

Ceci est une version préliminaire du chapitre 10 de l'ouvrage suivant : F. Pogliolesi et P. Wagner, éd., *Précis de philosophie de la logique et des mathématiques*, vol. 1, *Philosophie de la logique*, Paris, Éd. de la Sorbonne, 2021, p. 423-464. Merci de citer uniquement la version publiée.

10 Le phénomène d'incomplétude

Pierre Wagner

Existe-t-il un système arithmétiquement *complet*, au sens où il contiendrait la clef de toute question susceptible d'être soulevée en théorie des nombres ? Plus précisément : existe-t-il un ensemble A d'axiomes qui décide de tout énoncé arithmétique φ , c'est-à-dire tel que la négation de φ soit démontrable à partir de A si, et seulement si, φ ne l'est pas ? Si la réponse mathématique négative que Gödel donna en 1931¹ a reçu des interprétations diverses, parfois contestables, voire fantaisistes, nul ne nie la profondeur de ses conséquences philosophiques, qui dépassent la question de la complétude de l'arithmétique *stricto sensu* et dont la discussion requiert une intelligence des théorèmes d'incomplétude et des méthodes requises par leur démonstration. Les subtilités qui distinguent leurs multiples variantes risquent en effet d'induire des erreurs d'interprétation, malheureusement assez communes, si l'on n'y prend pas garde. Avant tout examen du phénomène d'incomplétude mis au jour par Gödel et de ses implications philosophiques, une série de définitions préalables est donc requise, ainsi qu'un aperçu des différentes versions des théorèmes en question et des méthodes utiles à leur démonstration. Dans un second temps, nous examinons quelques-uns des enjeux discutés dans la littérature, relatifs aux mathéma-

1. Annoncée publiquement en septembre 1930 à Königsberg à l'occasion du second congrès pour l'épistémologie des sciences exactes, la démonstration du premier théorème d'incomplétude fut publiée l'année suivante dans Gödel [1931]. Une nouvelle version fut présentée (Gödel [1934]) dans un cours donné par Gödel à Princeton dont les notes, recueillies par Kleene et Rosser, sont publiées dans Gödel [1986].

tiques ordinaires, à notre compréhension de l'arithmétique, aux remarques de Wittgenstein et à la philosophie carnapienne des mathématiques².

1. Variété du premier théorème d'incomplétude

Soit \mathcal{L}_A un langage formel du premier ordre avec égalité dont les signes non logiques sont « 0 », « s », « + » et « × », respectivement interprétés, dans le domaine \mathbb{N} des entiers naturels, par zéro et les fonctions *successeur*, *somme* et *produit*. La \mathcal{L}_A -structure \mathcal{N} ainsi définie est l'interprétation standard de \mathcal{L}_A .

Soit un système de dérivation formelle (système hilbertien, déduction naturelle ou autre) qui définit une relation \vdash de dérivabilité entre un ensemble Γ d'énoncés et un énoncé φ quelconques. On dit qu'un ensemble T d'énoncés d'un langage formel \mathcal{L} est une *théorie* si T est clos par dérivation formelle, c'est-à-dire si pour tout énoncé φ de \mathcal{L} , $\varphi \in T$ si $T \vdash \varphi$. Une partie A d'un ensemble U est dite *décidable* s'il existe une procédure effective applicable à un élément u quelconque de U permettant de déterminer si $u \in A$ ou $u \notin A$. Une théorie T est dite *axiomatisable* s'il existe un ensemble A d'énoncés du langage de T qui soit décidable et dont la clôture déductive A^+ soit égale à T . Une théorie est souvent définie par un ensemble d'axiomes dont elle est la clôture déductive. L'arithmétique de Robinson, Q , est ainsi définie par les sept axiomes suivants :

- $Q_1 \quad \forall x (0 \neq s(x))$
- $Q_2 \quad \forall x \forall y (s(x) = s(y) \rightarrow x = y)$
- $Q_3 \quad \forall x (x \neq 0 \rightarrow \exists y (s(y) = x))$
- $Q_4 \quad \forall x (x + 0 = x)$
- $Q_5 \quad \forall x \forall y (x + s(y) = s(x + y))$
- $Q_6 \quad \forall x (x \times 0 = 0)$
- $Q_7 \quad \forall x \forall y (x \times s(y) = (x \times y) + x).$

2. Nous prenons le parti de ne pas insister sur les implications de l'incomplétude pour la question du mécanisme en philosophie de l'esprit – trop éloignée de la philosophie de la logique proprement dite – et de ne traiter que brièvement de ses conséquences pour le programme de Hilbert, dont il est également question dans le volume 2 du présent *Précis*. On trouvera d'autres indications sur le premier point dans Shapiro [1998] et sur le second dans Sieg [2013]. L'espace manque pour rendre compte d'autres débats, notamment ceux qui touchent à la vérité, aux principes de réflexion et au déflationnisme aléthique (sur ces questions voir le chapitre 3).

L'arithmétique de Peano, PA , est une extension de Q obtenue par ajout d'un schéma d'axiomes dit « schéma d'induction » :

$$\left(\varphi(0) \wedge \forall x \left(\varphi(x) \rightarrow \varphi(s(x)) \right) \right) \rightarrow \forall x \varphi(x)$$

où $\varphi(x)$ est une formule quelconque de \mathcal{L}_A .

Soit T' une théorie dont le langage \mathcal{L}' inclut le langage \mathcal{L} d'une théorie T . On dit que T' est une *extension* de T si $T \subseteq T'$. Un ensemble Γ d'énoncés d'un langage formel \mathcal{L} est dit *consistant* si aucun énoncé φ de \mathcal{L} n'est tel que $\Gamma \vdash \varphi$ et $\Gamma \vdash \neg\varphi$, *inconsistant* s'il n'est pas consistant. Une théorie T est dite *complète* si pour tout énoncé φ du langage de T , $T \vdash \varphi$ ou $T \vdash \neg\varphi$, *incomplète* si elle n'est pas complète. Enfin, un énoncé φ du langage de T est dit *indécidable* dans T si $T \not\vdash \varphi$ et $T \not\vdash \neg\varphi$ (on ne confondra pas la décidabilité d'un énoncé avec celle d'un ensemble, définie ci-dessus³). On montre assez facilement qu'une théorie axiomatisable, consistante et complète est décidable (pour déterminer si un énoncé φ du langage de T est dans T , il suffit d'énumérer systématiquement les théorèmes jusqu'à trouver φ ou $\neg\varphi$).

Sur la base de ces définitions, une première version du théorème d'incomplétude peut être énoncée. Soit T une extension axiomatisable de Q ;

si T est consistante, elle est incomplète.

Le premier théorème d'incomplétude de Gödel (TIG_1) est cependant plus précis. Sa démonstration montre comment, pour toute extension axiomatisable T de Q , il est possible de définir un énoncé G de \mathcal{L}_A (dont la formulation dépend de T et qu'on note donc parfois G_T) tel que :

si T est consistante, alors G est indécidable dans T .

De surcroît, la définition de G est telle que si T est consistante, G est vrai (au sens de « vrai dans l'interprétation standard de \mathcal{L}_A »). En effet, en suivant l'interprétation standard des signes de \mathcal{L}_A , il est possible de lire G comme signifiant qu'il n'existe pas de preuve de G dans T^4 .

Le point remarquable, dans ce théorème, est le contraste entre l'extrême pauvreté du langage dans lequel G est formulé et la puissance démonstrative potentielle de T : quelle que soit la force des axiomes de l'extension T de Q , il existe un énoncé de l'indigent langage \mathcal{L}_A qui est indécidable dans T , pourvu que T soit axiomatisable et consistante. En outre, la complexité logique de

3. Cette notion de décidabilité ne dépend nullement de l'existence d'une procédure effective; « indécidable » en ce sens est une simple abréviation de « ni démontrable ni réfutable ».

4. Sur les conditions exactes auxquelles G se prête à une telle lecture, voir Milne [2007].

l'énoncé G est remarquablement faible : G est équivalent à un énoncé de la forme $\forall x \varphi(x)$ où $\varphi(x)$ est calculable (au sens où il existe une procédure effective permettant de déterminer, pour tout entier naturel n , si $\varphi(\bar{n})$ est vrai, \bar{n} étant la représentation standard de n dans \mathcal{L}_A ⁵). Les énoncés de cette forme sont dits Π_1^0 (ou Π_1). Plus généralement, un énoncé arithmétique est Π_1 s'il est de la forme $\forall x_1 \dots \forall x_n \varphi(x_1, \dots, x_n)$ où $\varphi(x_1, \dots, x_n)$ est sans quantificateur non borné, et donc calculable⁶.

La démonstration de Gödel publiée en 1931 prenait pour base non pas Q^7 mais un système T dérivé des *Principia Mathematica* de Russell et Whitehead et faisait l'hypothèse non de sa consistance mais de son ω -consistance. Plus précisément, Gödel montrait que l'énoncé G était tel que :

si T est consistant, alors $T \not\vdash G$;

si T est ω -consistant, alors $T \not\vdash \neg G$.

Par définition, T est ω -consistant s'il n'existe aucune formule $\varphi(x)$ à une variable libre telle que $T \vdash \exists x \varphi(x)$ et $T \vdash \neg \varphi(\bar{n})$ pour tout entier naturel n . Deux conséquences immédiates de cette définition sont d'une part que \mathcal{N} n'est pas modèle d'une théorie ω -inconsistante et d'autre part que toute théorie ω -consistante est consistante. En réalité, la démonstration de Gödel requiert seulement que T soit 1-consistante, c'est-à-dire que l'hypothèse d' ω -consistance vaille pour les formules dites Σ_1^0 (ou Σ_1), qui sont de la forme $\exists x_1 \dots \exists x_n \varphi(x_1, \dots, x_n)$, où $\varphi(x_1, \dots, x_n)$ est sans quantificateur non borné, et donc calculable⁸.

En 1936, un raffinement dû à Rosser permet de remplacer l'hypothèse d' ω -consistance par l'hypothèse plus faible de consistance simple, à condition toutefois de remplacer l'énoncé G de Gödel par un énoncé G_R , qui n'est pas équivalent à G . On parle alors du théorème de « Gödel-Rosser⁹ » :

si T est consistant, alors $T \not\vdash G_R$ et $T \not\vdash \neg G_R$.

La démonstration donnée par Gödel, aussi bien que celle de Rosser, n'a aucunement recours au concept de vérité ni à aucun autre concept sémantique.

5. Sur ce point, la terminologie est variable et on dit aussi qu'un tel $\varphi(x)$ est *décidable*.

6. Un exemple bien connu d'énoncé Π_1 est la conjecture de Goldbach : quel que soit l'entier naturel pair n supérieur ou égal à 4, il existe deux nombres premiers k et m (inférieurs à n) tels que $k + m = n$.

7. Introduite par Raphael Robinson en 1950, Q est une théorie arithmétique très faible, dont les ressources suffisent néanmoins pour la démonstration gödelienne de TIG_1 .

8. Comme exemple d'énoncé Σ_1 , on peut penser à celui qui affirme l'existence d'une solution de l'équation diophantienne $P(x_1, \dots, x_n) = 0$ (dans une telle équation, le terme $P(x_1, \dots, x_n)$ est un polynôme à coefficients entiers et la solution recherchée un n -uplet d'entiers).

9. Sur la définition de G , voir § 2.2; sur celle de G_R , voir Smith [2013], chap. 25.

On dit en ce sens qu'elle est purement syntaxique (bien qu'elle soit formulée dans des mathématiques informelles). Ce caractère de la démonstration historique donnée dans Gödel [1931] importait au plus haut point au regard des implications de TIG_1 pour le programme de Hilbert pour le fondement des mathématiques, lui aussi purement syntaxique. Par ailleurs, le scepticisme général qui entourait l'idée de vérité mathématique, lorsqu'elle est distinguée de celle de proposition démontrée, explique que Gödel ait tenu à donner une démonstration syntaxique de TIG_1 . Une version affaiblie de ce théorème est néanmoins possible, qui fait l'hypothèse non de la consistance de T mais de sa fiabilité :

si T est fiable, alors T est incomplète.

Au sens où nous l'entendons ici, une théorie arithmétique est dite *fiable* si tous ses théorèmes sont vrais. La fiabilité est donc une condition plus forte que la consistance car une théorie simplement consistante peut tout à fait démontrer des théorèmes faux. Cette version de TIG_1 est donc plus faible que les précédentes. Son intérêt est qu'elle en simplifie la démonstration. Le prix à payer est l'usage de la notion sémantique de vérité.

Une autre version, de force intermédiaire, a recours à la Σ_1 -fiabilité, ou fiabilité réduite aux énoncés Σ_1 :

si T est Σ_1 -fiable, alors T est incomplète.

Mais Σ_1 -fiabilité et 1-consistance sont en fait des conditions équivalentes.

En 1931, si la démonstration de Gödel était donnée pour un système formel particulier, il était clair que la méthode utilisée pouvait être transposée aux systèmes apparentés, ainsi qu'à tous les systèmes formels pour l'ensemble des mathématiques connus à l'époque (qui incluaient bien sûr l'arithmétique). Il fallut pourtant attendre 1936 et la caractérisation générale de la notion de procédure effective (par les machines de Turing, le lambda-calcul ou les fonctions récursives) pour qu'une version générale du théorème d'incomplétude pût être formulée, valable pour tout système formel possédant des ressources arithmétiques minimales (typiquement, celles de l'arithmétique de Robinson). C'est en effet sur la base de cette notion de procédure effective, ou d'algorithme, qu'une définition générale des théories formelles peut être donnée. Une telle théorie peut être définie comme un *ensemble effectivement énumérable d'énoncés*, sur un langage formel quelconque ; un ensemble étant dit *effectivement énumé-*

table s'il existe une procédure effective permettant d'énumérer ses éléments (c'est-à-dire, dans le cas d'une théorie, ses théorèmes)¹⁰.

En 1929, dans sa thèse de doctorat publiée l'année suivante (Gödel [1930]), Gödel démontra le théorème de *complétude* pour la logique du premier ordre. Dans ce théorème, c'est une *logique* (au sens d'une famille de langages formels munie d'une sémantique) qui est dite complète, non une *théorie*, ce qui signifie que les formules valides (*i.e.* vraies dans toute structure d'interprétation du langage) sont effectivement énumérables. Ou encore, selon une autre formulation : que toute formule valide du langage en question est un théorème, dans un certain système formel dont on dit alors qu'il est *sémantiquement complet*. Dans la discussion du phénomène d'incomplétude, on se gardera donc de confondre la complétude (sémantique) d'une logique et l'incomplétude (syntaxique) d'une théorie.

L'ensemble des énoncés vrais de \mathcal{L}_A forme une théorie (car il est déductivement clos) qui est assurément consistante (un énoncé et sa négation ne sont pas l'un et l'autre vrais) et complète (tout énoncé φ est tel que soit φ soit $\neg\varphi$ est vrai). Selon le théorème d'incomplétude, cette théorie (l'« arithmétique vraie ») n'est donc pas axiomatisable : aucun ensemble décidable d'énoncés de \mathcal{L}_A n'a pour clôture déductive les énoncés vrais de \mathcal{L}_A . Une théorie axiomatique comme PA n'est pas seulement incomplète si elle est consistante ; elle est surtout incomplétable par adjonction de nouveaux axiomes : toute extension consistante et axiomatisable de Q , et *a fortiori* de PA , est incomplète. C'est cette incomplétude essentielle qui autorise à parler d'un « phénomène d'incomplétude ».

2. Démonstrations du premier théorème d'incomplétude

2.1 Démonstrations fondées sur la théorie de la calculabilité

Afin de mieux saisir l'apport spécifique des démonstrations d'incomplétude dues à Gödel, nous esquissons deux exemples différents de démonstrations, qui reposent sur la théorie de la calculabilité. Le premier est librement adapté d'un ouvrage de Stillwell (Stillwell [2018], p. 76 et p. 80-81), le second d'un ouvrage de Franzén (Franzén [2005], p. 73-74).

10. Dans les démonstrations, les termes *décidable*, *calculable* et *effectivement énumérable* ont un sens technique précis, qui repose sur les fonctions récursives, les machines de Turing, ou d'autres notions apparentées étudiées en théorie de la calculabilité.

Les procédures effectives applicables à un entier naturel quelconque doivent pouvoir être décrites en une suite finie de signes et sont donc effectivement énumérables. Soit $A_0, A_1, A_2 \dots$ une énumération de ces procédures et $f_0, f_1, f_2 \dots$ les fonctions numériques partielles telles que pour tout $k \geq 0$, A_k est un algorithme pour le calcul de la fonction f_k , pour un entier n quelconque. Soit D (« D » comme « diagonalisation ») l'ensemble d'entiers naturels défini par

$$D = \{n \in \mathbb{N} \mid f_n(n) = 0\}$$

où l'on remarque que l'indice de f est identique à son argument. D est effectivement énumérable (pour énumérer les éléments de D , il suffit d'appliquer l'algorithme suivant : pour chaque entier naturel n , effectuer les n premières étapes des algorithmes A_0 à A_n appliqués à n et ajouter k à la liste des éléments de D dès que l'on trouve un entier naturel k tel que A_k appliqué à k produit le résultat 0 en n étapes). Montrons maintenant que D n'est pas calculable. Supposons qu'il le soit et nommons c_D sa fonction caractéristique. Pour tout entier n ,

$$c_D(n) = 1 \text{ si } n \in D$$

$$c_D(n) = 0 \text{ si } n \notin D$$

c_D est une fonction numérique et il existe donc un entier naturel k tel que pour tout $n \in \mathbb{N}$, $c_D(n) = f_k(n)$. On arrive à une contradiction. En effet, en appliquant les définitions qui viennent d'être données, on trouve que

$$k \notin D \text{ ssi } c_D(k) = 0 \text{ ssi } f_k(k) = 0 \text{ ssi } k \in D.$$

Notons \bar{D} le complémentaire de D dans \mathbb{N} . \bar{D} n'est pas effectivement énumérable, car si D et \bar{D} le sont l'un et l'autre, alors D est calculable. Il en résulte qu'il n'existe pas de théorie formelle fiable telle que tous les énoncés vrais de la forme $n \in \bar{D}$ (pour $n \in \mathbb{N}$) sont des théorèmes, car il existerait une énumération effective des énoncés $n \in \bar{D}$, et \bar{D} serait effectivement énumérable. D'où le théorème d'incomplétude suivant, pour une théorie T dont le langage admet les expressions $n \in \bar{D}$ (pour $n \in \mathbb{N}$) comme énoncés :

si T est fiable, alors T est incomplète.

Ce qu'on ne voit pas, dans cet exemple de démonstration, est comment les énoncés indécidables pourraient être formalisés dans \mathcal{L}_A . Il en va différemment de l'exemple suivant, qui requiert quelques nouvelles définitions préalables et fait appel à un théorème arithmétique démontré par Matiassevitch en 1970.

Les énoncés arithmétiques Π_1 (de la forme $\forall x_1 \dots \forall x_n \varphi(x_1, \dots, x_n)$, où $\varphi(x_1, \dots, x_n)$ est calculable) ont ceci de notable qu'ils sont vrais s'ils sont dé-

montrables dans une extension consistante T de Q . En effet, si un tel énoncé est faux, il existe n entiers k_1, \dots, k_n tels que $\varphi(k_1, \dots, k_n)$ est faux, et la formule $\neg\varphi(k_1, \dots, k_n)$ est démontrable dans T (car elle l'est déjà dans Q); la négation de l'énoncé initial est alors démontrable dans T , ce qui contredit l'hypothèse de consistance.

Une équation diophantienne est de la forme $P(x_1, \dots, x_n) = 0$, où le terme $P(x_1, \dots, x_n)$ est un polynôme à coefficients entiers. Une telle équation admet une solution s'il existe n entiers k_1, \dots, k_n tels que $P(k_1, \dots, k_n) = 0$. On considère les énoncés Π_1 de la forme suivante :

L'équation diophantienne $P(x_1, \dots, x_n) = 0$ n'a pas de solution.

Soit T une théorie arithmétique Σ_1 -fiable. T est consistante. L'ensemble des énoncés de la forme précédente démontrables dans T (et donc vrais car Π_1) est effectivement énumérable (par énumération effective des théorèmes). Or l'une des conséquences du théorème de Matiassevitch est que l'ensemble des énoncés vrais de cette forme n'est pas effectivement énumérable. Si l'on suppose que T est fiable pour les énoncés Σ_1 de la forme

L'équation diophantienne $P(x_1, \dots, x_n) = 0$ a une solution

on en déduit qu'il existe une infinité d'énoncés arithmétiques Π_1 qui sont indécidables dans T , et que l'ensemble de ces énoncés indécidables n'est pas effectivement énumérable.

Ce nouvel exemple d'un théorème d'incomplétude démontre l'existence d'énoncés arithmétiques Π_1 indécidables dans T si T est une extension de Q qui est Σ_1 -fiable. Il ne montre cependant pas encore comment construire un énoncé Π_1 spécifique qui soit indécidable, à la différence de la démonstration de Gödel, qui est, pour cette raison, supérieure à celles qu'on vient d'esquisser.

2. 2 Démonstration d'incomplétude par la méthode de Gödel

Voici maintenant les principales étapes d'une démonstration de TIG_1 inspirée de la méthode de Gödel.

Arithmétisation de la syntaxe

L'une des innovations majeures de la démonstration de Gödel est l'arithmétisation de la syntaxe, méthode qui requiert une théorie T qui soit une extension axiomatisable de Q . Les signes et les formules de \mathcal{L}_A sont codés par des entiers

naturels, ainsi que les preuves formelles dans T . Ce codage est tel qu'il existe une procédure effective permettant de déterminer, pour tout entier naturel, s'il est le code d'un signe, d'une formule ou d'une preuve et, le cas échéant, de quel signe, formule ou preuve il s'agit. Le code d'un énoncé E est noté $\ulcorner E \urcorner$. Sur la base de ce codage, on associe aux propriétés et relations *syntaxiques* telles que « x est une variable », « x est un axiome », etc. des propriétés ou relations *numériques* telles que « x est le code d'une variable », « x est le code d'un axiome », etc. jusqu'à parvenir, au terme d'un travail fastidieux (les définitions successives d'une longue liste de fonctions, propriétés et relations), à la définition de la relation « x est le code d'une preuve d'un énoncé dont le code est y », que l'on convient de noter $Preuve(x, y)$ (ou $Preuve_T(x, y)$ pour indiquer que sa définition dépend de T). On démontre que toutes ces propriétés et relations numériques sont décidables.

On introduit la propriété de prouvabilité, notée $Prouv(x)$, satisfaite par un entier naturel n ssi il existe un entier naturel m tel que $Preuve(m, n)$. Les entiers naturels qui la satisfont sont exactement les codes des théorèmes de T et l'ensemble des entiers naturels qui satisfont $Prouv$ est effectivement énumérable (une conséquence du théorème d'incomplétude est que cet ensemble n'est cependant pas *décidable*).

Théorème de représentation

Une propriété numérique $P(x)$ est *représentable dans* Q ssi il existe une formule $\varphi(x)$ de \mathcal{L}_A telle que, pour tout entier naturel n

si $P(n)$, alors $Q \vdash \varphi(\bar{n})$;

si non- $P(n)$, alors $Q \vdash \neg\varphi(\bar{n})$.

Cette définition est étendue aux propriétés $P(x_1, \dots, x_m)$ à m places et l'on prouve le théorème de représentation, selon lequel toute relation numérique décidable est représentable dans Q par une formule Σ_1 de \mathcal{L}_A . Au vu de l'extrême pauvreté de \mathcal{L}_A , il n'est rien moins qu'évident que toute relation calculable soit représentable par une formule de ce langage. C'est pourtant ce que Gödel réussit à prouver. Dans cette partie de la démonstration, une nouvelle innovation lui permet de montrer comment coder dans \mathcal{L}_A des suites finies quelconques d'entiers naturels, ce qui ouvre la voie d'une définition explicite dans \mathcal{L}_A de fonctions comme l'exponentielle, dont la définition naturelle est récursive. Gödel introduit pour cela la fonction numérique β , ingénieusement définie en ayant recours à des résultats connus de théorie des nombres¹¹.

11. On trouve une exposition de la fonction β par exemple in Franzén [2004], p. 25–29.

Le théorème de représentation étant acquis, on note $\text{PREUVE}(x, y)$ une formule Σ_1 de \mathcal{L}_A qui représente dans Q la relation numérique $\text{Preuve}(x, y)$. Un point essentiel de la démonstration du théorème de représentation est son caractère constructif, au sens suivant : la formule $\text{PREUVE}(x, y)$ dont le théorème démontre l'existence reflète la construction systématique de $\text{Preuve}(x, y)$ sur la base d'autres fonctions ou relations calculables. Lorsque cette condition est respectée, on dit que $\text{PREUVE}(x, y)$ est une représentation *canonique* de $\text{Preuve}(x, y)$ dans Q , étant entendu que plusieurs autres formules représentent également cette relation dans Q , qui ne satisfont pas cette condition.

L'écriture effective de $\text{PREUVE}(x, y)$ dans \mathcal{L}_A serait d'une longueur démesurée et une telle formule n'a en réalité aucune signification mathématiquement pertinente. Si $\text{PREUVE}(x, y)$ est une représentation canonique de $\text{Preuve}(x, y)$, elle présente cependant le caractère remarquable, essentiel du point de vue *logique*, que pour tous entiers naturels m, n , l'énoncé $\text{PREUVE}(\overline{m}, \overline{n})$ est vrai (dans l'interprétation standard de \mathcal{L}_A) si, et seulement si, m est le code d'une preuve de l'énoncé dont le code est n .

Nous abrègerons parfois la formule $\exists x \text{PREUVE}(x, y)$ par $\text{PROUV}(y)$. Pour tout entier naturel n , $\text{PROUV}(\overline{n})$ est vraie ssi n est le code d'un théorème de T . En ce sens, $\text{PROUV}(\overline{n})$ « dit » que n est le code d'un théorème de T . Comme $\text{PREUVE}(x, y)$ est un énoncé Σ_1 , $\text{PROUV}(y)$ est également Σ_1 .

Lemme de point fixe, ou lemme de diagonalisation

On démontre que pour toute formule $\varphi(x)$ de \mathcal{L}_A à une variable libre, il existe un énoncé E de \mathcal{L}_A tel que $Q \vdash E \leftrightarrow \varphi(\ulcorner E \urcorner)^{12}$.

Définition de G

Le lemme de point fixe est appliqué à $\neg \text{PROUV}(x)$ et l'on note G (ou G_T pour indiquer que cet énoncé est relatif à T) l'énoncé tel que

$$T \vdash G \leftrightarrow \neg \text{PROUV}(\ulcorner G \urcorner).$$

Comme $\text{PROUV}(x)$ est un énoncé Σ_1 , $\neg \text{PROUV}(x)$ est Π_1 . G est donc équivalent à un énoncé Π_1 .

Premier théorème d'incomplétude

a) Supposons que $T \vdash G$. Par le théorème de représentation, on a donc $T \vdash \text{PROUV}(\ulcorner G \urcorner)$ et, par la définition de G , $T \vdash \neg G$. T est donc inconsistent. En conséquence, si T est consistant, $T \not\vdash G$.

12. Par convention, afin de ne pas surcharger l'écriture, la représentation dans \mathcal{L}_A du code de E n'est pas notée $\overline{\ulcorner E \urcorner}$, mais simplement $\ulcorner E \urcorner$, comme le code de E lui-même.

b) Supposons que $T \vdash \neg G$. D'après la définition de G , $T \vdash \text{PROUV}(\ulcorner G \urcorner)$, et donc $T \vdash \exists x \text{PREUVE}(x, \ulcorner G \urcorner)$. Par ailleurs, comme T est consistant, $T \not\vdash G$. Aucun entier naturel n'est donc le code d'une preuve de G . D'après le théorème de représentation, pour tout entier naturel n , $T \vdash \neg \text{PREUVE}(\overline{n}, \ulcorner G \urcorner)$. Comme $T \vdash \exists x \text{PREUVE}(x, \ulcorner G \urcorner)$, T est 1-inconsistant. Par conséquent, si T est 1-consistant, $T \not\vdash \neg G$.

La démonstration historique de Gödel [1931] n'a pas recours au lemme de point fixe¹³. Gödel montre comment définir, dans le langage de l'arithmétique, un énoncé G dont la forme générale est :

Le résultat de la substitution de ... à __ dans ~~ n'est pas démontrable.

Le point clef est que le résultat de cette substitution n'est rien d'autre que l'énoncé G lui-même. G est donc équivalent à « G n'est pas démontrable ». On obtient un tel effet par exemple dans l'énoncé suivant :

Le résultat de la substitution de « Le résultat de la substitution de x à « x » dans x n'est pas démontrable » à « x » dans « Le résultat de la substitution de x à « x » dans x n'est pas démontrable » n'est pas démontrable.

G_T se distingue de cet énoncé en ce qu'il est formulé dans \mathcal{L}_A et qu'il est démontrablement équivalent, dans T , à $\neg \text{PROUV}(\ulcorner G_T \urcorner)$.

Par la construction de G , G est vrai si, et seulement si, il n'existe pas d'entier naturel n qui soit le code d'une preuve de G , c'est-à-dire si, et seulement si, G est indémontrable, à condition toutefois que l'on suive la construction de G indiquée par Gödel et que $\text{PREUVE}(x, y)$ soit une représentation canonique de $\text{Preuve}(x, y)$ ¹⁴. En conséquence, sous réserve que cette dernière condition soit satisfaite, G_T est vrai si T est consistant.

Nous ferons appel ci-dessous à un important corollaire du lemme de diagonalisation. Si T est une extension consistante de Q , il n'existe aucune formule $V(x)$ telle que pour énoncé φ , $T \vdash \varphi \leftrightarrow V(\ulcorner \varphi \urcorner)$. En effet, d'après le lemme appliqué à $\neg V(x)$, il existerait aussi un énoncé λ tel que $T \vdash \lambda \leftrightarrow \neg V(\ulcorner \lambda \urcorner)$, d'où l'on tirerait une contradiction à l'aide de $T \vdash \lambda \leftrightarrow V(\ulcorner \lambda \urcorner)$. En conséquence, le prédicat numérique $Vrai(x)$, satisfait par les codes des énoncés vrais de \mathcal{L}_A , et donc tel que pour tout énoncé φ , $Vrai(\ulcorner \varphi \urcorner)$ ssi φ , n'est pas représentable dans

13. Sur la différence entre les deux démonstrations (avec ou sans lemme de point fixe), voir les remarques éclairantes de Milne [2007], p. 193-194.

14. En effet, cette propriété n'est pas satisfaite par tout point fixe obtenu par application du lemme de diagonalisation à une formule $\neg \text{PROUV}(x)$.

T (théorème de Tarski)¹⁵. En effet, il existerait un λ tel que $T \vdash \lambda \leftrightarrow \neg \text{Vrai}(\ulcorner \lambda \urcorner)$. En outre, $\text{Vrai}(x)$ n'est pas non plus *définissable* dans \mathcal{L}_A .

3. Ce que le premier théorème dit ou ne dit pas

Il n'est pas rare qu'on attribue au premier théorème d'incomplétude un sens différent de ce que sa démonstration permet effectivement d'établir et des ouvrages comme Bouveresse [1999] ou Franzén [2005] ont le mérite de contribuer à dissiper bon nombre de confusions et erreurs d'interprétation. Par exemple, contrairement à ce qu'on trouve parfois écrit, TIG_1 n'affirme nullement que toute théorie arithmétique est incomplète – et de fait, l'arithmétique vraie est une extension complète de \mathbb{Q} (ce qui montre qu'elle n'est pas axiomatisable). TIG_1 ne dit pas non plus que toute théorie consistante et axiomatisable est incomplète – la théorie des corps réels clos, la théorie tarskienne de la géométrie euclidienne ou l'arithmétique de Presburger (formulée dans un langage muni de signes pour zéro, le successeur et l'addition, mais sans signe pour la multiplication) sont complètes. En fait, TIG_1 ne dit même pas que toute théorie arithmétique consistante et axiomatisable est incomplète – l'arithmétique élémentaire sans quantificateur, avec schémas d'axiomes pour l'addition et la multiplication est complète.

Une des confusions les plus communes consiste à penser que Gödel aurait prouvé que certains énoncés de l'arithmétique sont formellement indémontrables. Une telle allégation néglige le fait que l'énoncé G dont il est question dans la démonstration de Gödel est relatif à la théorie T considérée. Aucun énoncé d'un langage pour l'arithmétique n'est formellement indémontrable (au sens d'indémontrable dans toute théorie formelle) et si G_T est indémontrable dans T , il est trivialement démontrable dans $T + G_T$ (c'est-à-dire $T \cup \{G_T\}$). L'ordre des quantificateurs importe et TGI_1 ne dit pas qu'il existe un G indécidable dans toute théorie telle que ... , mais que pour toute théorie T telle que ... , il existe un énoncé G_T indécidable (et donc indémontrable) dans T .

Il est souvent affirmé que Gödel aurait démontré l'existence d'un énoncé vrai mais indémontrable. Admettons que l'erreur qui vient d'être signalée soit écartée et que l'on veuille dire par là « indémontrable dans T », pour une certaine théorie T , extension axiomatisable de \mathbb{Q} . L'affirmation est correcte si l'on a de bonnes raisons de penser par ailleurs que T est consistante. Ainsi, bon

15. Sur Tarski et la vérité, voir le chapitre 3 du présent volume.

nombre de mathématiciens seraient prêts à reconnaître que TIG_1 suffit à établir que G_{PA} est vrai (dans le modèle standard) bien qu'indémontrable dans PA , parce qu'ils estiment suffisamment justifiée la consistance de PA . Une affirmation similaire ne vaut certainement pas, en revanche, pour une théorie T dont la consistance n'est pas justifiée par ailleurs. De TIG_1 on ne peut déduire que G_T est vrai, seulement que G_T est vrai si T est consistante, condition qui n'est évidemment pas satisfaite en général. L'un des premiers arguments (Lucas [1961]) qui prétendaient recourir à TIG_1 pour réfuter le mécanisme en philosophie de l'esprit soutenait que l'esprit n'équivalait à aucun système formel S car G_S étant indémontrable dans S , S ne peut reconnaître la vérité de G_S , ce qui est à la portée d'un esprit humain. Ce dernier point est cependant incorrect car tout ce qu'un esprit humain peut reconnaître est que G_S est vrai si S est consistant, ce qui est, comme on le voit dans la démonstration du second théorème d'incomplétude, à portée de preuve des extensions de systèmes arithmétiques très élémentaires. Cependant, loin de réduire les anti-mécanistes à *quia*, cette réplique n'est que le premier élément d'une abondante littérature sur les usages légitimes du phénomène d'incomplétude dans un débat de philosophie de l'esprit que nous ne suivons pas plus avant ici.

Une autre erreur fréquente consiste à ne pas voir que l'incomplétude de T , dans l'énoncé du théorème, est en quelque sorte localisée dans sa partie arithmétique : l'énoncé G_T indécidable dans T est un énoncé de \mathcal{L}_A . Si l'on étend \mathcal{L}_A en un langage \mathcal{L}'_A et Q en une théorie T , rien ne permet d'établir que les axiomes adjoints à Q pour former T ne définissent pas, pris séparément, une théorie complète. L'incomplétude du théorème de Gödel est de nature *arithmétique*. Si les axiomes de T qui ne sont pas dans Q visent à la formalisation de la géométrie euclidienne, de la mécanique céleste ou de la démocratie participative, rien ne peut donc être conclu du théorème de Gödel sur l'incomplétude de ces domaines de la connaissance.

Il est certes possible qu'une théorie non arithmétique soit incomplète, et que son incomplétude se déduise du théorème de Gödel, mais uniquement dans le cas où l'arithmétique peut être interprétée dans cette théorie, au sens logique de l'interprétation d'une théorie dans une autre. L'arithmétique peut ainsi être interprétée dans la théorie axiomatique des ensembles de Zermelo-Fraenkel (ZF), qui est donc incomplète en vertu du théorème de Gödel, en dépit du fait que ses signes non logiques se réduisent au signe d'appartenance (\in) et qu'aucun signe arithmétique n'y figure. Selon une construction bien connue, zéro, le successeur, l'addition, la multiplication et l'ensemble des entiers naturels sont représentables dans ZF, et les axiomes de Q , lus au prisme de cette représenta-

tion, sont démontrables dans cette théorie. *A contrario*, la théorie des corps réels clos n'est pas incomplète, ce qui s'explique par le fait que son langage ne permet pas de définir l'ensemble des entiers naturels comme partie de l'ensemble des réels¹⁶.

L'un des objectifs du programme élaboré par Hilbert et son école dans les années 1920 était de montrer que tout énoncé arithmétique mathématiquement démontrable est démontrable sur la seule base d'une théorie arithmétique finitiste. Notons T_{AF} une telle théorie¹⁷. Il s'agissait d'établir que les extensions de T_{AF} sont conservatives au sens suivant : pour tout énoncé φ du langage de T_{AF} (typiquement, \mathcal{L}_A) et pour toute extension T de T_{AF} ,

si $T \vdash \varphi$, alors $T_{AF} \vdash \varphi$.

Or dès lors que T_{AF} est axiomatisable et consistante, TIG_1 permet justement d'affirmer qu'il existe un énoncé φ du langage de T_{AF} (à savoir $G_{T_{AF}}$) et une extension T de T_{AF} (à savoir $T_{AF} \cup \{G_{T_{AF}}\}$) tels que

$T \vdash \varphi$ et $T_{AF} \not\vdash \varphi$.

Dès lors que T_{AF} est une extension de Q , TIG_1 suffit donc à réfuter cet objectif précis du programme de Hilbert.

Selon un théorème bien connu, l'arithmétique de Peano du second ordre PA^2 (dans laquelle on dispose d'une quantification sur les *ensembles* ou les *propriétés* d'entiers naturels) est catégorique, ce qui signifie que tous ses modèles sont isomorphes¹⁸. Cela implique que tout énoncé arithmétique vrai est conséquence de PA^2 , qui est donc « complète » en un troisième sens de la complétude¹⁹. En particulier, l'énoncé de Gödel pour PA^2 , qui est vrai dans l'interprétation standard du langage de PA^2 , est vrai dans tous les modèles de PA^2 . G_{PA^2} est donc conséquence logique de PA^2 . Cette conclusion n'invalide évidemment pas TIG_1 , qui affirme que G_{PA^2} n'est pas formellement dérivable de PA^2 (si PA^2 est consistante) ; elle montre simplement que sous cette condition, il existe un

16. Cela peut sembler étrange au premier abord, mais on le comprend mieux si l'on essaie de trouver un moyen d'isoler les entiers naturels dans l'ensemble des réels.

17. Des interprétations diverses ont été proposées du terme *finitiste*. Plusieurs commentateurs s'accordent à identifier T_{AF} , au sens que lui donnait Hilbert, à l'arithmétique primitive récursive, *PRA*. Le langage de cette théorie dispose d'un signe primitif pour chaque fonction primitive récursive et la théorie elle-même comporte un schéma d'induction limité aux formules sans quantificateur non borné. Quoi qu'il en soit, T_{AF} est une extension de Q .

18. Sur la logique du second ordre, voir le chapitre 5.

19. On ne parle ici ni de la complétude syntaxique d'une théorie, ni de la complétude sémantique d'une logique (voir p. 6) mais de la complétude sémantique d'une théorie arithmétique, propriété que la théorie satisfait si tout énoncé vrai de \mathcal{L}_A en est conséquence logique.

énoncé φ tel que $PA^2 \vDash \varphi$ et $PA^2 \not\vDash \varphi$, et donc que la logique du second ordre n'est pas sémantiquement complète.

4. Le second théorème d'incomplétude

4.1 Énoncé du second théorème d'incomplétude

Soit T une extension axiomatisable de PA . La consistance de T est représentable dans T par $\neg\text{PROUV}(\ulcorner \perp \urcorner)$ ou, si \perp ne figure pas au nombre des signes logiques de \mathcal{L}_A , par une formule comme $\neg\text{PROUV}(\ulcorner 0 = 1 \urcorner)$. On notera CONS_T , ou CONS pour abrégé, l'une ou l'autre de ces formules. Exprimée dans \mathcal{L}_A , CONS_T est une formule démesurément longue dépourvue de tout intérêt mathématique ; son intérêt *logique* tient à ce qu'elle est vraie ssi T est consistante.

Il est alors aisé d'exprimer dans \mathcal{L}_A la première moitié du premier théorème d'incomplétude, par la formule $\text{CONS}_T \rightarrow \neg\text{PROUV}(\ulcorner G_T \urcorner)$, équivalente à $\text{CONS}_T \rightarrow G_T$ (puisque $T \vdash G_T \leftrightarrow \neg\text{PROUV}(G_T)$).

Si l'on réussit à montrer, en outre, que la *preuve* de cette partie de TIG_1 est elle-même formalisable dans PA , en sorte que $PA \vdash \text{CONS}_T \rightarrow \neg\text{PROUV}_T(\ulcorner G_T \urcorner)$, on en déduit que si T est consistante, alors $T \not\vDash \text{CONS}_T$. En effet, si $T \vdash \text{CONS}_T$, comme $T \vdash \text{CONS}_T \rightarrow G_T$, on a $T \vdash G_T$, ce qui contredit TIG_1 . D'où le second théorème d'incomplétude de Gödel, TIG_2 , selon lequel, si T est une extension axiomatisable de PA , alors

si T est consistante, $T \not\vDash \text{CONS}_T$.

Par ailleurs, si T est consistante, aucun entier naturel n ne code une preuve de \perp et l'on a donc, pour tout $n \in \mathbb{N}$, $T \vdash \neg\text{PREUVE}_T(\bar{n}, \perp)$. Si T est de surcroît ω -consistante, par définition de l' ω -consistance, $T \not\vDash \exists x \text{PREUVE}(x, \ulcorner \perp \urcorner)$. Or $\exists x \text{PREUVE}(x, \ulcorner \perp \urcorner)$ est équivalent à $\neg\text{CONS}$. En conséquence,

si T est ω -consistante, $T \not\vDash \neg\text{CONS}_T$.

4.2 Démonstration du second théorème d'incomplétude et corollaires

Pour démontrer TIG_2 , il suffit donc de montrer que

$$PA \vdash \text{CONS}_T \rightarrow \neg\text{PROUV}_T(\ulcorner G_T \urcorner).$$

L'exercice s'avère long et fastidieux, et la première version publiée, pour le cas où T est PA , dut attendre Hilbert et Bernays [1939], bien que le théorème ait été énoncé et informellement justifié dans Gödel [1931].

Une méthode classique consiste à montrer, comme lemme intermédiaire, que $\text{PROUV}_T(x)$ satisfait trois « conditions de prouvabilité ». Pour formuler ces conditions, nous adoptons une notation courante, par laquelle l'expression $\text{PROUV}_T(\ulcorner \varphi \urcorner)$ est abrégée en $\Box_T \varphi$, et même en $\Box \varphi$ (où le signe \Box indique à la fois le codage de φ en $\ulcorner \varphi \urcorner$ et la formule $\text{PROUV}_T(y)$ qui abrège $\exists x \text{PREUVE}_T(x, y)$). « $\Box \varphi$ » se lit donc « φ est formellement démontrable dans la théorie T (indiquée par le contexte) »²⁰.

CONS_T s'écrit alors $\neg \Box \perp$ et les trois conditions de prouvabilité sont les suivantes (pour toutes formules φ, ψ du langage de T) :

- C_1 si $T \vdash \varphi$, alors $T \vdash \Box \varphi$
- C_2 $T \vdash \Box(\varphi \rightarrow \psi) \rightarrow (\Box \varphi \rightarrow \Box \psi)$
- C_3 $T \vdash \Box \varphi \rightarrow \Box \Box \varphi$.

La partie longue et fastidieuse de la démonstration que $PA \vdash \neg \Box \perp \rightarrow \neg \Box G$ consiste à montrer que $\text{PROUV}_T(x)$ satisfait ces trois conditions. Cette démonstration requiert un schéma d'induction, de sorte que pour démontrer TIG_2 , le recours à Q ne suffit pas. (Nous ferons référence à PA bien qu'une théorie plus faible suffise, dans laquelle le schéma d'induction s'applique uniquement aux formules Σ_1).

D'autres différences entre les démonstrations de TIG_1 et TIG_2 méritent d'être soulignées. Tout d'abord, la formulation même de TIG_2 , dans laquelle figure CONS , requiert l'arithmétisation de la syntaxe, en sorte que contrairement à TIG_1 , il n'existe pas de démonstration de TIG_2 qui repose sur la théorie de la calculabilité. En outre, pour démontrer TIG_2 , il ne suffit pas de trouver un énoncé indécidable dans T ou de montrer qu'il en existe un, il faut également formaliser un raisonnement mathématique dans un système formel arithmétique, et donc non seulement démontrer l'existence d'un énoncé formalisé, mais également donner une *preuve* formelle (à savoir celle de $\text{CONS} \rightarrow \neg \text{PROUV}(\ulcorner G \urcorner)$). Enfin, il ne suffit pas d'appliquer le lemme de diagonalisation à la négation de n'importe quelle représentation de $\text{Preuve}(x, y)$ dans Q ; encore faut-il que $\text{PREUVE}(x, y)$ soit une représentation *canonique* (au sens défini p. 10) de $\text{Preuve}(x, y)$ afin que les trois conditions de prouvabilité satisfassent $\text{PROUV}(x)$.

20. En logique modale, $\Box \varphi$ se lit souvent « il est nécessaire que φ », mais en logique modale de la prouvabilité, l'expression signifie « φ est démontrable ». Voir Boolos [1993], chap. 2 et 3.

Dès lors qu'il est établi que $\text{PROUV}_T(x)$ satisfait bien les conditions de prouvabilité (pour une extension T de PA , axiomatisable et consistante), le second théorème d'incomplétude est aisément démontrable ainsi qu'une série de résultats remarquables, dont les suivants :

1. Théorème de Löb :

$$\text{si } T \vdash \Box\varphi \rightarrow \varphi, \text{ alors } T \vdash \varphi.$$

TIG_2 est une conséquence du théorème de Löb. En effet, dans le cas où φ est \perp , le théorème de Löb dit que si $T \vdash \Box\perp \rightarrow \perp$, alors $T \vdash \perp$, c'est-à-dire que si $T \vdash \neg\Box\perp$, alors $T \vdash \perp$. Or TIG_2 est la contraposée de cet énoncé. Le théorème de Löb est, par ailleurs, une conséquence de TIG_2 .

2. CONS_T est un point fixe de $\neg\text{PROUV}_T$:

$$T \vdash \text{CONS}_T \leftrightarrow \neg\text{PROUV}_T(\ulcorner \text{CONS}_T \urcorner).$$

3. Tout point fixe de $\neg\text{PROUV}_T(x)$ est équivalent à CONS_T :

$$\text{pour tout énoncé } \varphi, \text{ si } T \vdash \varphi \leftrightarrow \neg\Box\varphi, \text{ alors } T \vdash \varphi \leftrightarrow \neg\Box\perp$$

4. En particulier,

$$T \vdash G_T \leftrightarrow \text{CONS}_T$$

5. Pour tout énoncé φ ,

$$T \not\vdash \neg\Box\varphi$$

6. TIG_2 pour T est lui-même démontrable dans T :

$$T \vdash \text{CONS}_T \rightarrow \neg\Box\text{CONS}_T.$$

4.3 Ce que le second théorème d'incomplétude dit ou ne dit pas

Bien que CONS_T exprime la consistance de T , établir que $T \vdash \text{CONS}_T$ ne prouve évidemment pas que T soit consistante ; bien au contraire, si T est une extension axiomatisable de PA , d'après TIG_2 , établir que $T \vdash \text{CONS}_T$ est un moyen de démontrer que T n'est justement pas consistante. TIG_2 montre que l'on ne saurait espérer démontrer la consistance d'une extension axiomatisable de PA par des moyens formalisables dans cette extension. En particulier, si PA est consistante, CONS_{PA} est vrai et indémontrable dans PA .

Est-ce à dire que la consistance de PA est indémontrable? Nullement! Une démonstration relativement simple consiste à remarquer que tous les axiomes de PA sont vrais, que les règles d'inférence de la logique du premier ordre préservent la vérité, et à en conclure que tous les théorèmes de PA sont vrais, ce qui suffit à établir que PA est consistante²¹. TIG_2 ne montre pas que cette démonstration est incorrecte; seulement qu'elle n'est pas formalisable dans PA , dans laquelle $Vrai(x)$ n'est pas représentable.

En concluons-nous que la consistance de toute théorie mathématique T est en définitive démontrable, dans un système plus puissant que T (par exemple, trivialement, dans $T + \text{CONS}_T$)? En aucune manière. Une telle conclusion erronée résulte d'une confusion entre la démontrabilité *mathématique*, qui vise l'établissement d'une connaissance mais dont les frontières ne sont pas exactement circonscrites, et la démontrabilité *formelle*, précisément définie et relative à une théorie T , qui vise à l'exactitude des inférences sans pouvoir garantir par elle-même l'accès à aucune connaissance. La démonstration précédente de la consistance de PA est mathématiquement acceptable dès lors que l'on estime justifiée l'affirmation de la vérité des axiomes de PA , bien qu'elle ne soit évidemment pas de nature à dissiper les doutes éventuels que quiconque pourrait avoir au sujet de la consistance de PA puisqu'il y est fait appel à des principes théoriques strictement plus forts que ceux de PA . Un raisonnement similaire ne pourrait être tenu au sujet d'une théorie T dont les axiomes ne seraient pas mathématiquement justifiés ou acceptés. Quant à la dérivation de CONS_T à partir de $T + \text{CONS}_T$, elle est logiquement correcte, épistémologiquement vide, et mathématiquement inacceptable si CONS_T n'est pas déjà mathématiquement justifié. Les démonstrations mathématiques de consistance ont de tout autres formes. Leur intérêt *mathématique*, très variable selon les cas, est une question complexe qui fait l'objet de toute une littérature, sans rapport avec ce genre de dérivation triviale. Ce qu'il nous importe de noter ici est simplement que TIG_2 ne nous apprend rien sur la légitimité de telles démonstrations, dont la validité dépend de ce qui est considéré ou non comme des moyens démonstratifs mathématiquement acceptables.

La démonstration par Gödel que CONS_T n'est pas formellement dérivable de T (si T est une extension de PA , axiomatisable et consistante) a en revanche une signification mathématique majeure et des conséquences radicales pour la question du fondement des mathématiques. L'un des objectifs du programme

21. Cet argument prouve en fait bien plus que la consistance de PA puisqu'il établit sa fiabilité. La formulation mathématique de cette démonstration est plus intriquée qu'il n'y paraît, fût-ce parce qu'elle suppose une définition de « vrai », par exemple celle de Tarski.

de Hilbert (objectif distinct de celui que nous avons déjà considéré p. 14, au moment d'examiner TIG_1) était de démontrer la consistance de théories mathématiques comme l'arithmétique, l'analyse ou la théorie des ensembles (en considérant une version formalisée de ces théories), sur la base d'une métamathématique finitiste suffisamment sûre pour ne soulever aucune difficulté d'un point de vue épistémologique. Or TIG_2 montre que l'entreprise est vaine si la métamathématique finitiste à laquelle on a recours est formalisable dans la théorie dont on souhaite démontrer la consistance. En particulier, si la théorie finitiste en question est formalisable dans PA , elle ne saurait suffire à une démonstration de consistance de PA ni, *a fortiori*, de PA^2 (l'arithmétique de Peano du second ordre, qui peut servir de base à l'analyse mathématique) ou de ZF . Plus généralement, TIG_2 nous apprend qu'il est vain de chercher une preuve de consistance d'une extension axiomatisable T de PA en se limitant à des méthodes théoriques formalisables dans PA . Quant à la preuve de consistance de PA qu'on a esquissée ci-dessus, fondée sur la notion non finitaire de vérité et qui est formalisable dans un sous-système de PA^2 connu sous le nom d' ACA_0 (« *arithmetical comprehension axiom* », voir Simpson [2009]), elle ne saurait assurément pas satisfaire les exigences du programme de Hilbert, mais sera jugée acceptable par quiconque considère que les principes d' ACA_0 font partie des méthodes recevables en mathématiques, et donc par la plupart des mathématiciens.

Bien qu'on affirme souvent que TIG_2 constitue une réfutation du programme de Hilbert, Gödel eut la prudence de se garder d'une telle conclusion, au motif qu'il est parfaitement possible d'imaginer une méthode que l'on puisse considérer comme finitiste et qui ne soit pas formalisable dans « P » (le système formel utilisé dans Gödel [1931]), et *a fortiori* dans PA :

Je souhaite souligner expressément que [TIG_2] ne contredit pas le point de vue formaliste de Hilbert. Ce point de vue présuppose en effet uniquement l'existence d'une preuve de consistance dans laquelle on n'utilise rien d'autre que des moyens finitaires de preuve, et il est concevable qu'il existe des preuves finitaires qui *ne puissent pas* être exprimées dans le formalisme de P (Gödel [1931], § 4).

En 1935, Gentzen donna une preuve de consistance de l'arithmétique de Peano ayant recours à des moyens théoriques qui ne constituent pas une extension de PA (mais qui n'étaient pas pour autant formalisables dans PA). Quant à Gödel, il proposa lui-même (dans Gödel [1958]) une extension du point de vue finitiste.

Plus généralement, et indépendamment de toute considération finitiste, TIG_2 ne prouve nullement que la seule manière d'établir la consistance d'une théorie T serait d'avoir recours à une extension de T . Pour le montrer, Franzén (Franzén [2005], p. 99) donne l'exemple de la théorie $PA + \neg \text{CONS}_{PA}$, dont TIG_2 nous apprend qu'elle est consistante si PA l'est. Pour démontrer sa consistance, on voit qu'on a donc recours à une théorie qui non seulement n'en est pas une extension mais qui est même *incompatible* avec elle, à savoir $PA + \text{CONS}_{PA}$.

L'exemple de $PA + \neg \text{CONS}_{PA}$ est également éclairant sur ce qu'il est légitime, ou non, de déduire de la consistance d'une théorie. En notant T cette théorie, T est consistante si PA l'est (d'après TIG_2). De $T \vdash \neg \text{CONS}_{PA}$ on tire aisément que $T \vdash \neg \text{CONS}_T$ (Smith [2013], p. 252), d'où il appert que T , bien que consistante, n'est pas ω -consistante (voir p. 15). T est consistante sans être fiable et de fait, elle démontre l'énoncé faux $\neg \text{CONS}_T$. Enfin, bien que T soit consistante, $T + \text{CONS}_T$ est à l'évidence inconsistant. On voit donc qu'une théorie consistante peut démontrer l'énoncé faux de sa propre inconsistance et produire une théorie inconsistante par le simple ajout de l'énoncé vrai de sa propre consistance.

On aperçoit aussi, sur cet exemple, une nouvelle différence entre TIG_1 et TIG_2 . Alors que le raffinement dû à Rosser permet d'affirmer l'existence d'un énoncé G_R tel que $T \not\vdash G_R$ et $T \not\vdash \neg G_R$, si T est consistant, on ne peut faire l'économie de l'hypothèse d' ω -consistance en interrogeant la démontrabilité de $\neg \text{CONS}_T$. On a bien $T \not\vdash \neg \text{CONS}_T$ si T est ω -consistant, alors que tel n'est pas le cas pour certaines théories consistantes. $\neg \text{CONS}_T$ peut être dérivable de T si T est simplement consistant sans être ω -consistant.

5. Le phénomène d'incomplétude en mathématiques

Le phénomène d'incomplétude exclut la possibilité d'une théorie unique, complète et fiable des mathématiques tout entières, au sens d'un système maximal. Dans les termes de Gödel :

Nous cherchions un système formel pour les mathématiques et au lieu de cela nous en avons trouvé une infinité; et quel que soit le système formel qu'on choisisse dans cette infinité, il y en a un qui est plus englobant, c'est-à-dire dont les axiomes sont plus forts (Gödel [1933], p. 47²²).

22. La pagination des citations de Gödel est celle des *Collected Works*.

Si TIG_1 suffit à mettre en évidence ce phénomène, Gödel estimait que c'est TIG_2 , plus encore que TIG_1 , qui rend particulièrement manifeste cette incomplétude essentielle des mathématiques. Au sujet de TIG_2 , il écrit que

c'est ce théorème qui rend particulièrement évidente l'incomplétabilité des mathématiques. Car il rend impossible que quelqu'un conçoive un certain système bien défini d'axiomes et de règles et affirme de manière cohérente à son sujet : je perçois (avec une certitude mathématique) que tous ces axiomes et règles sont corrects, et je pense en outre qu'ils contiennent toutes les mathématiques. Si quelqu'un avance une telle affirmation, il se contredit lui-même. Car s'il perçoit que les axiomes en question sont corrects, il perçoit également (avec la même certitude) qu'ils sont consistants. Il a donc une intuition mathématique qui n'est pas dérivable de ses axiomes (Gödel [1951], p. 309).

Pourtant, pendant longtemps, il fut permis de douter que les théorèmes de Gödel, y compris TIG_2 , fussent l'indication d'une incomplétude authentiquement *mathématique*. Les énoncés indécidables que les démonstrations de TIG_1 et TIG_2 mettent en évidence peuvent certes, au moins en principe, être entièrement formulés dans \mathcal{L}_A ; mais ils n'ont aucune signification proprement mathématique en sorte qu'il est peu vraisemblable qu'aucun mathématicien eût jamais été amené à s'interroger sur leur vérité s'il n'avait trouvé des raisons *logiques* de le faire. Après la publication des théorèmes d'incomplétude, la question se posa donc de savoir s'il existait, pour des extensions axiomatisables et consistantes de théories arithmétiques comme celle de Peano, des énoncés arithmétiques indécidables qui fussent le fruit non de l'ingéniosité logicienne mais de la pratique mathématique effective.

Il fallut attendre près d'un demi-siècle pour qu'une réponse positive pût être donnée à cette question. En 1977, Paris et Harrington montrèrent qu'un certain énoncé de combinatoire finie, dont le sens est clairement mathématique et qui est à la fois formalisable dans \mathcal{L}_A et démontrable par des méthodes mathématiques communément acceptées, est néanmoins indémontrable dans PA . Une présentation simplifiée de ce résultat est donnée dans Kolata [1985], qui expose également un autre exemple, dû à Harvey Friedman, l'indécidabilité dans PA d'une version finie du théorème de Kruskal. En 1982, Paris et Kirby exhibèrent eux aussi un résultat mathématique arithmétiquement exprimable et indépendant de PA (dont une présentation aisément accessible est donnée dans Smith [2013], chap. 30).

Dans l'intervalle, Gödel avait soulevé le problème du statut épistémologique des énoncés indécidables, non en arithmétique mais en théorie des ensembles,

sur l'exemple de l'hypothèse du continu de Cantor (HC), hypothèse selon laquelle il n'existe aucun ensemble dont le cardinal est compris entre celui de \mathbb{N} et celui de \mathbb{R} . Une démonstration, due à Gödel, avait permis de conclure dès 1938 que cet énoncé était irréfutable dans *ZFC* et la preuve de son indépendance à l'égard de cette théorie fut acquise en 1963 lorsque Paul Cohen montra qu'il n'était pas non plus démontrable dans *ZFC*. Sans attendre ce résultat, Gödel exposa dès 1947 le fruit de ses réflexions sur l'interprétation et les conséquences de cette possible preuve d'indépendance. On sait que la plus grande partie des mathématiques peut être formalisée dans *ZFC* – et même, en réalité, dans un fragment de *ZFC* – et que les axiomes de cette théorie sont généralement acceptés par la communauté mathématique comme un fondement possible pour l'ensemble de la discipline. Or s'il en est ainsi, comment juger de la vérité d'un énoncé indépendant de *ZFC*, dont l'indécidabilité met en évidence l'incomplétude de cette théorie? Le problème se pose en particulier au sujet d'HC. La position que Gödel défend en 1947 est que quiconque estime que « les concepts et axiomes de la théorie classique des ensembles [...] décrivent une réalité bien déterminée » ne saurait se satisfaire du caractère indécidable d'un énoncé comme une solution définitive (Gödel [1947], p. 181), et qu'une telle indécidabilité à l'égard des axiomes de la théorie des ensembles « peut seulement signifier que ces axiomes ne contiennent pas une description complète de cette réalité » (*ibid.*). Ces axiomes ne formant pas « un système clos », Gödel suggère « leur extension par de nouveaux axiomes qui affirment l'existence de nouvelles itérations de l'opération "ensemble de" » (*ibid.*). Quoi qu'on pense de la tonalité réaliste de tels propos, ils soulèvent le problème d'une extension de *ZFC* par de nouveaux axiomes – problème qui est, depuis lors, objet d'une recherche active et fructueuse en théorie des ensembles – et posent corrélativement la question de leur justification. Le texte de Gödel de 1947 montre que l'indépendance d'énoncés bien précis de la théorie des ensembles peut tout à fait être interprétée positivement comme l'opportunité d'une extension des mathématiques au-delà de *ZFC*, plutôt que comme la mise en évidence d'un défaut ou de lacunes dans le système que forment les mathématiques formalisables dans cette théorie.

Les exemples de *PA* et de *ZF* montrent que l'incomplétude d'une théorie peut avoir des significations variées. Alors que les énoncés de \mathcal{L}_A reconnus comme indépendants de *PA* grâce aux travaux de Kirby, Paris, Harrington ou Friedman sont mathématiquement démontrables dans une théorie plus forte

que PA^{23} , il n'en va pas de même pour les énoncés du langage de ZF démontrablement indépendants de cette théorie, qui ne sont ni exprimables dans \mathcal{L}_A (ils ne relèvent donc pas du champ de TIG_1) ni démontrables dans le domaine de ce qu'on pourrait considérer comme les mathématiques « ordinaires ». Que pourrait donc être une démonstration d'un énoncé comme HC qui atteste de l'incomplétude de ZF ? Ajouter HC ou sa négation aux axiomes de ZF n'est pas une solution satisfaisante car dans l'état actuel des mathématiques, aucune raison universellement acceptée ne se présente qui puisse conduire à l'adoption de l'un ou l'autre de ces deux énoncés. La recherche de nouveaux axiomes ensemblistes qui puissent s'imposer comme acceptables – pour ne pas dire vrais – à la communauté des théoriciens des ensembles et dont l'adoption permettrait de prouver HC est active et, depuis la preuve de l'indépendance de HC en 1963, ce domaine a connu des progrès considérables, sans pour autant que l'incomplétude de ZF dont HC témoigne puisse être à ce jour résolue.

Eu égard au phénomène d'incomplétude, la différence entre des investigations de ce genre et celles d'une mathématique plus « ordinaire » mérite d'être approfondie. L'idée en est exposée dans les premières pages de Simpson [2009], où l'auteur distingue deux grandes parties des mathématiques : d'une part ce qu'il nomme les mathématiques « ensemblistes », d'autre part les mathématiques « ordinaires », par quoi il entend celles qui sont « indépendantes de, ou antérieures à, l'introduction de concepts ensemblistes abstraits », c'est-à-dire, typiquement, « des branches comme la géométrie, la théorie des nombres, le calcul infinitésimal, les équations différentielles, l'analyse réelle et complexe, l'algèbre dénombrable, la topologie des espaces métriques séparables complets, la logique mathématique et la théorie de la calculabilité » (p. 1). Si l'on accepte une telle distinction en dépit de son caractère quelque peu vague, on peut aller plus loin en ajoutant que la théorie des ensembles n'est pas un domaine auquel l'arithméticienne « ordinaire » se réfère couramment ou dont les résultats lui sont utiles (ce qui n'empêche évidemment pas que le mot « ensemble » soit pour elle d'usage courant dans son travail). Pour elle, la théorie des ensembles a tout au plus une signification fondationnelle, non un véritable usage dans sa pratique démonstrative. *A contrario*, des connaissances comme celles de l'analyse réelle et complexe font partie du bagage dont les mathématiciennes font communément usage, non seulement pour la démonstration de théorèmes d'analyse mais également en arithmétique. En prolongeant la distinction de Simpson, on peut considérer que la théorie des ensembles est,

23. Voir Smith [2013], chap. 30 pour un exemple de théorème arithmétique indémontrable dans PA et démontrable par une induction transfinie jusqu'à l'ordinal ϵ_0 .

du point de vue de ce qu'il nomme la « mathématique ordinaire », plutôt une référence en matière de rigueur démonstrative (lorsqu'un doute survient sur le sens d'une formulation, un critère généralement accepté pour la clarification de l'énoncé est sa formalisation dans le langage de la théorie des ensembles) et un horizon fondationnel (il est bon de connaître l'existence d'un fondement possible pour les mathématiques « ordinaires », fondement que l'on reconnaît comme mathématiquement acceptable même si l'on ne s'y réfère pour ainsi dire jamais dans sa pratique mathématique effective).

De ce fait, le phénomène d'incomplétude n'a pas les mêmes implications pour les mathématiques « ordinaires » et pour la théorie des ensembles. Les quelques énoncés arithmétiques qui ont un authentique sens mathématique et qui sont indécidables dans PA sont mathématiquement démontrables dans des extensions de PA qui restent mathématiquement « ordinaires » et leur vérité est donc connue par ailleurs. *A contrario*, nous n'avons pas de semblables raisons mathématiques de reconnaître comme vraie ou fausse l'hypothèse du continu de Cantor ; les mathématiques ordinaires ne nous sont sur ce point d'aucun secours. Gödel était convaincu de l'intérêt d'une recherche de nouveaux axiomes ensemblistes – dont TIG_1 montre qu'ils sont requis pour la solution de questions formalisables dans \mathcal{L}_A – sur la base d'une analyse approfondie de la signification de termes logiques ou mathématiques fondamentaux comme « ensemble », ou sur la base de ce qu'il appelle leurs « conséquences vérifiables » (Gödel [1947], p. 182). La question de savoir si les mathématiques ont besoin de nouveaux axiomes et comment ceux-ci pourraient être justifiés est cependant sujette à controverse et objet de débats²⁴.

Bien que nous sachions par TIG_1 qu'il existe des énoncés de \mathcal{L}_A indécidables dans ZFC (si ZFC est consistante), aucune des questions que les théoriciens des nombres se posent et qui demeurent à ce jour irrésolues (comme la conjecture de Goldbach ou celle des nombres premiers jumeaux²⁵) n'a pu faire l'objet d'une démonstration d'indécidabilité dans ZFC . Un objectif des recherches de Friedman (et des « mathématiques à rebours » – *reverse mathematics* – dont il est l'un des principaux promoteurs) est précisément d'élaborer une théorie mathématique élémentaire, qui relèverait sans ambiguïté de la mathématique « ordinaire » et soulèverait néanmoins des questions indécidables dans ZFC . La solution de telles questions requerrait des axiomes de grands cardinaux, censés venir compléter ceux de ZFC . L'objectif est donc de donner corps à l'idée que

24. Sur cette question, on trouvera une opposition de points de vue par exemple dans Feferman *et al.* [2000].

25. Deux nombres premiers m et n sont jumeaux si $m = n + 2$ ou $n = m + 2$. Selon la conjecture des nombres premiers jumeaux, il existe une infinité de couples de tels nombres.

les mathématiques ordinaires ne sont nullement confinées au domaine circonscrit par ZFC . Si le programme d'une telle lecture positive de l'incomplétude gödélienne était réalisé²⁶, la théoricienne des nombres aurait à faire appel, sur certaines questions qui sont de son ressort, aussi bien à des axiomes de $ZFC + A$ (où A est un ensemble approprié d'axiomes supplémentaires) ou à des résultats de théorie des ensembles qu'à des connaissances d'analyse pour la solution de problèmes arithmétiques, ce qui, incidemment, remettrait du même coup en question la distinction de Simpson entre mathématiques ensemblistes et mathématiques ordinaires. Il est aujourd'hui trop tôt pour décider des chances de succès d'une telle vision des mathématiques et de l'intégration de la théorie des ensembles dans la pratique ordinaire des théoriciens des nombres.

Qu'en est-il, enfin, du phénomène d'incomplétude mathématique dans les théories qui sont des extensions de PA strictement incluses dans ZF ? Y a-t-il un quelconque intérêt à interroger l'incomplétude de théories qui appartiennent à cet entre-deux? Les recherches menées dans le programme des mathématiques à rebours ont permis de donner une réponse positive parfaitement claire à cette question, en s'intéressant à certains fragments de PA^2 , obtenus par diverses restrictions du schéma de compréhension (voir Simpson [2009]). Ce schéma :

$$\exists X \forall n (n \in X \leftrightarrow \varphi(n))$$

affirme que pour toute formule $\varphi(n)$, il existe un ensemble X dont les éléments sont exactement les n tels que $\varphi(n)$ (où la variable X n'est pas libre dans $\varphi(n)$). Lorsque ce schéma est restreint, par exemple, aux cas où φ ne comporte aucune quantification du second ordre, on obtient le schéma de compréhension dit « arithmétique », qui définit le sous-système de PA^2 nommé ACA_0 . D'autres restrictions choisies imposées à $\varphi(n)$ définissent de nouveaux fragments de PA^2 qu'il est possible de hiérarchiser. En s'interrogeant sur le système minimal requis pour tel ou tel théorème mathématique connu (question caractéristique des *reverse mathematics*), on s'aperçoit qu'un grand nombre de théorèmes classiques de l'analyse peuvent être très précisément situés dans cette hiérarchie, en sorte qu'il est possible de déterminer, pour chacun de ces théorèmes, dans quels fragments de PA^2 il est indécidable et quel est le système minimal requis pour sa démonstration²⁷. La hiérarchie obtenue définit des degrés d'incomplétude dans une série de systèmes mathématiques. De ce fait, le phénomène d'in-

26. On trouvera une présentation de ce programme dans Friedman [2000].

27. L'ouvrage classique sur ces questions est Simpson [2009]. Stillwell [2018] en offre une approche pédagogique beaucoup plus facile d'accès.

complétude s'avère particulièrement instructif sur les relations de dépendance entre théorèmes démontrables au second ordre²⁸.

Les théorèmes dont il s'agit sont formalisables dans le langage \mathcal{L}_A^2 pour l'arithmétique du second ordre, à distinguer des théorèmes de la mathématique « ordinaire » qui sont formalisables dans \mathcal{L}_A et indécidables dans PA (et donc dans ACA_0), dont la liste demeure pour le moment extrêmement limitée (théorèmes de Paris et Harrington, Paris et Kirby, etc.). Quant au projet cher à Friedman de l'augmenter significativement par des énoncés arithmétiques démontrablement indécidables non seulement dans PA mais également dans ZFC , sa réussite reste pour le moment incertaine.

6. Incomplétude, arithmétique et entiers naturels

Les discussions philosophiques du phénomène d'incomplétude se sont souvent concentrées sur la version suivante de TIG_1 : dans tout système formel S qui inclut l'arithmétique élémentaire, il existe un énoncé G formellement indécidable que nous sommes néanmoins en mesure de reconnaître comme vrai (c'est-à-dire vrai dans le modèle standard), dès lors que nous sommes également en mesure de reconnaître S comme étant fiable. Les analyses que Dummett consacre à la « signification philosophique du théorème de Gödel » (Dummett [1963]) ne font pas exception, mais elles visent plus particulièrement les leçons que nous sommes en droit de tirer de TIG_1 touchant la signification du terme « entier naturel » et notre appréhension de ce concept. La critique de Dummett vise l'interprétation suivante de TIG_1 : ce que l'incomplétude gödélienne met en évidence, c'est l'impossibilité dans laquelle nous nous trouvons d'exprimer de manière explicite, sous la forme d'un système formel, l'intuition parfaitement claire que nous avons du concept d'entier naturel et du modèle standard de l'arithmétique ; toute tentative de caractérisation discursive (par un système axiomatique formalisé) de notre intuition du concept d'entier naturel et du modèle standard de l'arithmétique restera déficiente et c'est précisément cette déficience de nos usages possibles du concept d'entier naturel (caractérisé par des axiomes) que TIG_1 met en évidence.

Une telle interprétation du phénomène d'incomplétude est d'autant plus inacceptable, aux yeux de Dummett, qu'elle ouvre la voie à une sérieuse objection à l'un des principes de la philosophie intuitionniste qu'il défend, selon

28. On se gardera pourtant d'en conclure que certaines questions arithmétiques indécidables dans PA sont décidées dans ACA_0 ; de fait, ACA_0 est une extension *conservative* de PA .

lequel la signification des expressions d'un langage est donnée, en définitive, par leur usage, ou par la formulation de conditions d'assertabilité, et non en termes de conditions de vérité. Si l'interprétation selon laquelle aucun système formel n'est en mesure de caractériser tous les usages possibles du terme « entier naturel » était correcte, il pourrait en effet sembler légitime d'en conclure que ce n'est pas l'usage qui donne la signification de ce terme mais une intuition partagée du modèle standard de l'arithmétique, évidente à l'esprit, intuition que tout système formel demeurerait impuissant à exprimer, selon ce que nous apprend TIG_1 .

Le fond de l'erreur d'une telle interprétation, selon Dummett, réside dans l'idée selon laquelle la signification d'« entier naturel » serait donnée par l'intuition que nous avons du modèle standard de l'arithmétique, \mathcal{N} , modèle unique et que nous nous représentons comme clairement distinct des modèles non-standard. L'erreur serait à rechercher dans l'illusion dont est victime quiconque croit que ce modèle nous serait donné, dans sa singularité, par une sorte d'intuition intellectuelle, sans qu'aucune description discursive ne soit à l'œuvre. Or ce n'est pas du tout ainsi que nous est donné un modèle au sens mathématique du terme, dont la définition requiert non une intuition mais une description. Il y a donc confusion entre un modèle (mathématique) qui doit être décrit dans un métalangage et une idée de la structure des entiers naturels, dont nous sommes censés avoir une intuition claire. Le point crucial, dans l'argumentation dummettienne, est alors que tout cadre théorique dans lequel la description du modèle mathématique \mathcal{N} pourrait être donnée (par exemple celui de la théorie des ensembles) « comprend lui-même soit la notion d' "entier naturel", soit une notion équivalente ou plus forte comme celle d' "ensemble" » (*op. cit.*, p. 193). La référence à un tel cadre, inévitable si l'on doit pouvoir caractériser un authentique *modèle*, au sens mathématique du terme, ne saurait, en conséquence, avoir aucune valeur explicative pour la signification du terme « entier naturel ».

Où situer, alors, l'incomplétude révélée par TIG_1 , si elle n'est pas à chercher dans les limitations des formalismes (qu'on imagine incapables de donner forme discursive à une intuition claire)? La réponse de Dummett est qu'il faut se tourner vers le concept d'entier naturel lui-même pour trouver la solution de cette énigme. Or cela n'a rien d'évident car nous sommes persuadés d'avoir de la signification du terme « entier naturel » une conception parfaitement claire, pleine et sans reste, dépourvue de tout signe d'incomplétude. De fait, tant que nous ne considérons que l'*extension* du concept et le critère de son application, aucune difficulté ne semble se présenter, car nous sommes capables de reconnaître tout objet comme étant, ou n'étant pas, un entier naturel. En conséquence,

si la *signification* du terme n'enveloppait rien de plus que l'appartenance à l'ensemble des entiers naturels, nous serions justifiés à conclure que l'incomplétude essentielle ne réside pas du côté de cette signification mais du côté des moyens discursifs dont nous disposons pour l'exprimer.

Il en va différemment dès lors que nous comprenons que la signification d'« entier naturel » ne se réduit pas à l'extension du concept (de quelque manière qu'elle puisse être fixée) mais qu'elle dépend également des raisons que nous pouvons invoquer pour justifier une affirmation touchant la totalité des entiers naturels ; dès lors, donc, que nous comprenons qu'une signification se définit en termes de conditions d'assertabilité. Or ce que nous apprend TIG_1 est que nous ne saurions en aucune manière circonscrire l'ensemble de ces raisons ou de ces conditions. Quel que soit le système choisi pour la formalisation de l'arithmétique, il pourra être augmenté de nouveaux axiomes ou de nouvelles règles, dès lors qu'il est consistant. D'où il s'ensuit que c'est la notion même d'entier naturel dont la signification doit être considérée comme intrinsèquement vague, indéterminée, affectée d'une incomplétude essentielle.

Le point crucial, dans l'argumentation de Dummett, est le caractère suivant du concept d'entier naturel : « les entiers naturels forment une totalité à laquelle l'induction, relativement à toute propriété bien définie, est toujours applicable » ; or on sait que « le concept de propriété bien définie est indéfiniment extensible » (*op. cit.*, p. 198). La leçon principale de TIG_1 , en définitive, est que le concept d'entier naturel lui-même est intrinsèquement vague, bien que son extension ne le soit pas. Le théorème de Gödel ne révèle pas une incapacité discursive à formuler l'intégralité d'un concept dont la signification serait en elle-même parfaitement définie ; il manifeste le caractère vague de cette signification et l'indéfinie extensibilité des principes par lesquels nous sommes capables de reconnaître comme vraie une quantification universelle sur les entiers naturels.

Dummett estime par ailleurs que TIG_1 apporte aussi un nouvel argument en faveur de la philosophie intuitionniste des mathématiques, et plus précisément à la thèse selon laquelle une démonstration mathématique est une activité mentale qui ne saurait se réduire à sa représentation discursive dans un système formel. TIG_1 montre en effet qu'« aucun système formel ne pourra jamais parvenir à former le corps de tous les principes de preuve qu'il nous serait intuitivement légitime d'accepter » (*op. cit.*, p. 200) car la classe des méthodes de preuve intuitivement acceptables n'est pas susceptible d'être enfermée en un système formel. Lorsque les intuitionnistes soutiennent que « le sens des énoncés mathématiques doit être donné en termes de la notion de preuve ma-

thématique » (en termes de conditions d'assertabilité, non de conditions de vérité), ce qu'ils ont à l'esprit est « la notion intrinsèquement vague de preuve intuitivement acceptable », non celle de preuve formelle (*op. cit.*, p. 200). En définitive, TIG_1 met en lumière non seulement un caractère du concept d'entier naturel mais aussi certaines différences fondamentales entre démontrabilité mathématique et démontrabilité formelle.

Dans Isaacson [1987, p. 134–150], l'auteur s'interroge également sur les implications de TIG_1 pour le concept d'entier naturel, mais d'un tout autre point de vue, qui le conduit à distinguer plusieurs sens du terme « arithmétique ». Son point de départ est le constat que la définition d'« entier naturel » requiert une logique d'ordre supérieur. Typiquement, dans une formulation au second ordre, x est un entier naturel ssi, par définition, x possède toute propriété que possède 0 et que possède tout successeur d'un nombre qui la possède. En formule :

$$N(x) =_{\text{def}} \forall X \left(\left(X(0) \wedge \forall x \left(X(x) \rightarrow X(s(x)) \right) \right) \rightarrow X(x) \right).$$

Si PA échoue à caractériser la structure des entiers naturels²⁹, PA^2 est en revanche catégorique et sémantiquement complète (voir p. 15). Y a-t-il, dès lors, un quelconque intérêt à focaliser son attention sur PA si l'objectif est une appréhension du concept d'entier naturel ? Isaacson soutient que oui et que TIG_1 nous permet de comprendre pourquoi. Son idée est que PA traite d'une partie, et d'une partie seulement, des vérités qu'on décide habituellement de nommer « arithmétiques », qualificatif justifié par l'exprimabilité de ces vérités dans \mathcal{L}_A . Il suggère de circonscrire un sens plus resserré du terme « arithmétique », en distinguant d'une part les énoncés de \mathcal{L}_A dont la vérité ne peut pas être perçue sur la seule base de leur formulation dans \mathcal{L}_A parce que leur démonstration requiert la considération de concepts d'ordre supérieur et, d'autre part, ceux pour lesquels tel n'est pas le cas, qu'on pourra nommer « arithmétiques » en un sens plus étroit. Ce que révèle TIG_1 est que certains énoncés de \mathcal{L}_A , indémonstrables dans PA et qui ne sont donc pas « arithmétiques » au sens qu'on vient de dire, sont en fait, pour cette raison, implicitement porteurs de concepts d'ordre supérieur. Du coup, il y aurait un sens auquel PA pourrait être considérée comme *complète*, en dépit de TIG_1 , complète pour les énoncés *arithmétiques* au sens étroit. Pour Isaacson, « le système PA apparaît comme constituant le contenu purement arithmétique de notre pleine compréhension du concept

29. Comme nous l'apprend une application du théorème de compacité pour la logique du premier ordre, qui permet de montrer que PA a des modèles non standards, c'est-à-dire non isomorphes à \mathbb{N} (Van Dalen [2013]).

d'entier naturel, cette compréhension étant implicitement et intrinsèquement d'ordre supérieur. » (*op. cit.*, p. 154). Typiquement, l'énoncé de Gödel G_T ne vise pas à exprimer une propriété des entiers naturels, il n'est pas « à propos » des éléments de \mathbb{N} mais, via une méthode de codage, à propos du système T dont G_T dépend. Son contenu conceptuel est d'ordre supérieur en dépit du fait qu'il puisse être formulé dans \mathcal{L}_A . On pourrait donc dire que « de cette manière, le phénomène d'incomplétude n'est pas une incomplétude arithmétique. » (*op. cit.*, p. 158). $PA + G_{PA}$ n'est pas à proprement parler une extension arithmétique de PA mais, conceptuellement, une extension d'ordre supérieur. Pour Isaacson, « la découverte de Gödel, de l'incomplétude des systèmes formels arithmétiques, ne révèle pas tant leur faiblesse déductive que l'expressivité structurelle de l'arithmétique » (p. 159). La vérité des énoncés de \mathcal{L}_A qui sont en réalité au sujet d'autres choses ne dépend pas uniquement des principes qui régissent notre compréhension du concept d'entier naturel tel qu'elle est inscrite dans PA ; elle dépend également de principes d'ordre supérieur.

Une des difficultés auxquelles la proposition d'Isaacson doit faire face est celle d'une définition précise du sens étroit du terme « arithmétique » qui est ainsi suggéré. Que dire par exemple des énoncés de \mathcal{L}_A dont la signification n'est perceptible, pour nous, qu'au travers du codage gödélien, que nous devrions, à ce titre, considérer comme ayant un contenu d'ordre supérieur et qui sont pourtant démontrable dans PA ? Tel est le cas, par exemple, de l'énoncé qui exprime la consistance d'une partie finie quelconque de PA , qui est de fait démontrable dans PA . On sait par ailleurs, par un phénomène de « *speed-up* » mis en lumière par Gödel dans son bref texte sur la longueur des preuves³⁰ – et Isaacson n'ignore pas cette objection possible – que certains énoncés démontrables au second ordre le sont également en principe au premier, mais uniquement par des preuves d'une longueur si démesurée qu'elles en deviennent absolument illisibles et, dans les faits, parfaitement incompréhensibles. Il existe du reste des cas concrets d'un tel phénomène, comme le théorème de Dirichlet (selon lequel si a et k sont premiers entre eux, il existe une infinité de nombres premiers de la forme $kn + a$) dont la preuve originale fait appel aux logarithmes, aux intégrales et aux nombres complexes, et dont on sait qu'il pourrait être démontré dans PA au moyen d'une preuve qui serait probablement si longue qu'elle serait illisible³¹. Qu'en est-il de tels énoncés eu égard à la distinction introduite par

30. Gödel [1936] montre que pour toute fonction calculable f il existe une infinité d'énoncés φ de \mathcal{L}_A démontrables dans PA et PA^2 tels que la longueur de la plus courte preuve de φ dans PA soit strictement supérieure à $f(i)$, où i est la longueur de la plus courte preuve de φ dans PA^2 . Ce résultat montre que pour une infinité de formules de \mathcal{L}_A , le passage de PA à PA^2 permet de réduire drastiquement la longueur de leur preuve.

31. Sur cet exemple et d'autres cas apparentés, voir Franzén [2004], chap. 3.

Isaacson? Plus généralement, l'idée que les énoncés de \mathcal{L}_A indécidables dans PA contiennent des concepts de second ordre sous une forme pour ainsi dissimulée (et que TIG_1 permet justement de débusquer) donne du phénomène d'incomplétude une interprétation qui ne manque certes pas d'intérêt mais qui dépend d'un sens resserré du mot « arithmétique » qu'il n'est pas aisé de rendre précis.

7. Les remarques de Wittgenstein sur l'incomplétude

Les paragraphes des *Remarques sur les fondements des mathématiques* dans lesquels Wittgenstein commente les théorèmes d'incomplétude³² ont généralement suscité des réactions pour le moins réservées, voire franchement hostiles. Dans sa correspondance avec Karl Menger, Gödel écrit en 1972 qu'au vu des passages cités par Menger, il est clair que Wittgenstein « n'a pas compris » son « théorème sur les propositions indécidables », ou du moins qu'il « a fait comme s'il ne le comprenait pas », l'interprétant « comme une sorte de paradoxe logique alors qu'en fait c'est exactement le contraire, à savoir un théorème mathématique d'une partie absolument incontestée des mathématiques » (Gödel [2003], p. 133). En 1973, dans sa correspondance avec Abraham Robinson, Gödel ne se dément pas lorsqu'il ajoute que Wittgenstein « avance une interprétation erronée, complètement triviale et sans intérêt » (*op. cit.*, p. 201) de son théorème d'incomplétude. Les comptes rendus rédigés par Anderson, Bernays, Dummett ou Kreisel, parus à l'époque de la publication des *Remarques*, ne sont pas plus amènes. Pourtant, à l'encontre de ce flot de critiques par de fins connaisseurs des résultats de Gödel (mais assurément moins de la pensée de Wittgenstein), certains comme Shanker, Putnam ou Floyd portent à rebours un jugement positif sur les paragraphes incriminés³³. Sans chercher à rendre compte des abondants commentaires exégétiques touchant les pensées de Wittgenstein sur Gödel, nous soulignerons ici certaines difficultés et avancerons quelques tentatives d'éclaircissement, en essayant de comprendre pourquoi les remarques relatives à l'incomplétude ont pu susciter tant d'incompréhension.

Que Wittgenstein [1956] soit constitué d'un choix de fragments non destinés à la publication ne suffit pas à expliquer le caractère déconcertant de certaines des affirmations qu'on y trouve (notamment des formulations comme « Ma

32. Certains datent de 1937-1938, d'autres de 1941-1944. Les *Remarques* ont paru en 1956. Voir en particulier p. 111-118 et p. 308-311. La pagination est celle de la traduction française.

33. Voir par exemple Shanker [1988] et Floyd et Putnam [2000]. On trouvera les références des comptes rendus d'Anderson, Bernays, Dummett et Kreisel dans Rodych [1999], p. 200.

tâche n'est pas de parler de la démonstration de Gödel, par exemple, mais de l'esquiver dans mon discours. ») dont on ne peut guère espérer rendre compte hors contexte et sans autre connaissance de la pensée wittgensteinienne des mathématiques en général. En particulier, le lecteur des *Remarques* s'égarerait rapidement s'il cherchait à placer les notes de Wittgenstein relatives à l'incomplétude sur le même plan que la plupart des textes des autres commentateurs. Wittgenstein écrit par exemple : « Aussi étrange que cela paraisse, ma tâche concernant le théorème de Gödel consiste uniquement à mettre au clair ce que signifie en mathématiques une proposition comme "en admettant que l'on puisse prouver cela". » (p. 311). Parce que ses objectifs sont autres, les remarques de Wittgenstein suscitent souvent une impression de décalage. Ainsi, tandis que les commentateurs s'interrogent fréquemment sur les implications de TIG_1 et TIG_2 pour le programme de Hilbert, lorsque Wittgenstein traite de cette question, il s'agit plutôt de mettre en évidence ce qu'il y a de profondément erroné dans le programme de Hilbert lui-même et dans la conception des mathématiques qu'il présuppose, comme dans l'idée, jugée confuse, selon laquelle une recherche mathématique serait susceptible d'apaiser les interrogations philosophiques relatives à la question des fondements.

À l'encontre des interprétations qui font de TIG_1 et TIG_2 des énoncés spéciaux, qui auraient une signification différente de celle des mathématiques ordinaires, Wittgenstein replace ces théorèmes dans le contexte plus général des preuves d'impossibilité en se demandant ce que nous apprennent de telles preuves, comme celle « qui montre que telle construction ne peut être réalisée avec la règle et le compas » (p. 115). Le risque, avec TIG_1 et TIG_2 , est de se laisser abuser par l'idée mystifiante d'énoncés vrais qui resteraient inaccessibles à la démonstration, et cela pour n'importe quel système formel satisfaisant telle et telle conditions. On donne une idée différente de TIG_1 si on le comprend par analogie avec le cas de la géométrie où une preuve d'impossibilité contient un « élément de prédiction » qui conduit à une maxime du genre suivant : « Ne te fatigue pas à trouver une construction (de la trisection de l'angle par exemple³⁴) – on peut prouver que ça ne marche pas. » (*ibid.*). TIG_1 et TIG_2 ne dévoilent pas d'étranges vérités inaccessibles à la preuve ; leur signification peut être plus simplement comprise comme l'indication de certaines directions dans lesquelles il est tout simplement inutile de mener des recherches et donc comme une clarification des termes dans lesquels le problème de l'existence d'une preuve est posé.

34. Wittgenstein sous-entend ici : avec la règle et le compas.

Une des principales raisons du puissant pouvoir mystificateur des théorèmes d'incomplétude tient dans l'idée que G_T et CONS_T seraient *vrais* ; non pas vrais dans tel ou tel système logique – par exemple celui de Russell – mais vrais *simpliciter*. À l'encontre de ce genre de croyance, qu'on ne trouve pas seulement sous la plume des commentateurs, mais également au § 1 de Gödel [1931], Wittgenstein soutient que la vérité mathématique n'a pas de sens en dehors d'une certaine pratique mathématique. Il n'y a pas de vérité mathématique en soi, indépendamment de l'adoption de règles qui définissent un jeu de langage particulier, susceptible d'être abandonné au profit d'un autre :

Tout comme nous demandons : « "démontrable" dans quel système ? » nous devons aussi demander : « "vrai" dans quel système ? » "Vrai dans le système russellien" signifie comme nous l'avons dit : démontré dans le système russellien ; et "faux dans le système de Russell" signifie que l'on a démontré le contraire dans le système de Russell (Wittgenstein [1956], p. 114).

Que répondre, alors, à quelqu'un qui demanderait conseil au sujet d'une certaine proposition P qui, construite en symboles russelliens, pourrait être interprétée par « P n'est pas démontrable dans le système russellien » (p. 113) ; quelqu'un qui, raisonnant à partir de cette interprétation, se trouverait acculé à conclure que P est vraie et indémontrable ? Comment se sortir de ce mauvais pas si « vrai » signifie toujours « vrai dans un certain système S », caractérisé par des règles, et si « vrai dans S » n'a pas d'autre sens que « démontrable dans S » ? Le conseil de Wittgenstein est clair : renoncer à l'interprétation selon laquelle P se traduit en français par « P est indémontrable » (p. 114). Notons qu'il s'agit là d'un *conseil*, et donc d'une voie qu'il semble recommandable de suivre. Si « vrai » signifie « vrai dans le système de Russell », et que ta conclusion est que P est à la fois vrai en ce sens et indémontrable dans le même système, tu feras bien de renoncer à l'interprétation initiale de P . Si en revanche « vrai » signifie « vrai dans un autre système », la difficulté se dénoue, car « ce qui signifie "perdre" aux échecs, peut déterminer la victoire dans un autre jeu » (*ibid.*).

Le paragraphe où ce conseil est prodigué (Annexe III, § 8), objet de maints commentaires, est aussi aisément source de malentendus. D'un côté, en effet, on aura beau jeu de soupçonner Wittgenstein d'avoir tout simplement ignoré la preuve de TIG_1 , de s'en être tenu au § 1 de Gödel [1931], où l'aperçu de la preuve suit une voie sémantique qui n'est pas du tout celle qu'emprunte Gödel dans la suite du texte ; Wittgenstein se serait fourvoyé en pensant que la preuve de TIG_1 dépendait de l'interprétation de G_T en langue naturelle par « G_T n'est pas démontrable dans T ». D'un autre côté, selon une autre lecture,

plus charitable, Wittgenstein ne commettrait aucune des erreurs qui lui sont imputées, et ses détracteurs auraient simplement tort de vouloir lire ses remarques à la lumière de celles de la plupart des interprètes du phénomène d'incomplétude. En réalité, son objectif, explicitement autre, n'est pas de « parler de la démonstration de Gödel » (ce qui ne signifie pas qu'il l'ignore) mais de « l'esquiver », en indiquant les voies par lesquelles dénouer les problèmes de langage sur lesquels viennent si souvent buter les philosophes, pensant avoir affaire à d'authentiques difficultés théoriques. Or voilà exactement ce qui se passe lorsqu'ils croient devoir s'étonner de l'apparente énigme d'une vérité indémontrable, sans voir que la vérité dont ils parlent n'a de sens que dans un jeu de langage.

La lecture de Wittgenstein [1956] est compliquée par le fait que le texte ouvre aisément la voie à de multiples lectures, tel ce passage du même § 8 où Wittgenstein, raisonnant sur la vérité ou la fausseté de la proposition P interprétée par « P est indémontrable », écrit :

Que signifie, maintenant, ton « mettons qu'elle soit fausse » ? *Au sens russellien*, cela signifie : "mettons que le contraire soit prouvé dans le système russellien" ; *si c'est là ton présupposé*, tu renonceras à l'interprétation selon laquelle elle est indémontrable (Wittgenstein [1956], p. 114).

Faut-il lire ce passage comme un commentaire précis et rigoureux de la seconde moitié de TIG_1 , selon laquelle si $\neg G_T$ ($\neg P$ dans la notation de Wittgenstein) est démontrable dans T , alors T est ω -inconsistant (voir p. 15) ? Telle est la suggestion de Floyd et Putnam [2000], qui notent que si T est ω -inconsistant, l'interprétation standard de \mathcal{L}_A n'est pas modèle de T , en sorte qu'une autre interprétation de \mathcal{L}_A est requise. D'où la conséquence que tire Wittgenstein : tu renonceras à l'interprétation initiale de P . Un lecteur moins charitable que Floyd et Putnam notera que Wittgenstein, de fait, ne parle jamais d' ω -consistance, seulement de consistance simple, et qu'on ignore s'il avait connaissance du théorème de Gödel-Rosser, qui raffine la preuve de Gödel en montrant que l'hypothèse d' ω -consistance (qui figure dans l'énoncé original de Gödel) peut être affaiblie en une hypothèse de consistance simple (voir p. 4).

Le phénomène d'incomplétude est ordinairement perçu comme l'indication d'une limitation des formalismes, qui se révéleraient incapables de saisir l'en soi de la structure des entiers naturels, ou de donner une expression discursive de l'intuition que nous pourrions en avoir. Wittgenstein s'inscrit en faux contre une telle interprétation, qui fait des théorèmes de Gödel l'instrument d'une révélation. Sur ce point, ses *Remarques sur les fondements des mathématiques* sont

clairement l'expression d'une position déflationniste à l'égard de TIG_1 et TIG_2 . L'illusion est en fait présente à la source, dans l'idée même d'une représentation formalisée des mathématiques. Loin de nous éclairer sur la nature des mathématiques, la formalisation nous en fait précisément sortir. Dès lors que les signes sont privés de leur sens, un autre jeu est engagé. De ce fait, la métamathématique elle-même, croyant parler des mathématiques dans une version idéalisée, parle en réalité d'un autre objet. Dans les *Remarques philosophiques*, Wittgenstein écrit en ce sens : « En mathématiques on ne peut parler en général de systèmes, mais on peut seulement parler à l'intérieur de systèmes. [...] Il ne peut y avoir de métamathématique en aucun sens essentiel » (Wittgenstein [1964], § 152–153). Les théorèmes d'incomplétude ne montrent en aucune manière qu'il y aurait quelque chose d'incomplet dans les mathématiques elles-mêmes. En effet « la mathématique ne peut pas être incomplète; tout comme un *sens* ne peut pas être incomplet » (*op. cit.*, § 158), remarque qui ne contredit en rien le fait que des systèmes formels – qui jouent un autre jeu que celui des mathématiques – puissent être incomplets. Pour Wittgenstein, juger que TIG_1 et TIG_2 seraient de nature à nous apprendre quoi que ce soit de profond sur les mathématiques et leur fondement relève de la pure et simple illusion. Sur ce point comme sur d'autres, ses remarques vont manifestement à l'encontre du point de vue adopté dans la plupart des commentaires, suscitant souvent critiques et incompréhension.

8. Les mathématiques sont-elles la syntaxe du langage ?

Gödel n'a pas seulement démontré quelques-uns des plus célèbres théorèmes de l'histoire de la logique, il a également défendu avec persévérance et détermination une forme de réalisme mathématique dont l'exposé occupe une série de textes et conférences. Ce réalisme est-il d'une manière ou d'une autre en rapport avec les théorèmes d'incomplétude ? On distingue généralement assez bien les articles scientifiques et les textes dont la teneur est historique ou philosophique et il est souvent difficile d'établir une relation de conséquence précise et convaincante entre un résultat mathématique et une thèse comme le réalisme. Gödel s'est pourtant efforcé de faire usage de TIG_2 pour réfuter le déflationnisme mathématique défendu par quelques-uns des auteurs qu'on a coutume de regrouper sous l'étendard du positivisme logique. Sollicité pour composer une section du volume de la *Library of Living Philosophers* consacré à Carnap, Gödel a rédigé, pour cet ouvrage, pas moins de six versions différentes d'un texte intitulé « Is mathematics syntax of language ? » dans lequel il

a recours à TIG_2 dans l'un des arguments qu'il déploie pour établir la fausseté des idées de Carnap relatives à la nature des mathématiques. En définitive, Gödel ne se résolut pourtant jamais à envoyer son article à l'éditeur et il fallut attendre la parution du troisième volume des *Collected Works*, qui regroupe les inédits, pour que les versions III et V du texte soient rendues accessibles au public (Gödel [1995b]) et la parution de Gödel [1995c] pour les versions II et VI.

Les auteurs visés par la critique de Gödel (comme Hans Hahn ou Rudolf Carnap) ont défendu à l'égard des mathématiques des positions philosophiques non seulement différentes entre elles mais également en évolution rapide au cours du temps, surtout au début des années trente, période à laquelle Gödel se réfère au début la version III. Il n'est cependant pas ici nécessaire d'introduire de subtils *distinguos* ni d'entrer dans l'exégèse des textes car l'idée fondamentale à laquelle Gödel s'attaque peut être exposée en quelques mots. Elle est défendue par certains empiristes logiques comme Carnap, qui cherche à montrer qu'il est possible de reconstruire la science dans un système linguistique défini par des règles, en respectant, dans la reconstruction, une stricte distinction entre des propositions de la « science formelle » (mathématique et logique) qui soient vides de tout contenu empirique et des propositions propres à la « science du réel » dont le contenu, empirique, puisse être fondé sur leur rapport à l'expérience et la possibilité d'une vérification ou d'une confirmation.

C'est une telle conception des mathématiques comme vides de contenu et sans objet que Gödel cherche à réfuter ; plus précisément, une conception « syntaxique » caractérisée par l'idée que l'intuition mathématique par laquelle nous garantissons la vérité des axiomes pourrait être remplacée par des règles syntaxiques relatives à l'usage des signes. Pour Gödel, les énoncés mathématiques ont assurément un contenu, conceptuel, qui n'est pas du tout reconnu dans la conception des mathématiques comme syntaxe du langage. Les arguments qu'il avance contre cette conception ne présupposent cependant pas l'existence d'un tel contenu ; ils sont de nature réfutative et l'un d'eux (le seul qui nous intéresse ici) s'appuie sur TIG_2 . Gödel estime qu'on ne saurait prétendre fonder les mathématiques sur des règles syntaxiques à valeur conventionnelle sans apporter la preuve que parmi les conséquences de ces règles ne figure aucun énoncé à contenu empirique. Si un énoncé empirique est conséquence des règles logico-mathématiques, celles-ci ne sont en effet pas purement syntaxiques et elles n'ont donc pas valeur de simples conventions. Or prouver que certains énoncés du langage ne sont pas conséquences des principes mathématiques adoptés dans un système linguistique, c'est donner une preuve de consistance de ce dernier. Après avoir argumenté pour montrer que les ma-

thématiques requises dans un système pour la reconstruction rationnelle de la science satisfont bien les conditions d'application de TIG_2 , Gödel a recours à ce théorème pour montrer que les principes en question n'ont pas les ressources démonstratives suffisantes pour établir leur propre consistance ; ils n'apportent donc pas les preuves requises qu'ils sont de pures conventions. Nous savons en effet par TIG_2 que la preuve de consistance d'un système formel mathématique requiert des méthodes démonstratives qui ne sont pas formalisables dans ce système, dès lors que celui-ci enveloppe une partie suffisante de l'arithmétique élémentaire, ce qui est le cas d'un système linguistique pour la reconstruction rationnelle de la science.

Ces quelques lignes ne prétendent pas rendre compte de toute l'argumentation de Gödel, longue et passablement intriquée, contre ce qu'il appelle la « conception syntaxique des mathématiques » ; elles ne retiennent que l'usage qu'il entend faire de TIG_2 comme instrument de réfutation d'une thèse philosophique. Il n'est pas évident de comprendre pourquoi, après avoir rédigé six versions de son texte, Gödel ne put se résoudre à en faire paraître aucune mais plusieurs commentateurs (notamment Awodey et Carus [2004, p. 203-223]) ont imaginé ce que Carnap eût pu répondre à Gödel s'il avait eu connaissance de ses arguments critiques. L'un des points essentiels est que Gödel ne semble pas percevoir la radicalité de la position de Carnap, qui renonce purement et simplement à toute ambition de justification et de fondement des mathématiques, en sorte qu'il nierait simplement le point de départ de l'objection, à savoir que les principes des mathématiques, eu égard à l'usage qu'ils peuvent avoir dans un système linguistique pour la reconstruction de la science, devraient être justifiés par un argument qui implique une preuve de consistance. En renonçant à faire appel à l'intuition mathématique comme garante de la vérité des axiomes au profit d'un système de règles, Carnap renonce à l'idée de la description vraie d'un domaine d'objets mathématiques et même à toute exigence de justification du système de règles adopté. La science formelle n'est rien de plus qu'un système de signes auxiliaires régi par des règles utiles d'un point de vue pratique mais qui n'appelle aucune entreprise de fondation. On peut certes chercher à fonder les mathématiques sur une base logique (en un sens large du mot « logique », qui inclut par exemple une théorie des types) et parler de validité au sens d'une conformité à la conception classique des mathématiques, mais non de propositions mathématiques qui seraient en soi correctes ou vraies. Plus généralement, un système linguistique pour la reconstruction rationnelle de la science peut être commode, plus ou moins approprié aux fins qu'on s'est prescrites, mais non correct en soi. Au regard d'une telle conception

pragmatique, Carnap ne reconnaîtrait donc pas l'exigence de justification sur laquelle repose l'argument de Gödel.

Pour autant, le phénomène d'incomplétude n'est pas sans conséquence pour le projet philosophique d'une science unitaire tel qu'il était conçu par Carnap et d'autres empiristes logiques à l'époque de la *Syntaxe logique du langage*, où l'unité de la science est avant tout celle du langage de la science. Y a-t-il encore un sens, au regard de TIG_1 et de l'incomplétude qu'il implique, à vouloir persévérer dans l'idée d'un langage unitaire de la science? S'il demeure quelques rares traces d'un tel projet dans Carnap [1934]³⁵, Carnap tire clairement les conséquences du phénomène d'incomplétude arithmétique, lorsqu'il affirme l'impossibilité de construire aucun système linguistique pour la science dans lequel tout concept mathématique soit définissable et toute proposition mathématique vraie démontrable. D'où il conclut que « la mathématique requiert une série infinie de langages toujours plus riches³⁶ ». Pourtant, Carnap est bien en quête d'un « critère complet de validité pour les mathématiques » (Carnap [1937], § 34a) et d'une définition de la validité applicable, très exactement, « aux propositions valides (vraies, correctes), lorsqu'on prend comme base la logique et les mathématiques classiques » (*ibid.*). Comme d'autres logiciens en 1931, en prenant connaissance de TIG_1 , Carnap se mit en quête d'un moyen de réaliser malgré tout l'idéal d'un formalisme mathématiquement complet, en ayant recours à un système qui ne satisfît pas certaines des conditions d'application de TIG_1 et permit ainsi d'échapper aux conséquences du phénomène d'incomplétude. La solution qu'il expose dans la *Syntaxe logique du langage* est celle de langages L munis d'une relation de conséquence non finitaire, définie par exemple (d'autres méthodes sont également envisagées) par une règle dans laquelle le nombre de prémisses est infini. Carnap parvient ainsi, pour les deux langages qu'il expose dans Carnap [1934], à un théorème de complétude selon lequel « toute proposition logique³⁷ est soit analytique soit contradictoire » (§ 34). Tous les énoncés mathématiques qui sont *conséquences* (en ce sens infini) des axiomes de L ne sont certes pas *démonstrables* dans L ; mais selon une certaine définition d'« analytique dans L », qui repose sur une relation de conséquence du genre indiqué, tous ces énoncés sont *analytiques dans L* (sans être nécessairement *théorèmes* de L), ce qui permet à Carnap de formuler un

35. Au § 18, par exemple, où Carnap envisage de montrer qu'il est possible de « se tirer d'affaire avec un seul langage ».

36. Ce que Carnap entend ici par « langage » comporte en fait axiomes et règles d'inférence. On pourra comparer avec la citation de Gödel donnée p. 20.

37. Dans le vocabulaire de Carnap, cela inclut les mathématiques.

théorème de complétude relatif à cette notion d'analyticité, pour les deux langages exposés à titre d'exemples dans Carnap [1934].

Cette complétude carnapienne laisse néanmoins intacte une autre forme d'incomplétude, qui explique que Carnap reconnaisse la nécessité d'une « série infinie de langages toujours plus riches », en raison de deux défauts inhérents aux systèmes arithmétiques susceptibles d'être utilisés dans un langage pour la reconstruction rationnelle de la science : pour de tels systèmes S , il est toujours possible de spécifier d'une part des termes indéfinissables dans S (on pense par exemple à « analytique dans S », dont la définition requiert un métalangage essentiellement plus riche que S) et d'autre part (d'après TIG_1) des propositions indécidables dans S (comme l'énoncé de Gödel G_S). Que reste-t-il, alors, de la thèse de l'unité de la science, comprise comme celle du langage de la science, et quel sens y a-t-il à poursuivre la quête d'un langage unitaire pour la science après la mise au jour du phénomène d'incomplétude ? L'unité de la science du réel soulève des problèmes séparés, dont nous ne traitons pas ici. Quant à la science formelle, si Carnap ne vise pas pour elle une unité intégrale, ce n'est pas uniquement parce que dans les langages L auxquels il pense, « analytique dans L », ainsi que d'autres termes méthodologiques similaires, ne sont pas définissables dans L ; c'est aussi parce que dans bien des situations, on ne saurait se satisfaire de la complétude carnapienne fondée sur des méthodes infinitaires, et qu'on demande d'avoir accès à une *démonstration* des propositions mathématiques requises par la science du réel. Or ce que nous apprend le phénomène d'incomplétude gödelien est que dans le cas d'un système formel au sens usuel du terme (sans les méthodes infinitaires de Carnap [1934]), il est impossible de déterminer un ensemble complet d'axiomes pour les mathématiques, en sorte que les axiomes requis par la science dans son ensemble sont toujours susceptibles d'exiger le remplacement d'un système linguistique L provisoirement adopté par un autre, plus fort, lorsqu'un énoncé mathématique, arithmétique ou autre, s'avère indépendant de L , sans qu'il soit jamais possible d'assigner aucune limite au processus par lequel L pourra devoir être renforcé. Bien que la découverte d'énoncés mathématiques indépendants des théories en usage puisse être considérée comme un événement relativement exceptionnel à l'échelle de l'histoire, elle n'en justifie pas moins l'idée d'une série potentiellement infinie de langages mathématiques susceptibles d'être requis par les progrès futurs de la science, conséquence du phénomène d'incomplétude que Carnap prend en considération dans sa conception de l'unité de la science.

Indications bibliographiques. Nombreux sont les ouvrages qui exposent une démonstration des théorèmes d'incomplétude, comme Smith [2013], Franzén [2004] ou le classique Smullyan [1992]. Sur le phénomène d'incomplétude en général, on trouvera une bibliographie substantielle in Raatikainen [2015]. Raatikainen [2005] donne un aperçu de quelques-unes de ses implications philosophiques.

Bibliographie

- S. AWODEY et A. CARUS : How Carnap could have replied to Gödel. In S. AWODEY et C. KLEIN, édés : *Carnap Brought Home. The View From Jena*. Open Court, Chicago, 2004, p. 203-223.
- G. BOLOS : *The Logic of Provability*. Cambridge University Press, Cambridge, 1993.
- J. BOUVERESSE : *Prodiges et vertiges de l'analogie*. Raisons d'agir, Paris, 1999.
- R. CARNAP : *Logische Syntax der Sprache*. Julius Springer, Vienne, 1934.
- R. CARNAP : *The Logical Syntax of Language*. Open Court, Chicago and La Salle, 1937. Trad. angl. augmentée de Carnap [1934].
- M. DUMMETT : The philosophical significance of Gödel's theorem. *Ratio*, 5:140-155, 1963. Reproduit in M. Dummett, *Truth and Other Enigmas*, Duckworth, Londres, 1978, p. 186-201.
- S. FEFERMAN, H. M. FRIEDMAN, P. MADDY et J. STEEL : Does mathematics need new axioms? *Bulletin of Symbolic Logic*, 6:401-446, 2000.
- J. FLOYD et H. PUTNAM : A note on Wittgenstein's "notorious paragraph" about the Gödel theorem. *The Journal of Philosophy*, XCVII:624-632, 2000.
- T. FRANZÉN : *Inexhaustibility. A Non-Exhaustive Treatment*. A K Peters, Wellesley, 2004.
- T. FRANZÉN : *Gödel's Theorem. An Incomplete Guide to its Use and Abuse*. A K Peters, Natick, 2005.
- H. M. FRIEDMAN : Normal mathematics will need new axioms. *Bulletin of Symbolic Logic*, 6:434-446, 2000.

- K. GÖDEL : Die Vollständigkeit der Axiome des logischen Funktionenkalküls. *Monatshefte für Mathematik und Physik*, 37:349–360, 1930. Reproduit et traduit in Gödel [1986].
- K. GÖDEL : Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38:173–198, 1931. Reproduit et traduit in Gödel [1986].
- K. GÖDEL : The present situation in the foundations of mathematics. 1933. Publié in Gödel [1995].
- K. GÖDEL : On undecidable propositions of formal mathematical systems. Mimeographed lecture notes, taken by S. C. Kleene and J. B. Rosser. 1934. Publié in Gödel [1986].
- K. GÖDEL : Über die Länge von Beweisen. *Ergebnisse eines mathematischen Kolloquium*, 7:23–24, 1936. Reproduit et traduit in Gödel [1986].
- K. GÖDEL : What is Cantor's continuum problem. *American Mathematical Monthly*, 54:515–525, 1947. Reproduit in Gödel [1990].
- K. GÖDEL : Some basic theorems on the foundations of mathematics and their implications. 1951. Publié in Gödel [1995a].
- K. GÖDEL : Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes. *Dialectica*, 12:280–287, 1958. Reproduit et traduit in Gödel [1990].
- K. GÖDEL : *Collected Works, Vol. I, Publications 1929 - 1936*. Oxford University Press, New York et Oxford, 1986.
- K. GÖDEL : *Collected Works, Vol. II, Publications 1938 - 1974*. Oxford University Press, New York et Oxford, 1990.
- K. GÖDEL : *Collected Works, Vol. III, Unpublished Essays and Lectures*. Oxford University Press, New York et Oxford, 1995a.
- K. GÖDEL : Is mathematics syntax of language? 1995b. Publié in Gödel [1995a].
- K. GÖDEL : *Unpublished Philosophical Essays*. Birkhäuser, Bâle, 1995c. éd. par F. Rodríguez-Consuegra.
- K. GÖDEL : *Collected Works, Vol. V, Correspondence H-Z*. Oxford University Press, New York et Oxford, 2003.
- G. HEINZMANN : Wittgenstein et le théorème de Gödel. In J. SEBESTIK et A. SOULEZ, eds : *Wittgenstein et la philosophie aujourd'hui*. Méridien Klincksieck, Paris, 1992.
- D. HILBERT et P. BERNAYS : *Grundlagen der Mathematik, vol. II*. Springer, Berlin, 1939. Trad. fr. de M. Guillaume et F. Gaillard, *Fondements des mathématiques*, vol. 2, L'Harmattan, Paris, 2001.

- D. ISAACSON : Arithmetical truth and hidden higher-order concepts. In The Paris GROUP, éd. : *Logic Colloquium 85*. North-Holland, Amsterdam, 1987, p. 134–150.
- G. KOLATA : Does Gödel's theorem matter to mathematics? In L. A. HARRINGTON, éd. : *Harvey Friedman's Research on the Foundations of Mathematics*. Elsevier, New York, 1985.
- J. R. LUCAS : Minds, machines and Gödel. *Philosophy*, 36:112–127, 1961.
- P. MILNE : On Gödel sentences and what they say. *Philosophia Mathematica*, 15-2:193–226, 2007.
- P. RAATIKAINEN : On the philosophical relevance of Gödel's incompleteness theorems. *Revue internationale de philosophie*, 234:513–534, 2005.
- P. RAATIKAINEN : Gödel's incompleteness theorems. *Stanford Encyclopedia of Philosophy*, 2015.
- V. RODYCH : Wittgenstein's inversion of Gödel's theorem. *Erkenntnis*, 51:173–206, 1999.
- S. SHANKER : Wittgenstein's remarks on the significance of Gödel's theorem. In S. G. SHANKER, éd. : *Gödel's Theorem in Focus*. Croom Helm, Londres, 1988.
- S. SHAPIRO : Incompleteness, mechanism, and optimism. *The Bulletin of Symbolic Logic*, 4:273–301, 1998.
- W. SIEG : *Hilbert's Programs*. Oxford University Press, Oxford, 2013.
- S. G. SIMPSON : *Subsystems of Second Order Arithmetic*. Cambridge University Press, Cambridge, 2009.
- P. SMITH : *An Introduction to Gödel's Theorems*. Cambridge University Press, Cambridge, 2013. 2^e éd.
- R. SMULLYAN : *Gödel's Incompleteness Theorems*. Oxford University Press, Oxford, 1992. Trad. fr. de M. Margenstern, *Les théorèmes d'incomplétude de Gödel*, Masson, Paris, 1993; Dunod, Paris, 2000.
- J. STILLWELL : *Reverse Mathematics. Proofs From the Inside Out*. Princeton University Press, Princeton, 2018.
- D. VAN DALEN : *Logic and Structure*. 5^e éd., Springer, Berlin, 2013.
- J. VUILLEMIN : La philosophie des mathématiques et les théorèmes négatifs de Gödel. In F. NEF et D. VERNANT, édés : *Le formalisme en question. Le tournant des années trente*. Vrin, Paris, 1988, p. 169–189.
- L. WITTGENSTEIN : *Bemerkungen über die Grundlagen der Mathematik*. B. Blackwell, Oxford, 1956. Éd. d'E. Anscombe, R. Rhees et G. H. von Wright. Trad. fr. de M. A. Lescourret, *Remarques sur les fondements des mathématiques*, Gallimard, Paris, 1983.

L. WITTGENSTEIN : *Philosophische Bemerkungen*. B. Blackwell, Oxford, 1964. Éd. de R. Rhees. Trad. fr. de J. Fauve, *Remarques philosophiques*, Gallimard, Paris, 1975.