\Box\, \psi$ sont dans Γ_γ. $\Box\, \psi$ ne peut être un élément de Γ_ϵ, car, étant donné que $\epsilon \; R \; \delta$, cela impliquerait que $\Box\, \psi$ est dans Γ_δ qui est cohérent avec G. $\neg\Box\, \psi$ est donc dans Γ_ϵ et $\{\neg\psi, \Box\psi\} \subseteq \Gamma_\gamma$.

Lemme.
1. Si $X \subseteq \Gamma$ et X est cohérent avec G, alors il y a un ϵ tel que $X \subseteq \Gamma_\epsilon$.
2. Si $\neg\Box\, \psi \in \Gamma_\epsilon$, alors il existe un δ tel que $\epsilon \; R \; \delta$ et $\neg\psi \in \Gamma_\delta$.

Démonstration. Comme le premier point est assez classique, nous n'examinerons que le point 2.

Soit $X = \{\, \varphi \mid \Box\, \varphi \in \Gamma_\epsilon \,\} \cup \{\, \Box\varphi \mid \Box\, \varphi \in \Gamma_\epsilon \,\} \cup \{\Box\psi, \neg\psi\}$.

X est cohérent avec G. Sinon,

$$G \vdash \bigwedge\!\!\!\bigwedge \{\, \varphi \mid \Box\, \varphi \in \Gamma_\epsilon \,\} \wedge \bigwedge\!\!\!\bigwedge \{\, \Box\, \varphi \mid \Box\, \varphi \in \Gamma_\epsilon \,\} \rightarrow (\Box\psi \rightarrow \psi).$$

Dès lors,

$$G \vdash \bigwedge\!\!\!\bigwedge \{\, \Box\, \varphi \mid \Box\, \varphi \in \Gamma_\epsilon \,\} \wedge \bigwedge\!\!\!\bigwedge \{\, \Box\Box\, \varphi \mid \Box\varphi \in \Gamma_\epsilon \,\} \rightarrow \Box(\Box\psi \rightarrow \psi),$$

par les trois premiers points de la proposition.

Enfin,

$$G \vdash \bigwedge\!\!\!\bigwedge \{ \Box\varphi \mid \Box\varphi \in \Gamma_\epsilon \} \to \Box\psi,$$

par l'axiome de Löb et le point 4 de la proposition.

Γ_ϵ n'est donc pas cohérent avec G.

X étant cohérent avec G, on choisit un δ tel que $X \subseteq \Gamma_\delta$. Ce δ a les propriétés voulues.

Pour préciser le modèle que nous construisons, il reste à définir M par la condition : $M(p, \epsilon) = \epsilon(p)$.

Soit $\mathcal{M} = \langle I, R, M \rangle$. Nous montrons que pour tout ϵ de I et tout φ de Γ,

$$\mathcal{M} \models_\epsilon \varphi \qquad \text{ssi} \qquad \varphi \in \Gamma_\epsilon$$

La seule chose intéressante à considérer est le cas où φ est de la forme $\Box\psi$.

Si $\Box\psi \in \Gamma_\epsilon$, alors, pour tout δ tel que $\epsilon R \delta$, $\psi \in \Gamma_\epsilon$ (par définition de R), donc, par hypothèse d'induction, $\mathcal{M} \models_\delta \psi$, pour ces δ. Cela revient à dire que $\mathcal{M} \models_\epsilon \Box\psi$.

Si $\Box\psi \notin \Gamma_\epsilon$, alors, $\neg\Box\psi \in \Gamma_\epsilon$ et, par le lemme, il y a un δ tel que $\epsilon\ R\ \delta$ et $\neg\psi \in \Gamma_\delta$. Donc, $\mathcal{M} \not\models_\epsilon \Box\psi$. ■

Nous venons de démontrer, en même temps que le théorème de Solovay, un théorème de complétude pour G :

Si φ est vrai en chaque indice de tous les G-modèles finis, alors $G \vdash \varphi$.

En outre, en suivant la méthode prescrite pour construire $\mathcal{M}$ à partir de Γ, on constate que partant d'un énoncé φ, il suffit d'un nombre fini de démarches pour s'assurer de sa non-théorèmité. Nous venons donc également de suggérer une démonstration de :

$$G \text{ est décidable,}$$

et de

$$\{ \varphi \mid \text{ pour toute valuation } v, \mathsf{PA} \vdash \varphi^v \} \quad \text{est décidable.}$$

Remarques conclusives

1. φ est un théorème de G ssi φ est vrai à chaque indice de tous les G-modèles.

En vertu de ce qui précède, il reste à montrer que tout théorème de G est vrai à chaque indice dans chaque G-modèle. Examinons l'axiome de Löb, les autres cas étant aisés.

Supposons par absurde que pour un certain $\mathcal{M}$ et un i, $\mathcal{M} \models_i \Box(\Box\varphi \to \varphi)$ et $\mathcal{M} \not\models_i \Box\varphi$. Considérons l'ensemble $X = \{\, j \mid i\ R\ j$ et $\mathcal{M} \models_j \neg\varphi \,\}$. Si $j \in X$, alors $\mathcal{M} \models_j \neg\varphi$ et $\mathcal{M} \models_j (\Box\varphi \to \varphi)$ et, donc, $\mathcal{M} \models_j \neg\Box\varphi$. Il s'ensuit qu'il existe un indice k tel que $j\ R\ k$ et $\mathcal{M} \models_k \neg\varphi$. Comme R est transitive, $\{\, k \mid j\ R\ k \,\} \cap X \neq \emptyset$. La converse de R n'est donc pas bien fondée.

2. Dans son article *Extremely undecidable sentences*, J.S.L., 1982, pp.191–196, G. Boolos a renforcé le théorème de Solovay. Il a construit, par point fixe, **une** valuation v du langage modal dans PA, telle que :

$$G \vdash \varphi \qquad \text{ssi} \qquad \mathsf{PA} \vdash \varphi^v,$$

pour tout énoncé modal φ.