



HAL
open science

De quoi les "théorèmes de limitation des formalismes" : théorèmes de Gödel de 1931 et apparentés, sont-ils la limitation?

Patrice Pissavin

► To cite this version:

Patrice Pissavin. De quoi les "théorèmes de limitation des formalismes" : théorèmes de Gödel de 1931 et apparentés, sont-ils la limitation?. Philosophie. Université Panthéon-Sorbonne - Paris I, 2019. Français. NNT : 2019PA01H212 . tel-02493455

HAL Id: tel-02493455

<https://theses.hal.science/tel-02493455>

Submitted on 27 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Paris 1 Panthéon-Sorbonne, UFR de Philosophie

IHΦST

THÈSE

Pour l'obtention du grade de Docteur en Philosophie

De l'université Paris 1 Panthéon-Sorbonne

Présentée et soutenue publiquement

Le 15 février 2019 par

Patrice Pissavin

De quoi les « théorèmes de limitation des formalismes » :
Théorèmes de Gödel de 1931 et apparentés, sont-ils la limitation ?

Sous la direction du Professeur Pierre Wagner

Jury :

Hourya Bénis	Directeur de recherche au CNRS
Jean-Paul Delahaye	Professeur émérite de l'université de Lille
Gerhard Heinzmann	Professeur à l'université de Lorraine
Philippe de Rouilhan	Directeur de recherche émérite au CNRS
Pierre Wagner	Professeur à l'université Paris 1

Résumé

Nous cherchons à définir la nature des limitations révélées par les « théorèmes de limitation des formalismes » (Théorèmes de Gödel de 1931, de Church de 1936, et de Turing de 1936-1937). Pour répondre à cette question, nous avons retenu comme fil conducteur le programme de Hilbert (au sens large) : d'une part la réponse qu'Hilbert espérait apporter au problème des fondements, d'autre part la justification qu'il espérait apporter à l'absence de problèmes mathématiques insolubles. Cela nous a d'abord conduit à proposer une interprétation précise des deux aspects de ce programme. Nous avons ensuite analysé les propositions variées qui ont été faites en réponse à ce programme, incluant en particulier celle de Michael Detlefsen, tout en prenant en compte les résultats d'indécidabilité arithmétiques qui ont été obtenus dans les années 1970. Nous avons pour cela effectué une analyse détaillée de la thèse de Church-Turing. Nous avons également discuté les différentes positions qui ont été défendues dans le cadre du débat induit par l'argument de Lucas-Penrose. Nous avons enfin discuté les réponses que Post, Myhill et Ladrière ont successivement apportées à la question générale posée. Sur la base de l'ensemble de cette analyse, notre propre réponse est que ces théorèmes mettent en évidence une certaine relativité associée au recours à la formalisation elle-même, qui se caractérise par un ancrage nécessaire bien circonscrit dans la pratique empirique des mathématiques informelles.

Mots-clefs : argument de Lucas-Penrose, limitation de la connaissance, philosophie de la logique, philosophie des mathématiques, programme de Hilbert, systèmes formels, théorèmes de Gödel, théorie de la connaissance, thèse de Church-Turing.

Abstract

We want to define the limitations content revealed by the theorems of formalisms limitation (Gödel's theorems of 1931, Church's theorem of 1936 and Turing's theorem of 1936-1937). In order to answer this question, we have accepted as main theme Hilbert's program (in the broad sense) : on the one hand, the answer that Hilbert hoped to give to foundations problem, and on the other hand, the justification he hoped to give to the lack of insoluble mathematical problems. This first lead us to propose a precise interpretation of the two aspects of this program. We have then analyzed the various proposals which have been given in answer this program, including in particular Michael Detlefsen's one, taking into account arithmetical undecidability results obtained in the 1970's. In this aim we have made a detailed analysis of Church-Turing's thesis. We have also discussed the different positions which have been held within the framework induced by Lucas-Penrose's argument. We have then discussed Post, Myhill and Ladrière's successively answers given to the general question asked. On the basis on this whole analysis, our own answer is that these theorems show a kind of relativity in relation with the use of formalization itself, which must be rooted in a confined part of the empirical practice of informal mathematics.

Remerciements

Il n'est pas possible ici de ne pas faire état du contexte bien particulier dans lequel s'est déroulé ce travail, qui est celui d'une activité professionnelle de son auteur à plein temps dans un tout autre domaine, qui plus est nécessitant un fort investissement. Il en a naturellement résulté une durée totale considérablement plus importante que celle habituellement consacrée à une thèse de doctorat, avec les contraintes que cela implique pour tous ceux qui nous ont suivi durant ce parcours. Cela concerne d'abord l'Ecole Doctorale de Philosophie de Paris 1, dont nous remercions les directeurs successifs d'avoir bien voulu nous accueillir : le regretté Jean Gayon d'abord, Chantal Jaquet ensuite et enfin Pierre-Marie Morel. Cela concerne ensuite l'Institut d'Histoire de Philosophie des Sciences et des Techniques, dont nous remercions les directeurs successifs de nous avoir fait confiance : Jacques Dubucs d'abord, Jean Gayon ensuite et Maximilian Kistler enfin. Nous témoignons surtout notre gratitude à Philippe de Rouilhan et Pierre Wagner, qui ont eu la patience de nous suivre sur une aussi longue durée, et dont la pertinence des critiques et le haut degré d'exigence nous auront été particulièrement profitables. Nous tenons également à remercier toutes les personnes qui nous ont amené des éclaircissements très utiles sur des points techniques particuliers : Jacques Bouveresse, lors des discussions qui ont suivi son séminaire au Collège de France consacré aux théorèmes de Gödel de 1931, le regretté Marcel Guillaume, lors d'échanges concernant la traduction des *Grundlagen der Mathematik* de Hilbert et Bernays, Michael Detlefsen lors d'échanges concernant sa propre interprétation instrumentaliste du programme de Hilbert et Jean-Paul Delahaye lors d'échanges concernant la portée philosophique des résultats obtenus en Théorie Algorithmique de l'Information. Nous terminons par une pensée pour le regretté Paul Gochet, qui avait gentiment accepté de nous servir d'intermédiaire afin de rencontrer Jean Ladrière, ce qui s'est trouvé malheureusement juste avant que celui-ci ne décède.

Avertissement

Le sujet traité se caractérise par l'existence d'un très grand nombre de publications, dont une part significative est très récente (des années 2000), ainsi que la consultation de la bibliographie permet d'en juger. Certaines de ces publications en constituent déjà une synthèse, mais indéniablement plus succincte que l'analyse proposée ; ce sont les suivantes : [Sinaceur 1993], [Delahaye 1994, § 6, 7], [Blanché, Dubucs 1996, § 12], [Feferman 1998a], [Murawski 1999, § 4], [Potter 2000, § 23, 28, 34, 35], [Giaquinto 2002], [Zach 2003b], [Raatikainen 2013], [Sieg 2013c] et [Kahle 2015]. Ce sujet se caractérise également par une très forte technicité en logique mathématique (formelle), pour laquelle le choix retenu est celui d'un rappel systématique de toutes les définitions des notions importantes utilisées au fur et à mesure de leur introduction, avec un index récapitulatif à la fin du document. Lorsqu'un même terme existe à la fois dans le langage courant et en logique formelle (par exemple : « complet », « correct », « effectif », « énoncé »), avec évidemment un sens beaucoup plus précis et circonscrit dans le second cas, c'est par défaut celui de la logique formelle qui est retenu, sauf mention explicite contraire. L'essentiel de ces publications ayant eu lieu en anglais, sachant que la plupart des textes originaux en allemand ont été traduits en anglais ou en français, lorsqu'il existe à la fois une version française et une version anglaise (dans le cas d'un original en allemand ou en anglais), le choix retenu est celui de la version française (sachant qu'aucun original en allemand non traduit n'a été exploité). La traduction en français des principales notions utilisées est presque toujours consacrée par l'usage. Il convient toutefois de mentionner deux cas particuliers : 1) le terme « finitist » a été traduit par « finitiste » et quelquefois par « finitaire », conduisant, selon l'habitude, à la construction des adverbes « finitairement » et quelquefois « finitistement », 2) le terme de « computation » a été traduit par « calcul » et non par « processus sur ordinateur », suivant en cela la convention habituelle en logique (et non celle en informatique).

Table des matières

Résumé.....	1
Abstract.....	2
Remerciements	3
Avertissement.....	4
Liste des termes techniques introduits	8
Introduction.....	9
1 L'interprétation du programme de Hilbert au sens restreint.....	14
1.1 <i>Le programme de Hilbert au sens restreint.....</i>	16
1.2 <i>Justification de la prise en compte de l'ensemble des FRP et seulement elles, comme fonctions admises dans l'arithmétique « réelle ».....</i>	29
1.3 <i>Justification de la prise en compte des seuls énoncés Δ_0 et Π_1 comme énoncés de l'arithmétique « réelle ».....</i>	32
1.4 <i>Définition d'une procédure finitiste conduisant à la démonstration de consistance des théorèmes de la partie « réelle » de l'arithmétique du deuxième ordre.....</i>	36
1.5 <i>Expression technique du programme de Hilbert au sens restreint dans le cadre de l'interprétation retenue.....</i>	42
2 Le programme de Hilbert au sens restreint a-t-il été invalidé par les deux théorèmes de Gödel de 1931 ?.....	54
2.1 <i>Le programme de Hilbert au sens restreint était-il condamné d'avance ?.....</i>	55
2.2 <i>Le programme de Hilbert au sens restreint a-t-il été invalidé par le premier théorème ?</i>	58
2.3 <i>Le programme de Hilbert au sens restreint a-t-il été invalidé par le second théorème ?.....</i>	63

3 Peut-on sauver le programme de Hilbert au sens restreint en élargissant les critères du finitisme ?	77
3.1 <i>Les résultats « positifs » obtenus postérieurement aux théorèmes de Gödel de 1931</i>	79
3.2 <i>L'introduction de la règle-ω et la position de l'école de Hilbert sur l'extension du finitisme</i>	87
3.3 <i>L'inépuisabilité des mathématiques selon Franzen</i>	99
3.4 <i>L'instrumentalisme de Detlefsen</i>	102
3.5 <i>Le recours à la logique paraconsistante</i>	108
4 Peut-on réaliser partiellement le programme de Hilbert au sens restreint, en acceptant une réduction de la portée des mathématiques concernées ?	113
4.1 <i>Les résultats partiels obtenus avant 1931</i>	114
4.2 <i>Les résultats modernes obtenus dans le cadre des « reverse mathematics »</i>	117
4.3 <i>Les énoncés combinatoires finis, théorèmes de Z_2 et indécidables dans PA</i>	122
4.4 <i>La thèse d'Isaacson</i>	124
4.5 <i>La solution radicale de l'ultrafinitisme</i>	129
5 Le programme de Hilbert au sens large : les résultats « négatifs » en logique formelle et en informatique théorique s'opposent-ils à la justification de l'absence de problèmes mathématiques insolubles ?	137
5.1 <i>Le théorème de Tarski de 1940-1967 et 1948, sur la complétude syntaxique et la décidabilité de l'algèbre élémentaire ainsi que de la géométrie élémentaire</i>	138
5.2 <i>L'optimisme rationaliste de Hilbert confronté à l'incomplétude syntaxique et à l'indécidabilité de PA</i>	140
5.3 <i>L'affirmation, en Théorie Algorithmique de l'Information, d'une limite au domaine des vérités mathématiques accessibles</i>	149
5.4 <i>La thèse de Church-Turing et les tentatives de dépassement de celle-ci</i>	153

6 Les théorèmes de Gödel de 1931 permettent-ils de mettre en évidence une supériorité de l'esprit humain sur la « machine » ?	173
6.1 <i>La position de Gödel</i>	174
6.2 <i>L'argument de Lucas</i>	179
6.3 <i>Les deux arguments de Penrose</i>	190
6.4 <i>L'argument de Wright</i>	196
6.5 <i>L'argument de la logique paraconsistante</i>	197
7 Les théorèmes « de limitation des formalismes » représentent-ils une limitation de la connaissance ?	201
7.1 <i>Les « Remarques sur les fondements des mathématiques » de Wittgenstein, concernant les théorèmes de Gödel de 1931</i>	201
7.2 <i>La position de Post</i>	209
7.3 <i>La position de Myhill</i>	211
7.4 <i>La position de Ladrière</i>	214
Conclusion	221
Bibliographie	231

Liste des termes techniques introduits

Complétude sémantique par rapport à l'ensemble des modèles	§ 5.2	p. 147	n. 293
Complétude sémantique par rapport à un modèle	§ 5.2	p. 148	n. 296
Complétude syntaxique	§ 1.2	p. 29	n. 48
Complexité algorithmique	§ 5.3	p. 149	n. 297
Conservativité d'un système formel par rapport à un autre	§ 1.4	p. 41	n. 89
Consistance absolue	§ 1.1	p. 25	n. 38
Consistance sémantique	§ 1.1	p. 25	n. 40
Consistance simple	§ 1.1	p. 25	n. 37
Correction	§ 1.1	p. 26	n. 43
[Problème] Décidable	§ 5.2	p. 142	n. 280
[Système formel] Décidable	§ 5.2	p. 143	n. 285
Enoncés Δ_0 , Σ_1 , Π_1 de la hiérarchie de Kleene	§ 1.0	p. 15	n. 12
[Propriété] Exprimable (définissable) dans un langage formel	§ 2.2	p. 59	n. 129
Fonction Récursive Générale	§ 1.2	p. 31	n. 56
Fonction Récursive Primitive	§ 1.0	p. 15	n. 11
Nombres Ω_U	§ 5.3	p. 149	n. 298
Ordinal ε_0	§ 1.4	p. 39	n. 77
Ordinal Γ_0	§ 3	p. 77	n. 174
Prédicativisme au sens de Poincaré	§ 3	p. 78	n. 175
[Ensemble d'entiers] Récursif	§ 2.2	p. 59	n. 125
[Propriété] Récursive primitive	§ 2.2	p. 59	n. 126
[Propriété] Représentable (capturable) dans un système formel	§ 1.4	p. 37	n. 69
Système formel PA (ou Z_1 ou Z)	§ 1.2	p. 29	n. 49
Système formel PRA	§ 1.0	p. 15	n. 13
Système formel Q (de Robinson)	§ 1.4	p. 39	n. 74
Système formel Z_2 (ou PA_2)	§ 1.5	p. 52	n. 114
1-complétude	§ 1.5	p. 44	n. 100
1-consistance	§ 1.5	p. 43	n. 99
Δ_0 -correction, Σ_1 -correction, Π_1 -correction	§ 1.5	p. 43	n. 94
Δ_0 -complétude, Σ_1 -complétude, Π_1 -complétude	§ 1.5	p. 43	n. 94
ω -complétude	§ 1.3	p. 35	n. 66
ω -consistance	§ 1.1	p. 26	n. 41

Introduction

Pour les philosophes de la logique et des mathématiques de langue française, l'expression : « limitation(s) des formalismes » renvoie vraisemblablement au titre même de l'ouvrage majeur de Jean Ladrière publié il y a maintenant plus d'un demi-siècle : « Les limitations internes des formalismes » ([Ladrière 1957]). La notion de « limitation » ou de « limite »¹, a cependant été associée aux théorèmes de Gödel de 1931 et apparentés² beaucoup plus largement dès après leur publication [Fraenkel, Bar-Hillel 1958, p. 307, 344], [Quine 1972 (1984), p. 197], [Goodman 1979 (1986), p. 83-84] et [Webb 1980, p. 200]. Comme il n'est sans doute pas utile d'y insister, l'idée s'est même répandue au-delà du cercle limité des philosophes jusqu'à investir les domaines d'activité les plus variés, fort loin des mathématiques, avec toutes les dérives qui en ont résulté³. Il est donc naturel de se poser la question de la nature précise des limitations qui seraient révélées par ces théorèmes. S'agit-il seulement de restrictions techniques quant à l'aptitude des systèmes formels à répondre à ce pour quoi ils ont été créés par les mathématiciens ? Ou bien ces théorèmes expriment-ils une contrainte beaucoup plus forte sur la capacité humaine à pratiquer les mathématiques⁴, voire même, d'une manière plus générale, sur l'exercice de la raison ? A l'évidence, cette dernière possibilité ouvre la perspective d'un questionnement proprement philosophique qui dépasse le strict contexte mathématique, ou même de philosophie des mathématiques, qui a conduit à l'établissement de ces théorèmes. C'est à une réponse à une telle interrogation que la présente réflexion est dédiée.

Pour cela, il est avant tout nécessaire de revenir au contexte bien particulier de la philosophie des mathématiques du début du siècle précédent, dans lequel s'inscrivent ces théorèmes. La figure centrale en est évidemment Hilbert et son programme, par rapport auquel et contre lequel ces théorèmes se sont imposés comme un élément majeur. Encore faut-il s'entendre sur l'étendue et la mesure précise de cet impact, ce qui ne fait toujours pas l'objet d'un consensus en philosophie des mathématiques. Cela impose d'abord de définir le contenu de ce programme,

¹ Nous suivons l'usage établi en n'introduisant pas de différence de sens entre ces deux termes.

² C'est-à-dire les théorèmes métamathématiques « négatifs » concernant *l'arithmétique*, incluant en particulier les théorèmes de Church de 1936 et de Turing de 1936-1937 (article publié en 1936, un correctif ayant été publié en 1937), et dans une moindre mesure (parce que son objet dépasse le cadre de notre propos), celui de Tarski de 1933-1936 (article d'abord publié en polonais en 1933, puis en allemand avec un post-scriptum en 1936).

³ Conduisant au mieux (dans le cas où l'auteur sait de quoi il parle) à des analyses trop générales pour être instructives (exemple de [Boos 1998]), et au pire (dans le cas beaucoup plus fréquent où ce n'est pas le cas) à ce que l'on a du mal à distinguer du délire verbal (voir la critique de [Franzen 2005]).

⁴ Comme le laissent entendre, au-delà du classique : « Mathématiques, la fin de la certitude » ([Kline 1980]), les titres accrocheurs d'ouvrages plus récents : « Les limites des mathématiques » ([Chaitin 1998]), « L'inconnaissable » ([Chaitin 1999]), « Les limites de la connaissance » ([Zwirn 2000]) (dont la portée est plus large) et « Les limites de la raison » ([Chaitin 2006]).

ce pourquoi nous proposons à la suite de Georg Kreisel [Kreisel 1958b, p. 487] et Wilfried Sieg [Sieg 1999, p. 29], la distinction méthodologique entre ce qui peut être nommé : « le programme de Hilbert au sens large », rassemblant indistinctement l'ensemble des positions de Hilbert concernant la philosophie des mathématiques et « le programme de Hilbert au sens restreint », limitant la part de celui-ci qui vise à répondre au « problème des fondements ». Pour Hilbert, le finitisme et partant le programme au sens restreint lui-même n'avait sans doute pas à être précisé au-delà des indications informelles qu'il en a données. Une caractérisation précise et surtout formelle en est cependant indispensable pour pouvoir discuter de la mesure dans laquelle des résultats mathématiques précis portants sur des systèmes formels particuliers comme les théorèmes de Gödel de 1931, sont susceptibles de remettre en cause ce programme. Pour cette caractérisation formelle, le point de départ est naturellement la « thèse de Tait » : [Tait 1981] mais dont il convient de préciser le contenu et de discuter les objections qui lui ont été opposées.

Après avoir donné un contenu technique précis au programme de Hilbert au sens restreint, il est possible de confronter directement ce programme aux deux théorèmes de Gödel de 1931. Cela impose de traiter à la fois du premier et du second théorème, auxquels se heurte le programme de Hilbert au sens restreint d'une manière différente. Au préalable, il convient de répondre aux critiques variées qui ont été formulées à l'encontre de ce programme et qui visent à discréditer celui-ci *indépendamment* des objections essentiellement techniques apportées par les théorèmes de Gödel de 1931. L'objection majeure à la position dominante, qui a relancé le débat avec un certain retentissement dans la communauté des philosophes de la logique, est indiscutablement venue de Michael Detlefsen : [Detlefsen 1979], [Detlefsen 1986], [Detlefsen 1990] et [Detlefsen 2001] suivi (partiellement) par quelques rares auteurs, principalement Marcus Giaquinto : [Giaquinto 2002] et dans une certaine mesure Panu Raatikainen : [Raatikainen 2003]. C'est donc essentiellement la position défendue par Michael Detlefsen, concernant aussi bien le premier théorème de 1931 que le second, qu'il convient de discuter.

Comme on le sait, la voie qui consiste à élargir le finitisme pour sauver le programme de Hilbert au sens restreint et à laquelle invite Gödel à la fin de l'article de 1931, a été évoquée par Hilbert lui-même (dans le seul commentaire qu'on lui connaisse sur le sujet) après l'irruption des théorèmes de Gödel de 1931 [Hilbert, Bernays 1934 (2001), p. 43]. Une telle position amène naturellement deux questions, d'abord sur la circonscription du nouveau finitisme proposé, ensuite et surtout sur la justification épistémologique de celui-ci : que reste-il alors des caractéristiques du programme initial ? La tentative la plus détaillée de définition d'un finitisme élargi par l'école de Hilbert se trouve sans aucun doute dans [Hilbert, Bernays 1939, § 5.3], en relation en particulier avec l'introduction par Hilbert de la « règle- ω » en 1930. Il est donc

nécessaire de l'analyser de manière détaillée. Par ailleurs, de nombreuses autres tentatives ont été proposées sur une base finitiste élargie visant implicitement une sorte de « programme de Hilbert » modifié, en prenant acte des théorèmes de Gödel de 1931. Les plus connues et les plus anciennes sont au nombre de trois, à savoir, un résultat de Gödel lui-même : [Gödel 1933b], le théorème de Gentzen de 1936 et encore un résultat de Gödel : [Gödel 1958]. Une solution plus récente, se réclamant explicitement d'un « programme de Hilbert » modifié, est l'instrumentalisme de Michael Detlefsen : [Detlefsen 1986] qu'il convient d'analyser dans sa spécificité. Dans tous les cas il s'agit d'évaluer ces différentes solutions, en s'attachant à en déterminer à la fois la portée ainsi que la mesure dans laquelle elles constituent une extension raisonnable du finitisme de Hilbert.

A défaut de pouvoir atteindre l'objectif complet du programme de Hilbert au sens restreint, même en acceptant d'en étendre le finitisme initial, se pose légitimement la question de savoir s'il est au moins possible d'en obtenir une réalisation *partielle*, pour un domaine limité des mathématiques, tout en prenant en compte les deux théorèmes de Gödel de 1931. Bien entendu, le fait d'accepter de ne pas pouvoir rendre compte de l'ensemble des mathématiques sur une base finitiste, atténue grandement l'intérêt même du programme tel que l'avait conçu Hilbert. Cette réserve importante étant émise, essentiellement deux voies ont été adoptées en vue de proposer une réponse positive à la question posée, avec de nombreuses variantes. La première est d'ordre mathématique et conduit à déterminer, pour tout domaine mathématique donné, les systèmes formels les plus forts, conservatifs par rapport à un système formel de base fixé qui est typiquement PRA ou PA : c'est la démarche des « reverse mathematics », apparue au milieu des années 1970. Le caractère partiel de réalisation du programme de Hilbert au sens restreint y est alors défini, d'une part par l'étendue du domaine mathématique dont il est possible de rendre compte, et d'autre part par l'écart de la puissance du système formel de départ à celle de celui qui peut être rattaché au finitisme originel, dont on devine qu'il va s'agir de deux caractéristiques antagonistes. La seconde voie est d'ordre philosophique et elle est plus radicale. Elle consiste à tenter de justifier que les énoncés indécidables issus des théorèmes de Gödel 1931 ne sont pas des énoncés véritablement arithmétiques et que pour ces derniers un système formel comme PA est en fait syntaxiquement complet : c'est la thèse de Daniel Isaacson : [Isaacson 1987] et [Isaacson 1992]. Une telle position doit alors tenir compte non seulement des théorèmes de Gödel proprement dits, mais aussi de différents résultats également obtenus à partir du milieu des années 1970 montrant que certains énoncés *combinatoires finis* sont décidables dans Z_2 et indécidables dans PA.

L'autre grande préoccupation de Hilbert en philosophie des mathématiques, au-delà de la résolution finitiste du problème des fondements (le programme au sens restreint), était la justification de l'absence de problèmes mathématiques insolubles (l'ensemble conduisant au programme au sens large). Les principaux résultats « négatifs » obtenus concernant PA : son incomplétude syntaxique (premier théorème de Gödel de 1931) et son indécidabilité (théorèmes de Church de 1936 et de Turing de 1936-1937) sont habituellement présentés comme une réfutation de cette dernière position de Hilbert. Il s'agit d'analyser la pertinence d'une telle objection, ce qui nécessite de donner un contenu technique précis à la position de Hilbert, au moins pour l'arithmétique. Dans un contexte complètement différent, mais également à partir d'extensions du théorème de Turing de 1936-1937, les figures centrales de la Théorie Algorithmique de l'Information que sont Grégory Chaitin et Cristian Calude défendent quant à elles une position qui s'inscrit radicalement en faux contre celle de Hilbert. Ces auteurs n'affirment en effet rien de moins que l'existence de vérités arithmétiques « absolument » indécidables, et partant de *l'existence* de problèmes mathématiques insolubles. Une telle position a à son tour été contestée, principalement par Panu Raatikainen : [Raatikainen 1998] et [Raatikainen 2000]. Il s'agit d'analyser les arguments présentés par les deux parties. On peut enfin souligner que pour l'ensemble des débats qui viennent d'être évoqués, la prise en compte de la thèse de Church-Turing intervient de manière cruciale en permettant de donner un contenu technique précis à l'indécidabilité. Or un certain nombre de propositions ont été avancées récemment en faveur d'un dépassement de celle-ci. Il s'agit d'analyser ces travaux, susceptibles d'apporter un éclairage nouveau à la problématique de la justification de l'absence de problèmes mathématiques insolubles.

La question générale de savoir si la pensée à l'œuvre dans l'exercice des mathématiques doit être considérée comme « mécanique » (en un sens qui reste évidemment à préciser) a déjà été soulevée par Hilbert, qui y a répondu d'une manière positive mais avec une justification vague. Comme on le sait, c'est seulement en 1936 que Turing en a donné une justification précise via l'introduction des machines de Turing, mais en se limitant aux opérations de calcul. En 1961, la question est venue sur la scène philosophique au travers d'un article remarqué de John Lucas affirmant la supériorité de l'esprit humain sur la « machine », sur la base du premier théorème de Gödel de 1931 : [Lucas 1961]. Ce point de vue a ensuite été défendu à son tour par Roger Penrose à partir de la fin des années 1980 : [Penrose 1989] et [Penrose 1994]. Il s'avère alors que pratiquement tous les grands noms de la philosophie de la logique et des mathématiques se sont exprimés à ce sujet pour s'opposer à ces deux auteurs, au cours d'un débat qui a lieu quasiment sans discontinuité depuis cinquante ans. Il s'agit d'analyser les différents arguments

qui ont été présentés en faveur de et contre une telle thèse, en commençant par présenter la position de Gödel lui-même qui s'est engagé le premier dans un tel débat.

Cette discussion ramène naturellement à la question plus large, point de départ de la réflexion entreprise, qui concerne l'impact des théorèmes « de limitation des formalismes » sur la connaissance humaine. Cela nécessite d'analyser la position des rares auteurs qui se sont engagés en faveur d'une telle limitation de la connaissance humaine, voire de la raison, révélée par les théorèmes de Gödel de 1931 et apparentés, à savoir : Emil Post dans les années 1940, John Myhill dans les années 1950 et surtout Jean Ladrière dans les années 1960.

1 L'interprétation du programme de Hilbert au sens restreint

Pour l'essentiel, le programme de Hilbert⁵ est la réponse de Hilbert⁶ au « problème des fondements » [Hilbert 1928 (1992), p. 146] et [Hilbert 1931(1992), p. 189, 191]. Comme on le sait, ce dernier terme se rapporte aux conséquences en philosophie des mathématiques, de l'irruption des différents paradoxes (au premier rang desquels celui de Russell), dès la fin du XIX^{ème} siècle. Ainsi que l'ont fait remarquer Roger Martin [Martin 1964, p. 171] et Jean Toussaint Desanti [Desanti 1990, p. 713-714], il s'agit en fait plutôt d'un problème de « fondation » des mathématiques que de « fondement » (terme qui a clairement une connotation ontologique), mais l'expression (en français⁷) de « problème des fondements » est consacrée par l'usage. D'une manière plus précise, en reprenant les propositions de Georg Kreisel [Kreisel 1958b, p. 487] et Wilfried Sieg [Sieg 1999, p. 29], et en ne suivant pas Gentzen [Gentzen 1938a (1969), p. 238], nous introduisons une coupure terminologique entre ce que nous nommons respectivement : « programme de Hilbert au sens large », rassemblant indistinctement l'ensemble des positions de Hilbert concernant la philosophie des mathématiques⁸, et « programme de Hilbert au sens restreint », plus simplement : « programme de Hilbert⁹ », la part de celui-ci qui se limite à répondre au problème des fondements. Nous allons donc commencer par présenter le contenu de ce programme tel qu'il a été défini par Hilbert et son école, en relation avec le « finitisme » auquel il est rattaché, avec la distinction cruciale introduite entre les énoncés mathématiques « réels » et « idéaux ».

Pour Hilbert, le finitisme et partant le Programme lui-même, n'avait sans doute pas à être précisé au-delà des indications informelles qu'il en a données. Une caractérisation précise en est cependant indispensable pour pouvoir discuter de la mesure dans laquelle des résultats mathématiques précis, portant sur des systèmes formels particuliers comme les théorèmes de Gödel de 1931, sont susceptibles de remettre en cause le Programme. Cela impose d'abord de donner un contenu technique et *formel* au programme au sens restreint, et pour cela de fixer,

⁵ Pour ce qui concerne l'origine historique de la locution : « programme de Hilbert », le terme de « programme méthodique » est évoqué dès 1930 [Bernays 1930/1976 (2003), p. 78], mais la formule : « programme de Hilbert » est seulement explicite en 1936 [Gentzen 1936a (1969), p. 135].

⁶ Il convient de prendre garde au fait que lorsque Hilbert et ses collaborateurs parlent « d'arithmétique », ils donnent à ce terme un sens très large incluant l'analyse, et jusqu'en 1934, ils ne distinguaient pas radicalement le 1^{er} et le 2^{ème} ordre.

⁷ Sachant qu'en anglais, les deux mots sont traduits par un terme unique : « foundation », ce qui a d'ailleurs permis à Stephen Kleene le trait d'humour de [Kleene 1986a, p. 127].

⁸ Incluant la question de la justification de l'optimisme rationaliste de Hilbert (l'absence de *ignorabimus*) que nous traiterons au § 5.2.

⁹ Que nous désignerons également par le simple terme de « Programme » (avec une majuscule)

d'une part le contenu de l'arithmétique « réelle », et d'autre part l'étendue des preuves de consistance finitairement acceptables. Différentes solutions ont été proposées pour répondre à cette demande. Celle que nous voulons défendre est une interprétation libéralisée de l'interprétation « standard », celle-ci recouvrant quant à son contenu ce qui est en général nommé la « thèse de Tait » à la suite de l'article séminal de celui-ci¹⁰ : [Tait 1981]. Cette interprétation standard prend la forme de trois affirmations. Tout d'abord, les fonctions admissibles dans l'arithmétique « réelle » sont toutes les FRP¹¹ et seulement elles. Ensuite, les énoncés de l'arithmétique « réelle » sont les seuls énoncés Δ_0 et Π_1 selon la hiérarchie de Kleene¹². Enfin, les preuves finitairement acceptables se confondent avec les dérivations dans PRA¹³. C'est sur cette dernière affirmation que notre interprétation diffère quelque peu de l'interprétation standard, en nous limitant à une procédure finitiste conduisant à une démonstration *de consistance*. En abordant successivement chacune de ces affirmations, nous serons amenés à évoquer et à critiquer les solutions différentes qui ont été proposées comme

¹⁰ Qu'il soit cependant clair que si nous reprenons à notre compte pour l'essentiel la proposition technique de William Tait, nous ne suivons pas nécessairement celui-ci, ni dans son refus de considérer comme finitiste une fonction du seul fait d'être une fonction [Tait 1981, p. 527], ni dans sa contestation de l'idée que le Programme puisse assurer une sécurisation des mathématiques [Tait 1981, p. 546].

¹¹ Le terme FRP désigne les Fonctions Récursives Primitives, définies pour la première fois de manière précise par Gödel dans [Gödel 1931]. Les FRP sont des fonctions d'entiers dont toutes les valeurs peuvent être calculées par itération de proche en proche, de telle sorte que pour chaque valeur de l'argument il soit possible de déterminer, en un nombre fini d'étapes, la valeur correspondante de la fonction.

¹² La hiérarchie de Kleene permet de rendre compte de l'ensemble de tous les énoncés arithmétiques possibles selon leur complexité logique. Elle est exprimée d'une manière générale par les énoncés : « Δ^k_0 », « Σ^k_j », « Π^k_j », où l'indice inférieur traduit le degré dans la hiérarchie et l'indice supérieur l'ordre de la logique utilisée ; Par convention, Δ^k_0 désigne tout énoncé d'ordre k-1 (premier ordre : k = 0, deuxième ordre : k = 1, ...). Au premier ordre, nous n'avons donc affaire qu'aux énoncés : « Δ^0_0 », « Σ^0_j », « Π^0_j ». Nous utiliserons alors la simplification d'écriture habituelle qui consiste à omettre l'indice supérieur tant qu'on se limite au premier ordre et à les écrire respectivement : « Δ_0 », « Σ_j », « Π_j ».

Soit $\varphi_0(x_i)$ une formule de variables libres x_i (en nombre fini mais non borné), résultant de toute combinaison logique n'utilisant que des quantificateurs bornés d'équations ou inéquations arithmétiques ne faisant intervenir que des FRP. L'exemple le plus simple est : $f_0(x) = 0$ où f_0 est FRP.

Le niveau le plus bas dans la hiérarchie est constitué par les énoncés de type Δ_0 qui sont toutes les instances des formules $\varphi_0(x_i)$. La négation d'un énoncé de type Δ_0 est un énoncé de type Δ_0 .

Le niveau immédiatement supérieur est constitué par les énoncés $\Sigma_1 : \exists x_i \varphi_0(x_i)$ et $\Pi_1 : \forall x_i \varphi_0(x_i)$. La négation d'un énoncé de type Π_1 est un énoncé de type Σ_1 et réciproquement. Les énoncés Δ_0 sont inclus dans les énoncés Π_1 et dans les énoncés Σ_1 : ils sont justement leur partie commune.

La hiérarchie se poursuit de manière illimitée selon les deux relations de récurrence :

1) $\Sigma_{n+1} : \exists x_i \varphi_n(x_i)$ où φ_n est alors de type Π_n ;

2) $\Pi_{n+1} : \forall x_n \varphi_n(x_i)$ où φ_n est alors de type Σ_n .

Chaque niveau supérieur à 1 correspond ainsi à une succession alternée des quantificateurs existentiels et universels : $\Sigma_2 : \exists x_i \forall y_i \varphi_1(x_i, y_i)$, $\Pi_2 : \forall x_i \exists y_i \varphi_1(x_i, y_i)$,

¹³ PRA désigne l'arithmétique récursive primitive. C'est le système formel introduit dans [Skolem 1923] et repris dans [Hilbert, Bernays 1934]. Sa particularité est de ne pas comporter de quantificateurs logiques tout en incluant une infinité de symboles pour les FRP. Dans les limites des possibilités d'expression de son langage il ne comporte aucune restriction sur le schème d'induction.

alternative¹⁴, qui sont soit plus restrictives soit au contraire plus générales sous un aspect au moins. Nous pourrions alors justifier la pertinence de notre interprétation par rapport aux solutions différentes qui ont été proposées sur la base de la thèse de Tait, incluant en particulier l'objection de [Detlefsen 1990]. Pour cela, nous devons établir les relations existant entre les différents concepts métamathématiques pertinents dans le contexte du programme de Hilbert, d'une part dans un cadre classique, et d'autre part dans un cadre finitiste, ce qui est par ailleurs indispensable pour conduire la discussion qui va suivre sur l'impact des théorèmes de Gödel de 1931.

1.1 Le programme de Hilbert au sens restreint

Avec son programme, Hilbert s'est donné pour objectif « d'asseoir les mathématiques sur une base sûre, inaccessible au doute raisonnable », ce qu'il a exprimé à de nombreuses reprises dans des formules quelque peu emphatiques, souvent citées [Hilbert 1923 (1992), p. 131-132], [Hilbert 1926 (1972), p. 221], [Hilbert 1926 (1972), p. 227, 236], [Hilbert 1929 (1992), p. 184-185] et [Hilbert 1931 (1992), p. 196]. Une difficulté majeure surgit cependant lorsque l'on tente de rendre compte de la position de Hilbert en philosophie des mathématiques : il en existe très peu d'écrits publiés¹⁵, ceux-ci s'étalent sur une période d'une trentaine d'années au cours de laquelle la position de Hilbert a notablement évolué¹⁶, et à aucun moment celle-là n'a été présentée par son auteur comme une véritable position philosophique. Cela rend particulièrement délicate la prétention de rendre compte de son programme, comme la plupart des commentateurs de Hilbert n'ont pas manqué de le signaler [Kreisel 1958a, p. 155], [Resnik 1974, p. 117], [Sinaceur 1993, p. 271], [Mancosu 1998, p. 157], [Raatikainen 2003, p. 160], [Zach 2003a, p. 227] et [Zach 2005a, p. 505, 506].

L'intuition tient une place fondamentale dans la démarche adoptée par Hilbert à partir des années 1920, en liaison avec son attitude finitiste. On peut raisonnablement admettre que ce

¹⁴ En nous attachant alors à nous limiter aux solutions qui se rapportent au finitisme *originel*, c'est-à-dire avant les extensions qui ont été proposées par l'école de Hilbert concomitamment à la prise de connaissance des théorèmes de Gödel de 1931, que nous analyserons au § 3.

¹⁵ Il est certainement utile de préciser ici que les deux ouvrages à la fois les plus tardifs et les plus importants publiés sous le nom de Hilbert et Bernays, en 1934 et 1939, ont en fait été rédigés par Bernays seul, comme celui-ci l'a tardivement révélé longtemps après la mort de Hilbert [Sinaceur 2003, p. 8] et [Zach 2004, p. 91]. Cela explique que nous ferons parfois référence à Bernays seul lorsque nous serons amenés à évoquer ces deux ouvrages.

¹⁶ Sachant que l'exposition par Hilbert de son programme, lorsque le finitisme est pris au sens originel, est limité à la période 1922 – 1930 (la date du début étant définie de manière précise par Wilfried Sieg [Sieg 2013c, p. 1]).

terme, bien que n'ayant jamais été précisément circonscrit par lui, recouvre en fait *trois* contenus différents (même s'il a surtout insisté sur le premier cité¹⁷) dont le niveau d'évidence n'est certainement pas le même.

1) Il concerne d'abord l'intuition₁ *sensible* (en général visuelle) associée aux « traces sur le papier » des signes (dépourvus de sens) correspondant aux objets mathématiques ; son caractère d'évidence est difficilement contestable en tant que condition de possibilité de toute connaissance (et ce, quelle que soit l'ontologie retenue concernant la réalité sensible) :

Notamment en mathématiques, notre réflexion porte sur les signes concrets eux-mêmes, dont la forme est, de notre point de vue, immédiatement discernable et reconnaissable. Voilà le minimum de présupposition dont aucun scientifique ne peut se passer et auquel chacun doit donc se tenir, consciemment ou non. [Hilbert 1928 (1992), p. 146]

2) Il inclut ensuite l'intuition₂ « intellectuelle » correspondant à l'application de n'importe quelle série d'opérations *discrètes* effectuées sur des suites *finies* d'entiers au moyen de règles explicites fixées à l'avance ; la justification philosophique en est certainement plus problématique, mais il s'agit empiriquement d'une base minimale¹⁸ reconnue comme évidente pour tous les mathématiciens, quelle que soit leur inclination philosophique :

La condition préalable de l'application des inférences logiques et de l'effectuation d'opérations logiques est l'existence d'un donné dans la perception : à savoir l'existence de certains objets concrets extra-logiques en tant que sensations immédiates, précédant toute pensée. Pour que le raisonnement logique soit sûr, il faut que ces objets soient perçus dans toutes leurs parties et que leur occurrence, leur caractère distinct, leur succession ou leur juxtaposition se présentent à l'intuition en même temps que ces objets, comme quelque chose d'immédiat et qui ne réduit pas ou n'a pas besoin d'être réduit à quoi que ce soit d'autre. Telle est la conception philosophique fondamentale qu'à mon sens exigent les mathématiques et d'ailleurs toute pensée, toute compréhension et toute communication scientifique.

[Hilbert 1926 (1972), p. 228]

3) Il comprend enfin l'intuition₃, toujours « intellectuelle », associée au processus d'engendrement (ou de parcours) progressif *illimité* de la suite des entiers naturels¹⁹ : après un

¹⁷ Le fait que l'intuition s'étende au-delà de son seul aspect sensible est contesté par Philip Kitcher [Kitcher 1976, p. 113-114], mais est défendu par Hourya Sinaceur [Sinaceur 1993, p. 268].

¹⁸ A partir des années 1920 (ce n'était pas le cas en 1904), Hilbert admit que certaines idées intuitives minimales sont les conditions de possibilité de la connaissance mathématique.

¹⁹ Cette position implique évidemment la prise en compte de l'infini potentiel. Elle exclut par contre la prise en compte de l'infini actuel. Ce processus présente la caractéristique fondamentale d'être seulement « en principe », sans aucune contrainte sur les capacités d'engendrement de l'esprit humain ni même sur les possibilités de sa représentabilité physique, comme le montre la citation de Bernays suivante : « L'abstraction intuitive ne fait aucun

entier il y en a toujours un autre ; cette position, qui nécessite la précédente, la dépasse évidemment. Sa justification pose donc un problème encore plus aigu, même si, là encore, il s'agit d'une position reconnue par tous les mathématiciens du temps de Hilbert. Cette variabilité du contenu de l'intuition pose le problème fondamental (jamais abordé de front par Hilbert²⁰) de savoir (et de pouvoir justifier) jusqu'où aller dans l'abstraction.

D'une manière générale, la démarche de Hilbert s'appuie, dans un registre kantien, sur les conditions de possibilités (les fondements) de la connaissance mathématique [Hilbert 1926 (1972), p. 245] et [Hilbert 1931 (1992), p. 188, 189]. Toutefois, et comme le fait remarquer Hourya Sinaceur, l'a priori retenu par Hilbert est assez loin de celui de Kant [Sinaceur 1993, p. 260]. L'a priori y est constitué par les *signes* (mathématiques) qui remplacent l'intuition spatio-temporelle chez Kant. Ces signes sont « concrets » et objectifs (objets de l'intuition sensible comme : « 1 » et « + »), associés à un caractère d'évidence et de certitude (*intuition₁*). Ils sont également donnés intuitivement (avant toute pensée) et non pas par l'exercice de l'entendement. Ils sont encore formels, vides de sens (en dehors de tout présupposé axiomatique extra logique). Ce sont des abstractions de représentations effectives (indépendants des particularités graphiques de celles-ci) :

[...] quelque chose doit déjà être donné dans la représentation : certaines entités discrètes extra-logiques, qui intuitivement sont présentes en tant qu'expérience immédiate antérieure à toute pensée [...] puisque je considère que les objets de l'arithmétique sont les signes eux-mêmes, dont nous savons toujours et à coup sûr identifier la forme indépendamment du lieu, du moment et des circonstances particulières de leur production ainsi que des différences minimales dans leur tracé. Je tiens cette attitude philosophique ferme pour indispensable au fondement de la mathématique pure, comme de toute pensée, intelligibilité et communication scientifique en général. *Au commencement est le signe*, telle est la loi ici.

[Hilbert 1922 (1992), p. 117]

Enfin, ils sont sans limitation sur leur nombre possible au sein d'une succession : un réservoir²¹ inépuisable²² de signes (*intuition₃*).

cas de telles limites [se représenter des grands nombres entiers] à la possibilité de réalisation. Car les limites sont, du point de vue de la considération formelle, contingentes. L'abstraction formelle ne trouve pour ainsi dire aucune place pour une limite principielle, antérieure à la distinction entre fini et infini » [Bernays 1930/1976 (2003), p. 62 ; le commentaire entre crochets est nôtre].

²⁰ Selon Richard Zach, pour Hilbert c'est au philosophe et non au mathématicien de fournir une explication de l'intuition qui sert de base aux mathématiques finitistes [Zach 2005a, p. 505].

²¹ Le terme, associé aux formules, est explicitement mentionné par Hilbert [Hilbert 1926 (1972), p. 232].

²² Ce qui constitue évidemment une certaine abstraction, puisque, comme nous l'avons déjà indiqué, aucune limite *pratique* (au sens d'humainement accessible) n'est imposée : il suffit que les signes soient *en principe* accessibles à l'intuition sensible.

La question se pose alors évidemment de savoir ce que Hilbert entendait exactement par « concret ». Ce ne saurait être l'identification d'un symbole avec ses matérialisations particulières : ses « marques » (des traces de craie ou une tache d'encre), faute de quoi deux « a » ne seraient jamais identiques. Il s'agit donc nécessairement de l'identification d'un symbole avec l'abstraction de toutes ses matérialisations possibles. Voici la précision que Bernays a apportée sur ce point :

En ce qui concerne la structure figurée précise de ces codes, nous supposons comme d'habitude qu'on laisse une certaine latitude, c'est-à-dire qu'il ne faut pas se préoccuper d'une petite différence de réalisation, qu'il s'agisse de la forme du 1, ou de sa grandeur, ou encore de la distance entre les 1 qu'on ajoute. Ce dont nous avons besoin fondamentalement, c'est que [...] nous ayons un objet intuitif qui se reconnaisse toujours sans équivoque, et dont nous devons pouvoir toujours embrasser d'un coup d'œil dans un code les parties discrètes à partir desquelles il est construit. [Hilbert, Bernays 1934 (2001), p. 75]

D'une manière plus précise les signes mathématiques « concrets » peuvent être distingués en trois classes. La première classe est celle de l'arithmétique finitiste, qui se décompose elle-même en deux sous-classes. Il y a d'abord les chiffres (et suites de chiffres) désignant les entiers, chacun d'entre eux n'étant que l'abréviation de la représentation d'une suite donnée de bâtons (« 1 » pour I, « 2 » pour II, « 3 » pour III, ...). Il y a ensuite les signes fondamentaux de l'arithmétique finitiste (« + », « × », « = », « > ») ou des signes dérivés. La troisième classe est celle des symboles logiques utilisés dans les dérivations formelles. A partir des signes « concrets » de l'arithmétique finitiste, il est possible de construire des *propositions* mathématiques qui conservent le caractère « concret » et intuitif au moyen d'une *démarche* finitiste. Cette démarche consiste à appliquer les signes de l'arithmétique finitiste sur les chiffres en se limitant pour cela à des procédés constructifs, autorisant une induction ordinaire (*récurrence*) sur la construction effective de la suite des entiers. Cela suppose l'utilisation de raisonnements « intuitifs » (*intuition*₂), c'est-à-dire par considération directe, sans aucun axiome, d'assertions portant sur des collections finies (pour lesquelles les principes de non contradiction et du tiers exclu s'appliquent). Bien entendu toutes les propositions mathématiques ne sont pas « concrètes », en particulier lorsqu'elles font intervenir le transfini (au minimum un processus infini²³). Celles-ci sont alors dites abstraites.

²³ Sachant que pour Hilbert, comme le fait remarquer Hourya Sinaceur, l'opposition fini (finitiste) *versus* infini ne recouvre pas le sens *numérique* de cette expression : fini (finitiste) ne se limite pas nécessairement à une collection finie même non bornée d'entiers naturels [Sinaceur 1993, p. 264].

En effet, sensible aux critiques de Poincaré et surtout de Brouwer, Hilbert a été amené à introduire une distinction capitale entre les mathématiques et la métamathématique (le raisonnement sur les mathématiques formalisées) [Hilbert 1922 (1992), p. 128] et [Hilbert 1931 (1992), p. 192]. Par contre, pour lui, aucune limitation n'est introduite sur les mathématiques elles-mêmes (en opposition radicale sur ce point avec Brouwer). La métamathématique inclut au contraire les exigences constructivistes des intuitionnistes ainsi que des exigences finitistes particulières (qui ne sont pas exigées des intuitionnistes). Les exigences finitistes de Hilbert sont donc plus contraignantes que les exigences constructivistes de Brouwer²⁴ (en les incluant) mais ne s'appliquent qu'à la métamathématique²⁵ [Hilbert, Bernays 1934 (2001), p. 97]. Les exigences constructivistes prises en compte (pour la métamathématique) sont classiques. Elles consistent d'abord à refuser le tiers exclu appliqué à des collections infinies, donc de déduire de la seule négation d'un énoncé universel l'affirmation de l'énoncé existentiel correspondant. Elles consistent plus généralement dans le refus de l'affirmation de l'existence d'un objet sans disposer des moyens de le construire, et cela dès 1904 [Hilbert 1904 (1992), p. 266], [Hilbert 1923 (1992), p. 136] et [Hilbert 1926 (1972), p. 231]. Le contenu de la démarche finitiste²⁶ de Hilbert et ses élèves a évolué et Hilbert n'en a jamais donné une définition précise. Pour lui, cette notion tenue pour immédiate et irréductible, n'avait pas à être précisée²⁷. Si son programme avait réussi, une idée intuitive aurait sans doute suffi pour s'entendre sur son contenu. En sens inverse, une caractérisation précise est indispensable si l'on veut démontrer *l'impossibilité*, par exemple, d'une démonstration de consistance d'un système formel des mathématiques par des moyens finitistes (nous aurons à y revenir longuement). Ce finitisme inclut sans ambiguïté, toutes les propositions sans variables d'individus²⁸ (dont les arguments sont des chiffres²⁹). Il intègre raisonnablement les énoncés arithmétiques n'utilisant que des quantificateurs bornés, mais au-delà du fini borné, mis à part l'exclusion explicite des quantificateurs existentiels, la position de Hilbert est moins claire.

²⁴ Même s'il est généralement admis (avec une objection de William Tait [Tait 2006b] sur laquelle nous reviendrons au § 3.1) que l'école de Hilbert (von Neumann par exemple) ainsi que Herbrand, assimilaient jusqu'au début des années 1930 les deux points de vue, en nommant « intuitionniste » le finitisme pris en compte par Hilbert [Bernays 1935b, p. 89].

²⁵ Au-delà de la différence radicale sur le contenu de l'intuition dans les deux approches : abstraction de signes sur le papier *versus* constructions mentales (nous reviendrons sur le contenu de l'intuitionnisme au § 3.1).

²⁶ Le terme est apparu explicitement pour la première fois dans l'article de 1923.

²⁷ Richard Zach va même plus loin et considère que : « Hilbert était trop vague sur la question de ce qui constitue le finitisme pour définir la notion de manière univoque » [Zach 2003a, p. 227].

²⁸ Hilbert donne lui-même comme exemples : « $3 > 2$ », « $2 + 3 = 3 + 2$ », « $2 = 3$ », « $1 \neq 1$ » [Hilbert 1926 (1972), p. 229].

²⁹ Pour Hilbert (de par son attitude finitiste justement), les nombres entiers se ramènent à des chiffres (les signes concrets correspondants).

Comme nous l'avons indiqué, les énoncés « réels » (ou « élémentaires » ou « finitistes ») constituent pour Hilbert le cœur des mathématiques, et pour l'ensemble des mathématiciens, la partie non controversée³⁰ (en particulier ils sont acceptés par les intuitionnistes) [Hilbert 1926 (1972), p. 228]. Comme nous l'avons vu il s'agit d'énoncés élémentaires (ou primitifs), directement rattachés au contenu de l'intuition (au sens de Hilbert). Ils possèdent deux caractéristiques essentielles. Ils sont « concrets » (au sens de Hilbert), possèdent un sens (sont pourvus de contenu), et sont directement intelligibles. De ce fait les prédicats « vrai » et « faux » (au sens intuitif) leur sont applicables. Ils sont d'abord constitués par les propositions constituées à partir des chiffres désignant les entiers, associées aux opérateurs constitués par les signes de l'arithmétique finitiste (*intuition*₂). Ils sont ensuite étendus en utilisant des procédés constructifs autorisant l'induction ordinaire (récurrence), c'est-à-dire acceptant le processus d'engendrement *illimité* de la suite des entiers³¹ (*intuition*₃). Le point fondamental de Hilbert est que la pensée ne peut s'exercer (les prédicats « vrai » et « faux », ne peuvent s'appliquer) que sur des éléments « concrets » [Hilbert 1926 (1972), p. 232-233]. Les énoncés (informels) « réels » étant « concrets », appartiennent d'emblée à la mathématique « contentuelle » (qui possède un contenu intuitif³²) lorsqu'ils sont vrais (théorèmes informels). C'est une partie restreinte des mathématiques dont on ne peut pas mettre en doute la validité (sa cohérence est supposée assurée par évidence). L'arithmétique (classique) et a fortiori l'algèbre dépasse le domaine « contentuel » [Hilbert 1922 (1992), p. 118-119] et [Hilbert 1928 (1992), p. 151-152].

Les énoncés « idéaux » (ou « abstraits ») s'ajoutent, pour Hilbert, aux énoncés « réels » de manière à constituer l'ensemble des mathématiques [Hilbert 1928 (1992), p. 153]. D'une manière générale, les énoncés « idéaux » incluent tous les énoncés présupposant l'infini actuel (au minimum, un processus infini) mais en un sens large dont la circonscription exacte reste à préciser. En pratique, ils constituent presque toutes les mathématiques existantes. Ils n'ont pas de sens intuitif (ils sont d'ailleurs refusés par les intuitionnistes) mais sont introduits pour le gain pratique et l'élégance qu'ils apportent (ils permettent en particulier d'étendre le principe du tiers exclu à des collections infinies) [Hilbert 1926 (1972), p. 231] et [Hilbert 1926 (1972), p. 232]. Les énoncés (informels) « idéaux » (en particulier les propositions faisant appel au

³⁰ Philip Kitcher pose la question légitime de la justification épistémologique de ce choix [Kitcher 1976, p. 110-111] : comment assurer que la somme des éléments constitutifs constitue bien un corps de connaissance intersubjectif ? Mais ce n'est pas une question à laquelle Hilbert a voulu (ou prétendu) répondre.

³¹ En parallèle avec son refus de donner la définition précise d'une proposition finitiste, Hilbert n'a jamais non plus fourni la limite de ce qu'il entendait par énoncés « réels ».

³² Selon Hourya Sinaceur, « contentuel » signifie pour Hilbert : « qui a un modèle fini » [Sinaceur 1998, p. 139]. Mais, toujours selon Hourya Sinaceur, dans la mesure où on ne peut appréhender « concrètement » ou complètement que des structures finies, « contentuel » devient synonyme d'intuitif.

transfini) sont en effet dénués de sens « concret ». Leur statut est très particulier : ce ne sont ni des objets ni de pures fictions mais des « quasi objets », entités virtuelles dont nous n'avons d'expérience que par l'esprit [Hilbert 1926 (1972), p. 221]. Avec cette position, Hilbert s'oppose ainsi à la fois aux intuitionnistes qui en refusent l'existence et aux réalistes « platoniciens » qui les considèrent comme des objets réels. Comme nous l'avons déjà indiqué, l'origine de ce choix est avant tout pragmatique³³. Ces énoncés ne sont donc *ni vrais ni faux* (au sens intuitif). Le sens concret est en fait retrouvé par *formalisation* des mathématiques (« réplique » des mathématiques), l'établissement d'un système formel (au sens moderne), dont les symboles constituent bien des signes « concrets ». La dérivation au sein d'un système formel constitue alors un assemblage « concret » de signes « concrets », un « dessin » sur lequel on peut raisonner, ce qui est l'objet de la *métamathématique*³⁴ [Hilbert 1926 (1972), p. 233, 234] et [Hilbert 1928 (1992), p. 147]. Toutefois, comme il n'est alors plus possible de s'appuyer sur l'intuition, leur manipulation doit être effectuée au travers de règles mécaniques d'inférence (des règles « sans pensée » en un sens que nous allons préciser) et leur cohérence doit être *démontrée* [Hilbert 1926 (1972), p. 233, 235]. Cette démonstration doit être elle-même réalisée au moyen d'une démarche finitiste. On obtient alors des propositions vraies (au sens intuitif) pour des systèmes dont la nature est indifférente³⁵. Pour tenter de clarifier une situation certainement complexe, nous proposons ci-après un schéma général des liens entre les différents concepts introduits par Hilbert selon notre reconstitution (schéma A).

³³ Il a été associé à un engagement instrumentaliste, en particulier par Philip Kitcher : [Kitcher 1976], et surtout par Michael Detlefsen dans [Detlefsen 1986] et [Detlefsen 1990], association contestée par Michael Hallet [Hallet 1990, p. 226].

³⁴ Comme le remarque Jacques Dubucs, il n'y a pas de remontée infinie : d'une mathématique vers une métamathématique, puis d'une métamathématique vers une « méta-métamathématique », ... [Dubucs 1979, p. 118]. En effet, la métamathématique repose elle-même sur un domaine concret, qui constitue le socle de base de l'édifice.

³⁵ Comme le fait remarquer Jacqueline Boniface, les mathématiques n'en sont pas pour autant réduites à un jeu vide de sens selon la critique habituelle faite au formalisme [Boniface 2004, p. 270-272]. Même les mathématiques idéales sont en effet ancrées sur le concret à travers l'expression de leur formalisation et pour l'établissement de leur non-contradiction.

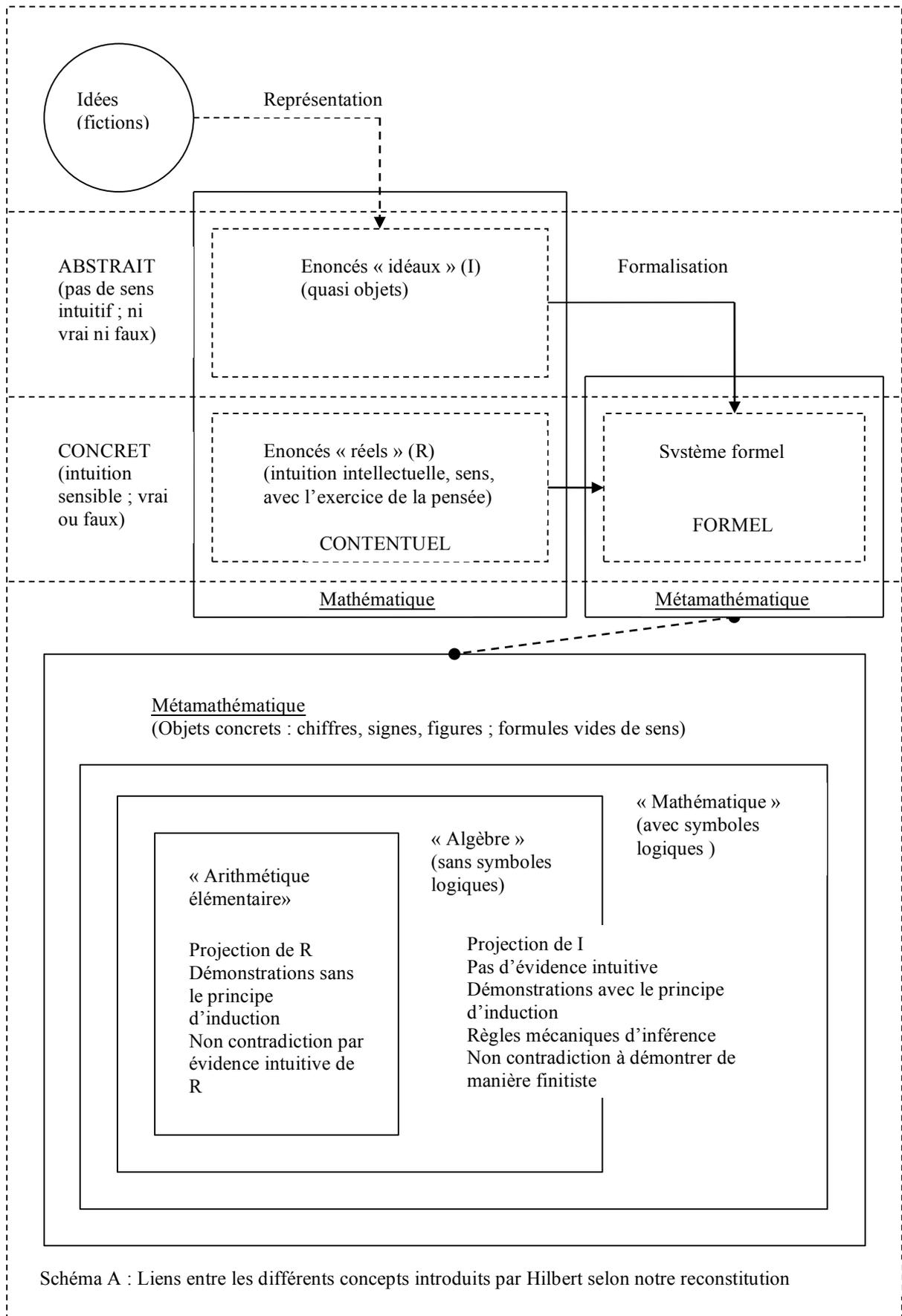


Schéma A : Liens entre les différents concepts introduits par Hilbert selon notre reconstitution

Le cœur du programme de Hilbert au sens large, qui est la clef de voûte de toute sa construction selon l'expression de Hilbert lui-même [Hilbert 1926 (1972), p. 236] et qui correspond à ce que nous avons appelé le programme *au sens restreint*, consiste ainsi à fournir une preuve *finitiste* de non-contradiction d'un système formel formalisant l'ensemble formé par les énoncés « réels » et les énoncés « idéaux » des mathématiques, en s'appuyant sur l'évidence intuitive des énoncés « réels » vrais. Plus précisément, eu égard aux résultats de non-contradiction relative obtenus par ailleurs, l'objectif se limite à fournir une preuve « absolue » de non-contradiction (du système formel) de la seule arithmétique *du deuxième ordre*³⁶ [Hilbert 1926 (1972), p. 236]. De cette manière le Programme répondait à l'objectif d'établir définitivement la sûreté de l'ensemble des mathématiques sans en exclure aucune partie (en conservant le « paradis de Cantor ») [Hilbert 1926 (1972), p. 227] et [Introduction de Hilbert, Hilbert, Bernays 1934 (2001), p. 43]. Il conduisait par ailleurs à une formulation de la pensée mathématique sans avoir besoin de reconnaître la réalité de tous ses objets. En effet, le finitisme de Hilbert conduit à une neutralité ontologique concernant les objets « idéaux », l'engagement ontologique se limitant aux seuls énoncés « réels ». Enfin, la problématique des « fondements » ne relevait alors plus de la spéculation philosophique, mais de la seule preuve mathématique [Bernays 1922b, p. 221].

Il s'avère cependant que l'on ne trouve pas de description précise du contenu du programme au sens restreint dans les écrits de Hilbert et de Bernays. Les seuls éléments pratiquement disponibles sont une mention de Hilbert qui sera très abondamment reprise par tous les commentateurs, ainsi que deux autres de Bernays :

[...] une condition, une seule, mais qui est absolument indispensable à laquelle est liée l'emploi de la méthode des éléments idéaux, à savoir l'obligation de donner une *démonstration de non contradiction*. Une extension par adjonction d'éléments idéaux n'est recevable que s'il n'en résulte pas de contradiction à l'intérieur de l'ancien domaine, c'est-à-dire par conséquent si les relations entre les anciens objets en termes desquels on élimine les objets idéaux sont toujours valables dans l'ancien domaine.
[Hilbert 1926 (1972), p. 235]

³⁶ Il convient de rappeler une ambiguïté dans la formulation des écrits de Hilbert et de Bernays que nous avons déjà mentionnée, qui font état « d'arithmétique » (sous-entendu pour eux du *deuxième* ordre), pour désigner à la fois, la théorie des nombres, l'analyse, et la théorie des ensembles [Hilbert, Bernays 1934 (2001), p. 57]. Michael Potter [Potter 2000, p. 241] justifie alors la limitation à l'arithmétique du 1^{er} ordre dans les tentatives de réalisation du Programme par l'école de Hilbert, par une attitude pragmatique : « commencer par cela » eu égard aux difficultés techniques rencontrées (illustrées par la reconnaissance tardive de la limitation de portée du théorème d'Ackermann 1924, sur laquelle nous reviendrons au § 4.1). Cela est plus ou moins confirmé par Hilbert et Bernays [Hilbert, Bernays 1934 (2001), p. 75]. Dans le même esprit, Stephen Kleene suggère qu'en 1931 Gödel a quant à lui choisi de diviser la difficulté de la démonstration de consistance de l'analyse, en commençant par celle de la théorie des nombres [Kleene 1986a, p. 127].

Mais la tâche de cette preuve [de non contradiction de l'arithmétique] consiste exclusivement à nous procurer la certitude pleine et éclairée que ce système de pensées, déjà motivé par des raisons internes de la systématique et éprouvé le mieux possible dans son efficacité en tant qu'outillage de l'esprit, ne puisse être ruiné par une discordance dans ses conséquences [...] il n'est question dans cette preuve de rien d'autre que de reconnaître la non contradiction, au sens littéral de l'expression, de la Théorie mathématique, i.e. *l'impossibilité de son auto réfutation interne*.

[Bernays 1930/1976 (2003), p. 76 ; le commentaire entre crochets est nôtre]

[...] nous aurions l'assurance que ces méthodes arithmétiques usuelles ne peuvent jamais conduire à un résultat dont on pourrait démontrer la fausseté ; et plus précisément, nous aurions l'assurance que les résultats de l'application de ces méthodes sont d'accord aussi bien les uns avec les autres qu'avec chaque fait évident du point de vue finitiste. Mais ce problème n'est rien d'autre que celui de la non contradiction de notre arithmétique usuelle. [Hilbert, Bernays 1934 (2001), p. 98]

La formulation la plus précise en est incontestablement celle du tout jeune Kurt Gödel, lors de la Table ronde du 7 septembre 1930, à Königsberg :

Cette vision [du finitisme de Hilbert] suppose que si l'on adjoint au système S des formulations pleines de sens, le système T des axiomes et formulations transfinies et alors prouve un théorème de S dans un détour par les théorèmes de T que ce théorème est aussi contentuellement correct et ainsi qu'aucun théorème faux ne devient prouvable à travers l'adjonction des axiomes transfinis.

[Gödel en 1930, dans Dawson 1984b, p. 125 ; le commentaire entre crochets est nôtre]

D'un point de vue purement technique, et en anticipant les discussions à venir, il est utile de commencer par apporter une précision quant au contenu de la consistance en jeu dans le programme au sens restreint. Du fait de vouloir ramener la vérification de consistance de l'ensemble des mathématiques à celle de leur seule partie « réelle », la consistance simple³⁷ est nécessaire tandis que la consistance absolue³⁸ est en principe insuffisante³⁹. En sens inverse, l'expression même du programme n'implique pas de requérir une forme de consistance plus forte, qui serait même contraire à son esprit : ni la consistance sémantique⁴⁰, ni a fortiori l' ω -

³⁷ Nous rappelons la définition de la consistance « simple » (ou non contradiction, ou encore cohérence) : il n'existe pas d'énoncés : φ du système formel, telle qu'à la fois φ et $\neg\varphi$ soient des théorèmes (c'est une définition syntaxique, qui n'a de sens que pour les systèmes formels qui possèdent un opérateur de négation).

³⁸ Nous rappelons la définition de la consistance « absolue » (ou non trivialité) : il existe au moins un énoncé du système formel qui n'est pas un théorème (c'est encore une définition syntaxique).

³⁹ En remarquant qu'en logique classique, pour la plupart des systèmes formels intéressants (incluant la logique des propositions), la consistance simple et la consistance absolue sont équivalentes.

⁴⁰ Nous rappelons que la consistance « sémantique » est définie par le fait qu'il existe au moins une structure d'interprétation du système formel qui est un modèle. Au 1^{er} ordre, elle est équivalente à la consistance simple par le théorème de Gödel de 1930, mais à la suite d'une démonstration non constructiviste (donc non finitiste).

consistance⁴¹. En effet, l' ω -consistance fait intervenir un élément « idéal » sous la forme d'un quantificateur existentiel, et la consistance sémantique fait intervenir la notion de modèle totalement étrangère à la démarche de Hilbert. La position de Hilbert et de Bernays est d'ailleurs sans ambiguïté sur ce point : ils se réfèrent de manière constante à la non-contradiction, c'est-à-dire à la consistance simple⁴². Comment doit être alors traduit le caractère *finitiste* de la démonstration de consistance ? Bernays a été là aussi très clair sur ce point. La démonstration finitiste de consistance implique d'utiliser une forme de correction, mais, comme le souligne Richard Zach [Zach 2003, p. 218-219], qui se réfère à la vérité au sens intuitif⁴³ (parfaitement définie pour la partie réelle) [Hilbert, Bernays 1934 (2001), p. 295]. Cette « correction » est alors exprimée au moyen du critère que Bernays appelle « vérifiable » et qui est défini de la manière très simple suivante : un énoncé « réel » est vérifiable si chacune de ses instances est vraie au sens intuitif [Hilbert, Bernays 1934 (2001), p. 304, 313, 330]. Cela s'applique d'abord à la quantification universelle (au premier ordre) [Hilbert, Bernays 1934 (2001), p. 313], même si Bernays prend bien soin de mentionner que la quantification universelle en tant que telle en est exclue [Hilbert, Bernays 1939 (2001), p. 85, n. 1]. De manière peut-être un peu plus surprenante, l'application est également étendue à la quantification *existentielle* (toujours au premier ordre) mais *avec des restrictions* :

Celui-ci [le théorème de la réduction partielle] dit bien, pour une formule $(\exists x) \mathbf{A}(x)$, qui ne comporte aucune variable de formule ni aucune variable d'individu libre, que cette formule est vérifiable si et seulement si, pour un certain code \mathbf{z} , la formule $\mathbf{A}(\mathbf{z})$ est vérifiable, la découverte de \mathbf{z} découlant de la réduction de cette formule dans le cas de la vérifiabilité de $(\exists x) \mathbf{A}(x)$.

[Hilbert, Bernays 1934 (2001), p. 330 ; le commentaire entre crochets est nôtre]

⁴¹ Nous rappelons que l' ω -consistance est elle-même définie par l'exigence suivante (avec les notations métamathématiques modernes habituelles, où « \vdash » renvoie à la dérivabilité) : il n'y a pas d'énoncé $\varphi(x)$ telle que l'on ait à la fois, d'une part $\vdash \varphi(0), \vdash \varphi(1), \vdash \varphi(2), \dots$, et d'autre part $\vdash \neg \forall x \varphi(x)$. Bien que sa forme soit purement syntaxique, on montre qu'elle est équivalente à l'exigence d'existence d'un modèle *naturel* (standard) du système formel. C'est donc une condition plus forte que celle de la consistance sémantique.

⁴² Comme l'ont fait remarquer de nombreux commentateurs de Hilbert, par exemple : Jacques Dubucs [Dubucs 1979, p. 93] et [Blanché, Dubucs 1996, p. 379], John Dawson [Dawson 1984b, p. 126] ainsi que Stephen Kleene [Kleene, 1986a, p. 135], on conçoit que ce choix puisse poser un problème qui n'avait certainement pas été imaginé par Hilbert, dans le cas de systèmes formels (théories) « pathologiques » consistants mais ω -inconsistants, qui n'ont pas de modèle naturel. Nous reviendrons sur ce point, qui a été clairement perçu a posteriori (c'est-à-dire après Gödel 1931) par Bernays lui-même [Hilbert, Bernays 1939 (2001), p. 344, 345].

⁴³ Nous rappelons en effet qu'au sens moderne, la correction d'un système formel est définie, soit par rapport à la validité, soit par rapport à la vérité dans un modèle, en exprimant que tout théorème du système formel est valide (pour tout énoncé φ du système formel, si $\vdash \varphi$ alors $\models \varphi$, où « \models » renvoie à la conséquence logique), respectivement vrai dans le modèle considéré.

Comment se ramène-t-on alors à l'objectif même du Programme, qui porte sur la consistance simple d'un système formel *de l'ensemble des mathématiques* (« réelles » et « idéales »), et non seulement des mathématiques « réelles » ? Tout simplement par l'effet de « contagion » en logique classique (*ex falso sequitur quodlibet*) qui fait que toute éventuelle contradiction dans la seule partie « idéale » se répandrait inéluctablement jusqu'à « contaminer » la partie « réelle ». La prise en compte du critère de vérifiabilité conduit ainsi à transformer la formulation du Programme, qui devient : montrer de manière finitiste que tout théorème de la partie « réelle » de l'arithmétique du deuxième ordre, obtenu au sein d'un système formel formalisant l'ensemble des mathématiques « réelles » et des mathématiques « idéales » (donc possiblement via des axiomes et des règles « idéaux »), est vérifiable.

Dans cette formulation, au-delà de la nécessité de préciser le contenu de la partie « réelle », il reste encore à définir ce qui est entendu par « de manière finitiste ». Sur ce point, Hilbert et Bernays disent très peu de choses : ils mentionnent seulement le caractère communicable d'une preuve du début à la fin [Bernays 1922a (1998), p. 195], et la nécessité d'évaluer à l'avance le nombre maximum de pas nécessaires [Hilbert 1928 (1992), p. 162]. En nous fondant sur l'esprit de l'ensemble de leurs propos, on peut sans doute résumer les caractéristiques générales requises pour toute démonstration finitiste en général⁴⁴ par deux exigences :

- 1) celle-ci ne doit pas contenir d'éléments « idéaux », donc ne doit pas contenir de quantificateur (pas même le quantificateur universel) ;
- 2) elle doit se traduire par un algorithme à aboutissement certain en un nombre fini d'étapes, sur la base de règles fixées à l'avance (ce qui correspond aux critères modernes d'effectivité que l'on applique à un système formel).

Dans le cas particulier, qui concerne directement le Programme, d'une démonstration finitiste de *consistance* d'un système formel formalisant l'arithmétique du 2^{ème} ordre, les éléments mentionnés par Hilbert et Bernays se réfèrent exclusivement à la démonstration d'Ackermann de 1924 (sur laquelle nous reviendrons au § 4.1). La démarche évoquée se ramène alors à quatre étapes :

- 1) montrer que les axiomes du système formel sont vérifiables ;
- 2) montrer que les règles de dérivation conservent le caractère vérifiable ;
- 3) établir directement que la dérivation que « $0 \neq 0$ » ne peut pas être un théorème du système formel ;

⁴⁴ Sachant que la nécessité de définir le contenu d'une démonstration finitiste *en général*, n'est pas une requête du Programme (qui ne porte que sur une démonstration finitiste *de consistance*).

4) éliminer dans cette dérivation, tous les éléments non finitistes, en se ramenant à une dérivation ne contenant que des éléments « réels ».

Cette démarche est bien résumée par Hilbert :

La démonstration habituelle de la consistance consiste à transformer, conformément à une procédure bien déterminée, toutes les formules de la démonstration donnée en des formules numériques ; ensuite il faut établir que toutes les formules initiales sont vraies. [Hilbert 1931 (1992), p. 194]

Cependant, au-delà de la référence explicite à la démonstration d'Ackermann de 1924⁴⁵, et sans doute avec une référence implicite à la démonstration de Presburger de 1929⁴⁶, aucune indication plus précise à caractère général n'est donnée sur le contenu des étapes (3) et (4), évidemment cruciales. Or, en étant simplement passé de : « montrer de manière finitiste » à : « suivre une procédure bien déterminée », on peut se demander si l'on a vraiment progressé. Il est clair que pour aller plus loin, il est nécessaire de définir de manière précise le contenu, des énoncés réels d'une part, et de la procédure finitiste qui doit être associée au critère de vérifiabilité pour conduire à une démonstration finitiste de consistance, d'autre part, ce qui revient à proposer une interprétation technique du Programme.

Comme nous l'avons indiqué, la consistance (informelle) de l'ensemble des théorèmes « réels » est admise a priori sur la base de leur caractère intuitif. Cependant, si l'on introduit un système formel spécifique pour formaliser cette seule partie « réelle »⁴⁷, il faut encore bien évidemment s'assurer au préalable du caractère vérifiable de l'ensemble de ses théorèmes, et ce de manière finitiste. Une telle exigence devient alors une *condition de possibilité* du Programme. Par ailleurs, chaque énoncé « réel » étant intuitivement connu comme vrai ou faux, la décidabilité (informelle) de l'ensemble de ces énoncés est assurée, comme l'indique Hilbert sans ambiguïté [Hilbert 1926 (1972), p. 232-233]. Dans le même cas où l'on introduit un système formel spécifique pour formaliser cette seule partie « réelle », la question se pose donc de savoir si l'on doit ou non exiger que ce système formel soit non seulement simplement

⁴⁵ Qui ne porte en fait que sur un domaine partiel et limité de l'arithmétique du 2^{ème} ordre, qui plus est (en anticipant sur l'analyse qui sera effectuée au § 4.1) dont le caractère finitiste de la démonstration est problématique.

⁴⁶ Même si celle-ci n'est pas une démonstration de consistance, mais de complétude syntaxique (dont la définition est rappelée à la note 48) d'un système formel de l'arithmétique du 1^{er} ordre sans l'opérateur de multiplication. Elle utilise cependant une technique analogue à celle qui vient d'être évoquée, conduisant à supprimer la quantification qui intervenait naturellement.

⁴⁷ Tel que le Programme est formulé, il n'y apparaît en effet aucune nécessité de disposer d'un tel système formel *spécifique* pour traduire la seule partie réelle des mathématiques. De fait, Hilbert n'a jamais formulé une telle exigence.

consistant mais également syntaxiquement complet⁴⁸. Si tel devait être le cas, comme cela est suggéré, explicitement par Stephen Kleene [Kleene 1976 (1990), p. 51] et implicitement par Gödel selon Solomon Feferman [Feferman 1995a, p. 38], une telle exigence constituerait alors une seconde *condition de possibilité* du Programme. Ce point sera donc à préciser sur la base de l'interprétation technique qui sera retenue pour le Programme.

1.2 Justification de la prise en compte de l'ensemble des FRP et seulement elles, comme fonctions admises dans l'arithmétique « réelle »

Les fonctions les plus élémentaires de l'arithmétique : l'addition et la multiplication font indiscutablement partie de l'ensemble R des mathématiques « réelles » sans limitation quant à leur portée. Elles ont été mentionnées explicitement par Hilbert et Bernays à de nombreuses reprises et sont acceptées par tous les commentateurs. Elles sont traduites dans les langages des deux principaux systèmes formels de l'arithmétique : PA⁴⁹ et PRA. Pour des fonctions plus complexes, la justification n'est plus aussi évidente et de fait le consensus n'est plus général. L'exponentiation a été jugée non compatible avec le finitisme par Charles Parsons : [Parsons 1998] et Michael Potter : [Potter 2000]. Jeremy Avigad l'accepte mais ne va pas au-delà, en refusant donc l'ensemble des FRP⁵⁰ : [Avigad 2003]. Toutefois, à la différence des précédents, son choix⁵¹ n'est pas présenté comme une remise en cause du finitisme « standard ». Enfin, Georg Kreisel a suggéré au contraire d'aller au-delà des FRP.

Charles Parsons⁵² accepte comme fonctions finitistes, sans hésitation l'addition (illimitée), à la rigueur la multiplication (illimitée), mais refuse l'exponentiation (illimitée) [Parsons 1998, p. 260-262]. Il s'appuie pour cela sur [Hilbert, Bernays 1934] avec deux arguments. Il fait d'abord valoir que la croissance très rapide de l'exponentielle conduit à des nombres qui ne sont très vite physiquement plus représentables (en évoquant l'interrogation de Bernays sur

⁴⁸ Nous rappelons que la complétude syntaxique est définie de la manière suivante : pour tout énoncé ϕ du système formel, ou bien $\vdash \phi$ ou bien $\vdash \neg\phi$.

⁴⁹ PA (pour « Peano Arithmetic » en anglais) désigne la théorie (moderne) du calcul des prédicats du 1^{er} ordre appliquée à l'arithmétique récursive, issue des travaux de Peano de 1889. Il est plus rarement également dénommé Z_1 (par exemple dans [Hilbert, Bernays 1939]) ou encore Z.

⁵⁰ Plus précisément il se limite aux fonctions élémentaires de Kalmar, pour lesquelles il précise : « [...] grossièrement, ce sont les fonctions qui peuvent être calculées dans un temps et/ou un espace borné par une itération fixée de la fonction exponentielle. » [Avigad 2003, p. 263].

⁵¹ Son objectif, sur lequel nous serons amenés à revenir, est d'introduire un système formel intéressant plus faible que PRA.

⁵² Michael Potter lui a emboîté le pas sur ce point dans [Potter 2000].

l'évidence du caractère intuitif d'un nombre comme $67^{257^{229}}$). L'argument principal est cependant basé sur le mode de construction de la fonction exponentielle qu'il prend en compte : par récursion. En s'appuyant sur une justification par Hilbert et Bernays de la récursion primitive, il en vient à affirmer que celle-ci impose le recours à l'induction \sum_1^{53} qui n'est pas finitiste [Parsons 1998, p. 262-264]⁵⁴. Le premier argument cité n'est évidemment pas recevable, d'abord parce que l'auteur ne voit apparemment pas qu'un problème identique se pose déjà dans le cas d'un nombre entier suffisamment grand (même s'il est évidemment amplifié dans le cas de l'exponentiation), alors qu'il a pourtant accepté le principe de l'addition de nombres entiers sans limitation. Surtout, il revient à introduire des considérations « en pratique », alors que Hilbert ne prend en compte pour le finitisme que celles « en principe »⁵⁵ :

L'abstraction intuitive ne fait aucun cas de telles limites [10^{10} et a fortiori $10^{10^{20}}$] à la possibilité de réalisation. Car les limites sont, du point de vue de la considération formelle, contingentes.

[Bernays 1930/1976 (2003), p. 62 ; le commentaire entre crochets est nôtre]

L'argument principal nous paraît, quant à lui, une manière détournée d'aborder la question posée qui est de savoir si oui ou non la fonction exponentielle (illimitée) est finitairement (intuitivement) acceptable. La question de savoir comment celle-ci peut être construite dans un système formel donné vient seulement après. Dans le cadre strict de la présentation du finitisme effectuée par Hilbert, cette question nous semble pouvoir être traduite de manière pertinente par la suivante : pour tout couple donné de numéraux a et b , traduits chacun par la suite correspondante de bâtons, peut-on rendre compte du numéral associé à $\exp_a(b)$, lui-même exprimé sous la forme d'une nouvelle suite de bâtons, cela uniquement à partir de mouvements visualisables des deux suites de bâtons initiales ? La réponse est évidemment positive, par simple continuation et combinaison des manipulations de suites de bâtons proposées par Hilbert pour l'addition et pour la multiplication. On notera que cette réponse pour l'exponentiation est d'ailleurs explicitement mentionnée par Bernays :

Nous avons de cette façon la procédure de multiplication par dix d'un nombre z . De là nous obtenons la procédure de passage d'un nombre a au nombre 10^a , en faisant correspondre 10 au premier 1 dans a puis à chaque nouveau 1 la procédure de multiplication par dix et ainsi de suite jusqu'à ce que nous obtenions

⁵³ Il s'agit d'une induction faisant intervenir une quantification existentielle non bornée.

⁵⁴ En évoquant le même argument Michael Potter parle, quant à lui, de « récursion emboîtée » [Potter 2000, p. 236-237].

⁵⁵ A notre sens William Tait tombe dans le même travers [Tait 1981, p. 539-540].

à la fin la figure *a*. Le nombre obtenu par la dernière procédure de multiplication par 10, nous l'écrivons 10^a . Ce procédé n'offre principalement aucune difficulté à l'attitude intuitive.
[Bernays 1930/1976 (2003), p. 62]

Qu'il s'agisse de l'addition, de la multiplication ou de l'exponentiation sur des numéraux, ces trois opérations sont donc tout aussi visualisables, la seule différence étant que la première peut l'être sur une seule dimension (disons sur un ruban) par simple concaténation, alors que les deux autres en nécessitent deux (feuille de papier ou tableau). La visualisation s'étend alors à toute combinaison de ces trois opérations, selon le même mode opératoire de fractionnement et de regroupement des suites de bâtons.

La justification de l'affirmation que les fonctions finitistes ne vont pas au-delà des FRP et en particulier n'incluent pas les FRG⁵⁶ totales, soit la seconde partie du premier volet de la thèse de Tait, est plus délicate. En effet, comme l'ont fait remarquer d'une part Wilfried Sieg [Sieg 2003a, p. 9, n. s] et [Sieg 2005, p. 176] et d'autre part Richard Zach [Zach 2007, p. 425-426], les données historiques disponibles sont ambiguës. Celles-ci sont doubles. Elles concernent d'abord les positions affichées par les différents protagonistes juste après la prise de connaissance des théorèmes de Gödel de 1931. Elles consistent ensuite dans les commentaires épars que l'on trouve sur ce sujet dans [Hilbert, Bernays 1934]. Sur le premier point la situation peut être résumée de la manière suivante. En 1931, au cours des échanges épistolaires qui ont suivis la diffusion du résultat de Gödel, aussi bien Gödel lui-même (la remarque incluse à la fin de son article selon laquelle les théorèmes ne remettent pas en cause le programme de Hilbert), que von Neumann (lettre à Gödel du 12 janvier : [Neumann 1931]) et Herbrand (lettre à Gödel du 7 avril : [Herbrand 1931c]), considèrent que le finitisme va au-delà des seules FRP et inclut en particulier la fonction d'Ackermann, donc les FRG. Il faut cependant remarquer qu'en 1931 la notion de FRG n'était pas disponible, et par voie de conséquence les critères d'effectivité que l'on doit exiger d'un système formel n'étaient pas clairement perçus par les protagonistes, faute de disposer d'un équivalent de la thèse de Church-Turing. Bien entendu, cette remarque dépasse la seule question des fonctions finitairement acceptables et concerne celle, plus globale, du système formel qui doit rendre compte de

⁵⁶ Le terme FRG désigne les Fonctions Récursives Générales, introduites par Gödel en 1934 et analysées en détail par Kleene en 1936. C'est une classe de fonctions d'entiers plus large que celle des FRP, tout en conservant l'exigence de pouvoir en déterminer le résultat pour chaque valeur de l'argument. Alors que les FRP sont équivalentes à des possibilités de programmation limitées aux opérations : « si ... alors » et : « allez à ... », soit l'utilisation de boucles ayant seulement un nombre d'étapes fini et annoncé, les FRG sont équivalentes à l'utilisation de la séquence : « allez vers ... », ce qui revient à autoriser des boucles ayant un nombre fini mais non borné d'étapes.

l'arithmétique finitiste (donc « réelle »). Sur le second point, la situation était encore ambiguë en 1934 comme l'illustre le commentaire de Bernays suivant (le seul sur le sujet, sauf omission de notre part) :

Bien sûr, sans enlever à la théorie réursive des nombres [PRA] ce que ses méthodes ont de caractéristique, nous pouvons autoriser certaines *extensions du schéma de récurrence* et du schéma d'induction. [Hilbert, Bernays 1934 (2001), p. 407 ; le commentaire entre crochets est nôtre]

On peut tout de même remarquer que si la construction de fonctions récursives primitives est détaillée d'une manière explicitement finitiste [Hilbert, Bernays 1934 (2001), p. 82], ce n'est pas le cas de celle de fonctions plus générales. Enfin, il faut également signaler la proposition de Georg Kreisel de 1958⁵⁷, selon laquelle les fonctions finitistes pourraient s'étendre à tout ce qui peut être exprimé dans une théorie analogue à PA. Cette proposition a cependant été l'objet d'une volée de bois vert de la part de William Tait [Tait 1981, p. 544-545], entérinée par Richard Zach [Zach 2007, p. 427], dont il est peu vraisemblable qu'elle puisse se remettre. Finalement et bien qu'il ne s'agisse certainement pas d'un argument dirimant comme le souligne Richard Zach⁵⁸ [Zach 2007, p. 425], nous retiendrons la position de limiter les fonctions « réelles » aux seules FRP, dans la mesure où ces fonctions sont les seules dont la justification finitiste (au sens de Hilbert) est indiscutable⁵⁹.

1.3 Justification de la prise en compte des seuls énoncés Δ_0 et Π_1 comme énoncés de l'arithmétique « réelle »

Nous avons à définir les énoncés « réels », c'est-à-dire ceux qui appartiennent en propre à l'ensemble R des mathématiques « réelles » au sens de Hilbert. Les énoncés de l'ensemble R étant uniquement arithmétiques, pour définir la limite de R, la hiérarchie de Kleene s'avère particulièrement opportune. Nous allons donc en balayer successivement les éléments constitutifs. La prise en compte de l'ensemble des énoncés de type Δ_0 comme éléments « réels » est évidente, dans la mesure où ceux-ci traduisent à peu près littéralement les différentes

⁵⁷ Mentionnée à la fois par William Tait [Tait 1981, p. 545], Charles Parsons [Parsons 1998, p. 250] et Richard Zach [Zach 2007, p. 426].

⁵⁸ Celui-ci mentionne un autre argument, en faisant remarquer que les élèves d'Hilbert ont utilisé des méthodes qui dépassaient la récursivité primitive dans les différentes démonstrations de consistance qui ont été établies avant 1931. Nous y viendrons au § 4.1 en abordant les résultats partiels obtenus avant 1931.

⁵⁹ En effet, pour toute FRP $f(x)$ et pour tout entier a , il est possible de calculer $f(a)$ en un nombre fini d'étapes.

formulations de Hilbert que nous avons évoquées. En effet, ayant assimilé les fonctions finitistes aux FRP, comme les formules $\varphi_0(x_i)$ que nous avons introduites pour définir la hiérarchie de Kleene sont des combinaisons logiques n'utilisant que des quantificateurs bornés d'équations ou d'inéquations arithmétiques ne faisant intervenir que des FRP, ces formules sont nécessairement finitistes. Comme les énoncés Δ_0 sont les instances des formules $\varphi_0(x_i)$, ils sont eux-mêmes finitistes. Un tel choix minimal est donc non problématique et a de fait été accepté par tous les commentateurs de Hilbert. De la même manière, le *refus* de prendre en compte les énoncés de type Σ_1 , traduit naturellement l'exigence réitérée d'Hilbert d'exclure toute totalité infinie (l'infini actuel) de l'arithmétique « réelle ». L'exclusion des énoncés correspondant de fait aux énoncés Σ_1 est d'ailleurs explicite de la part de Hilbert et de Bernays⁶⁰ [Hilbert 1931 (1992), p. 194] et [Bernays 1930/1976 (2003), p. 58]. Bien entendu, cette exclusion des énoncés Σ_1 s'étend à tous ceux de type supérieur ($\Sigma_2, \Sigma_3, \dots$). Là encore, un tel choix a été entériné par presque tous les commentateurs de Hilbert. En sens inverse, la prise en compte des énoncés Π_1 comme éléments « réels » ne présente pas du tout le même degré d'évidence, et ce sont en fait ces énoncés qui ont cristallisé les discussions sur le contenu des énoncés de R (voir par exemple William Tait [Tait 1981, p. 524] et Richard Zach [Zach 2004, p. 89]). Cela s'est traduit en particulier quelquefois par le *refus* de les prendre en compte [Niebergall, Schirn 1998, p. 294] et toujours par l'affirmation du caractère « problématique » de ces énoncés [Detlefsen 1986, p. 352-353], [Giaquinto 2002, p. 145-147] et [Zach 2004, p. 90]. A l'appui de ce refus ou de cette réserve sont mentionnés deux arguments : 1) d'abord le fait que de tels énoncés ne sont pas clos par application de toutes les lois de la logique classique, en particulier la négation⁶¹, 2) ensuite l'affirmation que de tels énoncés font implicitement appel à la totalité infinie des nombres naturels, en contradiction avec les exigences de Hilbert.

Que l'ensemble des énoncés Π_1 ne soit pas clos pour la négation n'est pas niable, c'est une complication reconnue par Hilbert [Hilbert, Bernays 1934 (2001), p. 88] qu'il faut prendre en compte⁶². Cela ne nous paraît toutefois pas une raison suffisante pour appliquer une solution

⁶⁰ Nous avons pourtant précédemment mentionné l'ouverture par Bernays du finitisme à des énoncés de type Σ_1 *sous conditions* [Hilbert, Bernays 1934 (2001), p. 330], la même idée étant évoquée par Gentzen [Gentzen 1936a (1969), p. 164-165]. Mais c'est justement la condition de disposer d'un moyen finitiste qui n'est pas donné d'emblée, pour définir les instances d'un énoncé Σ_1 qui sont vérifiées, qui n'est pas remplie par les énoncés Σ_1 d'une manière générale. L'introduction éventuelle d'un tel moyen constituerait évidemment une extension de l'interprétation finitiste retenue, ce qui semble avoir été envisagé par Gödel [Sieg, Parsons 1995, p. 65].

⁶¹ Nous rappelons que la négation d'un énoncé Π_1 conduit à un énoncé Σ_1 , qui n'appartient donc pas à R compte tenu de ce qui vient d'être dit.

⁶² Michael Detlefsen considère également « problématiques » certains énoncés Σ_0 comme : « $\exists n$ premier tel que $p < n < p! + 1$ » dans la mesure où leur généralisation naturelle conduit à un énoncé « idéal », en

aussi radicale que leur élimination pure et simple, à l’instar de ce que préconisent Karl-Georg Niebergall et Matthias Schirn. Nous ne pouvons non plus suivre Craig Smorynski : [Smorynski 1977] et [Smorynski 1988], qui propose une tripartition des énoncés en introduisant une catégorie spéciale pour les énoncés \prod_1 : les énoncés « généraux finitistes », qui ne seraient donc ni « réels » ni « idéaux ». Cela ne nous paraît en effet pas compatible avec l’esprit du programme de Hilbert, pour lequel ou bien un énoncé est intuitif (ou peut se ramener à un tel énoncé) et susceptible d’un attribut de vérité ou bien il ne l’est pas (se reporter au schéma A du § 1.1). Nous rejoignons ainsi sur ce point la conclusion de la critique détaillée de la position de Craig Smorynski effectuée par Michael Detlefsen [Detlefsen 1990, p. 347-357].

Quant à l’objection principale de l’appel implicite des énoncés \prod_1 à une totalité infinie, il se trouve qu’Hilbert y a répondu explicitement avec une réponse simple, mais cruciale pour l’ensemble de notre réflexion. De tels énoncés doivent être compris exclusivement comme un schème, un prototype d’énoncé, un *jugement hypothétique*, en suivant alors la position intuitionniste [Weyl 1927-1949, p. 51]. Herbrand et Bernays sont sans ambiguïté sur ce point :

Nous entendons par raisonnement intuitionniste [finitiste] un raisonnement qui satisfait aux conditions suivantes : [...] ; on n’y considère jamais l’ensemble de tous les objets x d’une collection infinie ; et quand on dit qu’un raisonnement (ou un théorème) est vrai pour tous ces x , cela signifie que pour chaque x pris en particulier, il est possible de répéter le raisonnement général en question, qui ne doit être considéré que comme le *prototype* de ces raisonnements particuliers.

[Herbrand 1931b, p. 225, n. 3 ; le commentaire entre crochets est nôtre ; souligné par nous]

Un jugement *universel* sur des codes [numéraux] ne peut être interprété de façon finitiste que dans un sens hypothétique, c’est-à-dire comme un énoncé sur n’importe quel code [numéral] donné. Un tel jugement exprime une loi qui doit se trouver vérifiée dans chaque cas isolé qui se présente.

[Hilbert, Bernays 1934 (2001), p. 87 ; les commentaires entre crochets sont nôtres]

La formulation : $\forall x_1, x_2, \dots, x_n \varphi_0(x_1, x_2, \dots, x_n)$ n’y signifie donc pas autre chose que l’affirmation *conditionnelle* suivante : si l’on affecte un ensemble quelconque de numéraux a_1, a_2, \dots, a_n à la formule générique φ_0 , alors celle-ci fournit l’énoncé $\varphi_0(a_1, a_2, \dots, a_n)$. Comme nous avons montré que les formules $\varphi_0(x_i)$ sont finitistes, tout énoncé \prod_1 étant défini par la relation $\forall x_i \varphi_0(x_i)$, compris dans le sens de Hilbert, il est nécessairement finitiste. En aucun cas l’on ne rend compte ainsi de l’ensemble infini des valeurs possibles de la formule générique φ_0 , ni des

l’occurrence \sum_1 [Detlefsen 1990, p. 353]. Cette nuance nous paraît artificielle. Dans l’exemple donné, si p est une constante, l’énoncé est Δ_0 et « réel », si p est une variable, l’énoncé est \sum_1 et « idéal ».

fonctions qu'elle contient⁶³ : on se limite à l'infini potentiel en excluant l'infini actuel⁶⁴ (la distinction entre ces deux formes d'infini par Hilbert et par Bernays est explicite [Hilbert 1926 (1972), p. 225] et [Bernays 1930/1976 (2003), p. 59]). La formulation $\forall x_i \varphi_0(x_i)$ ne doit donc être prise que comme un raccourci de notation, qui peut également se traduire par l'emploi de variables libres : $\varphi_0(x_i)$ ⁶⁵. Si l'on s'intéresse à la dérivabilité d'un énoncé de la forme $\varphi_0(x_i)$ dans un système formel donné, on n'accède qu'à ses instances successives, mais non à la quantification universelle de celles-ci. On accède ainsi éventuellement (lorsque chaque instance est dérivable) à : $\vdash \varphi_0(1), \vdash \varphi_0(2), \vdash \varphi_0(3), \dots$, donc à $\forall x (\vdash \varphi_0(x))$ mais *non* à $\vdash \forall x \varphi_0(x)$ ⁶⁶. Cette signification très particulière attribuée aux énoncés Π_1 , que les commentateurs du programme de Hilbert n'ont pas toujours à l'esprit⁶⁷, nous paraît ainsi une réponse dirimante aux objections « infinitaires » qui ont été avancées contre la prise en compte de ces énoncés parmi les énoncés « réels » à l'instar d'Edward Nelson [Nelson E. 1992, p. 157]. Il n'est par contre pas possible d'étendre une telle signification aux énoncés de types supérieurs (Π_2, Π_3, \dots), dans la mesure où ceux-ci font nécessairement intervenir au moins une quantification existentielle non bornée. Ces derniers énoncés seront donc quant à eux *exclus* des énoncés « réels »⁶⁸. Finalement, les énoncés « réels » sont donc uniquement les énoncés de type Δ_0 et Π_1 , ces derniers étant pris dans le sens d'un « jugement hypothétique ». Pour n'être pas unanime, ce choix est tout de même celui de la grande majorité des commentateurs de Hilbert (voir [Zach 2004, p. 91]). Il nous semble surtout tout à fait en accord avec les exigences précises synthétisées a posteriori par Bernays :

⁶³ Une autre manière possible d'exprimer cette convention serait d'affirmer, en suivant Michael Detlefsen [Detlefsen 1990, p. 354], que $\forall x_i \varphi_0(x_i)$ est « idéal » pris dans sa généralité et « réel » en se limitant à chacune de ses instances. Mais cette formulation est une source possible de confusion et d'erreur. Nous ne la retiendrons donc pas.

⁶⁴ C'est justement parce que Karl-Georg Niebergall et Matthias Schirn refusent l'infini potentiel au sein du finitisme de Hilbert [Niebergall, Schirn 1998, p. 293, n. 32], qu'ils sont légitimement amenés à refuser les énoncés Π_1 au sein de R , même avec la signification limitée qui vient d'être introduite pour ceux-ci.

⁶⁵ Bernays souligne d'ailleurs le risque d'erreur en cas d'oubli du fait qu'il ne s'agit que d'un raccourci de notation [Hilbert, Bernays 1939 (2001), p. 336, n. 4]. Il cite l'exemple très simple de la relation avec variable libre : $x = 0$, qui peut être traduite par l'énoncé : $\forall x (x = 0)$, mais dont la négation n'est alors pas : $x \neq 0$, qui pourrait se traduire de la même manière par l'énoncé : $\forall x (x \neq 0)$, mais l'énoncé : $\exists x (x \neq 0)$.

⁶⁶ Le passage de $\forall x (\vdash \varphi_0(x))$ à $\vdash \forall x \varphi_0(x)$ n'est autre que l'exigence associée, d'une manière générale (sans se limiter aux énoncés Π_1), à la définition de l' ω -complétude. Nous rappelons en effet que l' ω -complétude est définie par la relation suivante : si $\vdash \varphi(0), \vdash \varphi(1), \vdash \varphi(2), \dots$, alors $\vdash \forall x \varphi(x)$. En l'occurrence, en se limitant aux énoncés Π_1 , il s'agit en fait de la 1-complétude, où les instances de $\varphi(x)$ sont limitées à des énoncés Δ_0 .

⁶⁷ On la retrouve néanmoins, explicitement dans : [Resnik 1974, p. 120-121, n. 10], [Prawitz 1981, p. 255-256] et [Mancosu 1998, p. 161-162, 166-167] et implicitement dans : [Giaquinto 2002, p. 146-147] et [Zach 2004, p. 90].

⁶⁸ Bernays donne un exemple d'énoncé finitiste, qui se présente sous la forme d'un énoncé Π_2 [Hilbert, Bernays 1939 (2001), p. 414] : voir la citation qui suit. Mais une analyse attentive des caractéristiques qui lui sont associées montre que celui-ci se ramène en fait à un énoncé Π_1 (la quantification existentielle étant bornée).

Le concept originel étroit d'énoncé finitiste, dans le domaine de la théorie des nombres, revient à n'admettre, comme énoncés de la théorie finitiste des nombres, que les suivants : ceux qui se présentent dans le formalisme de la théorie des nombres, sous adjonction éventuelle de symboles de certaines fonctions calculables de la théorie des nombres, mais sans usage de variable de formule [élimination du deuxième ordre], ou bien ceux qui sont susceptibles d'une interprétation renforcée par un énoncé de cette forme.

Aux énoncés de la première espèce appartiennent bien ceux de la forme : "s'il y a un nombre x qui a la propriété $A(x)$, alors on a B "; en effet (si la variable a ne figure pas dans B) un tel énoncé est représentable par la formule $A(a) \rightarrow B$.

Des exemples d'énoncés de la deuxième espèce sont ceux de la forme : "pour tout nombre x , il y a un nombre y tel qu'on ait $A(x,y)$ " pour lesquels nous disposons d'une interprétation renforcée, fournie par un énoncé représentable sous la forme $A(c, \varphi(c))$ où $\varphi(c)$ est une certaine fonction calculable.

Par contre, n'appartiennent pas aux énoncés finitistes, au sens indiqué, des propositions où une proposition universelle se trouve comme conditionnelle.

[Hilbert, Bernays 1939 (2001), p. 414 ; le commentaire entre crochets est nôtre]

1.4 Définition d'une procédure finitiste conduisant à la démonstration de consistance des théorèmes de la partie « réelle » de l'arithmétique du deuxième ordre

Le troisième volet de la thèse de Tait selon son interprétation standard vise à donner un contenu technique à la notion de preuve finitiste (en général). Nous avons déjà défini les caractéristiques de la procédure correspondante au § 1.1. Celle-ci ne doit d'abord pas contenir d'élément « idéal », ce qui exclut en particulier tout quantificateur : première exigence. Cette procédure doit ensuite se traduire par un algorithme à aboutissement certain en un nombre fini d'étapes, sur la base de règles fixées à l'avance (ce qui correspond à la notion moderne d'effectivité) : seconde exigence. Marcus Giaquinto exprime cela d'une manière particulièrement claire :

Quand une prétendue preuve de proposition générale finitaire [énoncé Π_1] est-elle acceptable du point de vue finitiste ? La réponse est : quand nous avons une procédure finitaire pour obtenir à partir de la prétendue preuve, une preuve finitaire de toute instance donnée de la proposition générale.

[Giaquinto 2002, p. 155 ; le commentaire entre crochets est nôtre]

Le troisième volet de la thèse de Tait proposé par celui-ci [Tait 1981], est une réponse globale à ces exigences qui peut se résumer par la double affirmation forte suivante : 1) les preuves finitistes se confondent avec les dérivations dans le système formel PRA (et par voie de conséquence, les théorèmes finitistes se confondent avec les théorèmes de PRA), 2) les méthodes de raisonnement finitistes (par extension le finitisme) sont représentées (au sens formel⁶⁹) par PRA. Une fois acceptés les deux premiers volets de cette thèse : l'assimilation des fonctions admises dans R avec les FRP et l'assimilation des énoncés de R avec les énoncés Δ_0 et Π_1 (pris au sens de « jugements hypothétiques »), la solution de Tait répond clairement aux deux exigences formulées. C'est évident pour la première et cela se montre sans difficulté pour la seconde. Par contre, en limitant le finitisme à ce qui peut être représenté au sein de PRA, il en résulte des contraintes fortes sur le finitisme lui-même, qui ont été soulignées par William Tait et surtout par John Burgess dans [Burgess 2010]. Etant enfermé dans PRA, le finitiste ne peut en effet reconnaître la validité de PRA comme un tout, ni le fait que les théorèmes de PRA soient finitairement acceptables ; il ne peut en particulier pas reconnaître l'équivalence entre la dérivabilité dans PRA et la prouvabilité finitiste, soit la thèse de Tait elle-même [Tait 1981, p. 545-546] et [Burgess 2010, p. 133]. Bien entendu, le finitiste ne peut pas non plus reconnaître la non-contradiction de PRA (ce qui est la condition de possibilité du programme). Ce troisième volet de la thèse de Tait correspond en fait à deux affirmations enchaînées : 1) les preuves finitistes se confondent avec les dérivations dans un système formel S, formalisant l'arithmétique « réelle » R, 2) S est PRA. A l'instar de ce qu'il en est pour l'assimilation des fonctions « réelles » aux seules FRP, cette proposition bénéficie d'un large consensus, sans pour autant que celui-ci soit total. Là encore, les objections qui ont été formulées concernent d'abord la prise en compte de PRA pour rendre compte de S (seconde affirmation), en suggérant pour S, soit un système formel plus faible que PRA, soit au contraire plus fort que celui-ci. Sans surprise, on retrouve à leur origine, principalement les acteurs déjà cités qui se sont prononcés en faveur, d'une part d'une réduction des FRP, d'autre part d'une extension de celles-ci, comme fonctions finitistes. Elles consistent plus rarement en la contestation de la possibilité de rendre compte du caractère finitiste des preuves dans un système formel donné (première affirmation). Notre propre objection est plus globale, en contestant la nécessité de définir une preuve finitiste

⁶⁹ Nous rappelons qu'une propriété P ou une classe de nombres C_p est représentable (on dit aussi capturable) dans un système formel T si et seulement si il existe dans ce système formel une formule F(n) telle que :

1) si n vérifie P (appartient à la classe C_p) alors $\vdash_T F(n)$;

2) si n ne vérifie pas P (n'appartient pas à la classe C_p) alors $\vdash_T \neg F(n)$.

Cette définition s'étend aux fonctions numériques à une variable f(m), en faisant intervenir une formule à deux variables F(m,n) : 1) si f(m) = n, alors $\vdash_T F(m,n)$, 2) si f(m) \neq n, alors $\vdash_T \neg F(m,n)$.

en général, dans le but de donner un contenu technique au programme de Hilbert au sens restreint. Nous avons vu en effet (§ 1.1) que le strict besoin consiste seulement à définir une procédure finitiste *de consistance* (d'un système formel formalisant l'arithmétique du deuxième ordre), en utilisant le critère de vérifiabilité, ce que nous proposons de faire sur la base d'une interprétation *libéralisée* (du troisième volet) de la thèse de Tait.

Les commentateurs qui se sont prononcés en faveur d'un système formel plus faible que PRA pour rendre compte de l'arithmétique « réelle » obtenue de manière finitiste, acceptent évidemment le principe d'une réduction des preuves finitistes aux dérivations dans un système formel donné, simplement ils proposent pour cela un système formel plus faible que PRA. Compte tenu de la confusion qui règne sur la désignation des systèmes formels de l'arithmétique du 1^{er} ordre⁷⁰ et pour faciliter les discussions ultérieures, nous avons rassemblé les caractéristiques les plus importantes des principaux d'entre eux dans le tableau ci-après⁷¹.

⁷⁰ Par exemple le système EA de [Avigad 2003] ne recouvre certainement pas EA de [Potter 2000].

⁷¹ Pour la désignation de ces systèmes formels, nous avons retenu celle qui a été fixée par l'ouvrage de référence classique : [Hajek, Pudlak 1993].

Tableau I : Bilan des caractéristiques des principaux systèmes formels de l'arithmétique du 1^{er} ordre⁷²

Système formel	ERA	Q	$I\Delta_0, I\Sigma_0$	EA, EFA, $I\Sigma_0 + \text{Exp}$	$I\Sigma_1, \text{QF-IA}$	PRA	PA, Z_1, Z
Origine	Thistlewaite, ... 1988 ⁷³	Robinson 1952 ⁷⁴	Parikh 1971	Friedman 1980	Paris, ... 1978 ⁷⁵	Skolem 1923	Tarski, ... 1953 ⁷⁶
Langage	PRA	PA	PA	PA + Exponentiation	PA	PRA	PA
Quantificateurs	Bornés	Non bornés	Non bornés	Non bornés	Non bornés	Bornés	Non bornés
Portée du principe d'induction	SANS	SANS	Enoncés Δ_0	Enoncés Δ_0	Enoncés Σ_1	Toute formule (variables libres)	Toute formule
Limite <i>supérieure</i> de l'ordinal accessible	ω	ω	?	ω^3	ω^ω	ω^ω	ε_0 ⁷⁷
Démonstration du premier théorème de Gödel de 1931 ?	NON	OUI ⁷⁸	OUI	OUI	OUI	OUI ⁷⁹	OUI
Complétude syntaxique ?	OUI	NON ⁸⁰	NON	NON	NON	NON ⁸¹	NON
Démonstration du second théorème de Gödel de 1931 ?	NON	NON	NON	OUI ⁸²	OUI	OUI	OUI
Démonstration finitiste de consistance ?	OUI	OUI ⁸³	OUI	OUI ⁸⁴	NON	NON	NON

⁷² La plupart des informations contenues dans ce tableau seront en fait utiles pour les discussions ultérieures.

⁷³ Thistlewaite Paul, McRobbie Michael, Meyer Robert, *Automated theorem-proving in non-classical logics*, London, J. Wiley and sons, 1988.

⁷⁴ Robinson Raphael, « An essentially undecidable axiom system », *Proceedings of the International Congress of Mathematicians*, Cambridge (USA), Vol. 1, 1952, p. 729 – 730.

⁷⁵ Paris Jeffrey, Kirbie Laurie, « Sigma_n collection schemas in arithmetic », dans McIntyre Angus, Pacholski Leszeh, Paris Jeffrey (ed.), *Logic Colloquium '77*, Amsterdam, North Holland, p. 199-209.

⁷⁶ Tarski Alfred, Mostowski Andrzej, Robinson Raphael, *Studies in logic on the foundations of mathematics*, Amsterdam, North Holland, 1953.

⁷⁷ Nous rappelons que $\varepsilon_0 = \omega$ à la puissance ω , à la puissance ω , ...

⁷⁸ Q est la théorie minimale à laquelle s'applique le premier théorème de Gödel de 1931.

⁷⁹ Bien que les indécidables mis en avant par le premier théorème de Gödel de 1931 soient des énoncés Π_1 , on peut tout de même mettre en évidence un énoncé indécidable de PRA à travers l' ω -incomplétude. Mais il faut pour cela un théorème plus fort que le premier théorème de Gödel de 1931 ; nous traiterons ce point au § 4.1.

⁸⁰ L'incomplétude syntaxique de Q s'établit trivialement (sans passer par le premier théorème de Gödel de 1931).

⁸¹ Par contre, il existe une procédure de décision pour chaque instance donnée d'un énoncé général (qui s'apparente à un énoncé Π_1).

⁸² EA est la théorie minimale à laquelle s'applique le second théorème de Gödel de 1931.

⁸³ Au-delà du fait que le second théorème de Gödel de 1931 ne s'y applique pas, l'on peut tout de même en établir un équivalent et montrer que la consistance de Q ne peut pas être dérivée dans Q : [Bezboruah, Shepherdson 1976]. Mais la consistance de Q peut s'établir finitement (dans PRA), par des moyens qui vont au-delà de ceux de Q.

⁸⁴ La consistance de EA s'établit finitement (dans PRA), par des moyens qui vont au-delà de ceux de EA.

Sur cette base, au-delà des disparités sur les désignations utilisées par les auteurs, il s'avère que les systèmes formels qui ont été proposés en alternative faible à PRA se ramènent en fait à deux⁸⁵ : d'une part $I\Delta_0$ à la fois par Karl-Georg Niebergall et Matthias Schirn dans [Niebergall, Schirn 1998], par Charles Parsons dans [Parsons 1998] et semble-t-il⁸⁶ par Michael Potter dans [Potter 2000], d'autre part EA par Jeremy Avigad dans [Avigad 2003]. Le choix de EA est techniquement intéressant, comme le signale Richard Zach [Zach 2005a, p. 511], dans la mesure où ce système formel présente des caractéristiques métamathématiques analogues à celles de PRA tout en étant un peu plus faible. Par contre, il n'y a aucune justification d'un tel choix sur la seule base des critères finitistes, sachant que ce n'est pas pour rendre compte du finitisme de Hilbert qu'il a été introduit par Jeremy Avigad. En sens inverse, $I\Delta_0$ est techniquement beaucoup plus faible que PRA au point que la dérivabilité n'y est pas représentable, ce qui l'immunise⁸⁷ contre l'application du second théorème de Gödel de 1931. Karl-Georg Niebergall et Matthias Schirn sont conduits au choix de ce système formel par leur refus de prendre en compte les énoncés Π_1 parmi les énoncés de R et Charles Parsons est amené à $I\Delta_0$ par son refus de prendre en compte l'exponentiation parmi les fonctions de R. Dans les deux cas, nous avons contesté le caractère abusif (du point de vue finitiste) de ces restrictions. Au-delà de la proposition de Georg Kreisel de 1958, que nous avons déjà évoquée et sur laquelle nous ne reviendrons pas⁸⁸, les suggestions de rendre compte des preuves finitistes par un système formel au contraire *plus fort* que PRA, ont essentiellement quatre sources. La première est le constat, mentionné par exemple par Richard Zach [Zach 2005a, p. 510] et [Zach 2007, p. 425-426], que les démonstrations de consistances obtenues par l'école de Hilbert dans les années 1920 utilisaient des méthodes de preuve qui dépassaient déjà PRA. Nous l'analyserons en détail au § 4.1 en soulignant les limites. La deuxième qui lui est liée, est basée sur l'interprétation des propos de [Hilbert, Bernays 1934] dont nous avons déjà relevé l'ambiguïté (§ 1.2). La troisième est liée à l'utilisation de la *règle ω* introduite par Hilbert en 1930, que nous analyserons en détail au § 3.2 dans le cadre des extensions du finitisme originel, en en

⁸⁵ Panu Raatikainen suggère également la possibilité d'utiliser un système formel plus faible que PRA pour rendre compte de l'arithmétique finitiste de Hilbert, mais sans en définir le contenu [Raatikainen 2003, p. 160-161].

⁸⁶ Compte tenu de la description que l'auteur en donne [Potter 2000, p. 232, 238-239] et sachant qu'il renvoie explicitement à [Parsons 1998], on peut penser qu'il se réfère à $I\Delta_0$ (même s'il le nomme EA). D'un autre côté, il précise que ce système formel est à la fois complet [saturé] et décidable [Potter 2000, p. 232], ce qui n'est évidemment pas compatible avec $I\Delta_0$ mais ferait plutôt penser à ERA.

⁸⁷ Nous serons amenés à aborder dans le détail ce point technique au § 3.3.

⁸⁸ Cette position qui conduit à étendre S jusqu'à PA est peu ou prou reprise par Solomon Feferman [Feferman 1988b, p. 374] ; nous serons amenés à y revenir au § 4.2, en analysant les réalisations partielles du programme de Hilbert.

soulignant également les limites. La quatrième se fonde sur les échanges épistolaires de 1931, que nous avons déjà analysés (§ 1.2). Cette dernière est clairement la plus radicale, puisqu'elle revient à refuser l'idée même que les preuves finitistes puissent être traduites par un système formel unique. Nous en avons toutefois mentionné les limites, nous n'y reviendrons donc pas. Sur la base de l'expression générale du Programme à laquelle nous sommes parvenus au § 1.1, à savoir : *montrer de manière finitiste que tout théorème « réel » de T est vérifiable*, son expression technique dans le cadre de l'interprétation standard de la thèse de Tait devient : montrer dans PRA que tous les théorèmes de la partie « réelle » de T sont également des théorèmes de PRA, donc (via le deuxième volet) que T est conservatif par rapport à PRA pour les énoncés Π_1 ⁸⁹.

Comme nous l'avons déjà indiqué, le gros reproche que l'on peut faire à cette interprétation standard du troisième volet de la thèse de Tait, est que le fait de définir ce qu'est une preuve finitiste *en général*, va au-delà du seul but de donner un contenu technique au programme de Hilbert au sens restreint et contraint inutilement ce dernier. En particulier, nous contestons à John Burgess que le Programme (ou la nécessité d'en définir le contenu technique) impose que le contenu de la base intuitive du finitisme ait lui-même besoin d'être justifié par une *preuve finitiste* [Burgess 2010, p. 137]. Le strict besoin du Programme, consiste en effet seulement à définir une procédure finitiste permettant *d'établir la consistance* de la partie « réelle » du système formel T retenu formalisant l'arithmétique du deuxième ordre, en utilisant le critère de vérifiabilité, la base intuitive du finitisme étant reconnue a priori par évidence. Il s'agit donc de démontrer la consistance de l'ensemble des théorèmes « réels » de T (possiblement obtenus par le détour d'axiomes et de règles « idéaux »), c'est-à-dire, sur la base des deux premiers volets de la thèse de Tait que nous reprenons à notre compte, de tous les énoncés de T qui sont de la forme Δ_0 et Π_1 (et sont donc parfaitement circonscrits), et qui sont des théorèmes. Nous proposons pour cela une voie qui réponde *directement* à l'exigence globale du Programme, sur la même base des deux premiers volets de la thèse de Tait, mais en utilisant une version *libéralisée* de son troisième volet, qui s'exprime sous la forme de deux affirmations :

1) la consistance de PRA (qui est finitiste par construction) est supposée assurée par évidence intuitive⁹⁰ ;

⁸⁹ Nous rappelons qu'un système formel B est dit conservatif par rapport à un système formel A pour un certain type d'énoncés φ , si pour tout énoncé de ce type on a : si $\vdash_B \varphi$, alors $\vdash_A \varphi$.

⁹⁰ Une telle position, que nous justifierons au § 3.1, rend naturellement compte de l'affirmation de Bernays selon laquelle PRA est trop restrictif pour rendre compte de la théorie intuitive des nombres [Hilbert, Bernays 1934 (2001), p. 406-407].

2) il est de plus requis que tous les théorèmes de la partie « réelle » de T soient également des théorèmes de PRA, donc que T soit conservatif par rapport à PRA pour les énoncés Π_1 (ceux-ci étant compris comme des « jugements hypothétiques »), *la vérification devant elle-même être effectuée au sein de PRA* ; cette solution est à la fois parfaitement définie et répond explicitement à l'exigence finitiste du Programme via l'affirmation précédente⁹¹.

Nous aboutissons bien ainsi à une procédure *finitiste* parfaitement définie de *démonstration de consistance* de la partie réelle de T, même si elle tout à fait différente de celle qu'imaginait Hilbert (que nous avons présentée au § 1.1). Si, sur un plan purement technique, la seconde affirmation recouvre strictement l'expression technique du Programme obtenue dans le cadre de l'interprétation standard de la thèse de Tait, sa différence majeure (au-delà de rappeler la spécificité de la compréhension des énoncés Π_1) est qu'elle ne limite pas le finitisme à PRA, celui-ci n'étant qu'un élément de celui-là. Contre cette interprétation standard, elle n'interdit nullement au finitiste de reconnaître ce fait, et plus généralement la thèse de Tait ainsi libéralisée elle-même, bien entendu sur une base intuitive qui va au-delà de PRA. Également contre l'interprétation standard, le finitiste peut savoir et *sait* que PRA est consistant⁹² (condition de possibilité du Programme).

1.5 Expression technique du programme de Hilbert au sens restreint dans le cadre de l'interprétation retenue

Nous présentons d'abord les relations techniques un peu subtiles existant entre les différentes notions métamathématiques de consistance, de correction, et de complétude, *dans un cadre classique* (non finitiste)⁹³. Eu égard à notre centre d'intérêt, nous nous limitons à l'arithmétique du premier ordre (en logique classique) et même aux seuls énoncés Δ_0 , Π_1 , et Σ_1 . Plus particulièrement, nous nous intéressons à la correction et à la complétude par rapport au modèle standard de l'arithmétique. Pour cela, l'on utilise classiquement les notions

⁹¹ Si tous les commentateurs du programme de Hilbert font bien état d'un tel critère de conservativité, aucun d'entre eux ne le présente comme *l'expression d'une exigence finitiste* qui en minimise les contraintes, appliquée à la requête exprimée par la formulation standard du Programme.

⁹² Sachant que ce n'est pas parce que le finitisme selon notre interprétation n'est pas limité à PRA qu'il n'est pas pour autant parfaitement défini et circonscrit, en comprenant : PRA, la connaissance de la consistance de PRA et possiblement, la connaissance de son propre contenu.

⁹³ Leur connaissance est d'abord nécessaire pour la discussion qui va suivre sur les différentes interprétations qui ont été proposées en prolongement de la thèse de Tait. Elle est également indispensable, dans la mesure où elle intervient dans tous les résultats métamathématiques qui ont été obtenues postérieurement aux théorèmes de Gödel de 1931, et que nous allons aborder dans la suite de notre analyse.

respectives de Δ_0 -correction, Π_1 -correction, Σ_1 -correction et de Δ_0 -complétude, Π_1 -complétude, Σ_1 -complétude, étant sous-entendu (dans ce paragraphe) que celles-ci se rapportent à l'un des modèles (standard ou non) de n'importe quel système formel raisonnable de l'arithmétique. La définition de chacune de ces notions est assez intuitive⁹⁴ ; elle est illustrée sur le schéma B ci-après en se limitant alors aux seuls énoncés Δ_0 et Π_1 ⁹⁵ et au modèle standard (noté \mathcal{M}_0). Nous avons par ailleurs représenté sur le schéma C ci-après, les liens logiques⁹⁶ existant entre toutes ces notions et les autres notions métamathématiques plus générales que nous avons déjà introduites⁹⁷. La situation peut être résumée en distinguant cinq niveaux, dont trois principaux⁹⁸, d'exigence ou de contrainte croissante pour le système formel. Le niveau le plus faible (1 sur le schéma C) est associé à la consistance simple (la non-contradiction), lui sont rattachées (de manière coextensives) à la fois, la Δ_0 -correction, la Δ_0 -complétude (donc la complétude syntaxique pour les énoncés Δ_0), la Π_1 -correction, et la Σ_1 -complétude. Un niveau d'exigence légèrement plus élevé (2 sur le schéma C) est nécessaire pour obtenir la Σ_1 -correction : il correspond à la 1-consistance⁹⁹. Un niveau d'exigence beaucoup plus élevé (4 sur le schéma C) est nécessaire pour obtenir la Π_1 -complétude et partant la complétude syntaxique

⁹⁴ Nous la rappelons tout de même. La correction relative à la famille d'énoncés d'un type donné (pour un système formel donné de l'arithmétique et par rapport à un modèle, consiste dans le fait que tout énoncé de ce type, dérivable dans ce système formel, est vrai dans ce modèle). De manière duale, la complétude (sémantique) relative à la famille d'énoncés d'un type donné (pour un système formel donné de l'arithmétique et par rapport à un modèle), consiste dans le fait que tout énoncé de ce type, vrai dans ce modèle, est dérivable dans ce système formel.

⁹⁵ Sur ce schéma, les flèches en trait plein caractérisent les relations qui peuvent être établies en supposant la 1-consistance du système formel. Quant aux flèches en traits pointillés, elles caractérisent des relations qui nécessitent la condition supplémentaire forte de complétude syntaxique du système formel pour les énoncés Π_1 et Σ_1 .

⁹⁶ Ces liens ne sont jamais mentionnés intégralement dans les ouvrages classiques (on en trouve une part notable dans [Smorynski 1977], [Isaacson 1992] et [Smith 2007]), bien qu'ils soient cruciaux pour la discussion qui va suivre. Il est vrai qu'ils s'établissent facilement à partir des définitions mêmes des notions introduites, en prenant en compte les symétries implicites de ces définitions ainsi que la dissymétrie fondamentale existant entre les énoncés Σ_1 et Π_1 (en supposant seulement la consistance simple) :

1) tous les énoncés Σ_1 intuitivement vrais sont dérivables ;

2) tous les énoncés Π_1 intuitivement faux sont réfutables.

⁹⁷ Ce schéma doit se lire de la manière suivante : les notions situées au sein d'un même niveau sont coextensives ; si un niveau est inclus dans un autre, les notions du niveau le plus large s'y appliquent à fortiori.

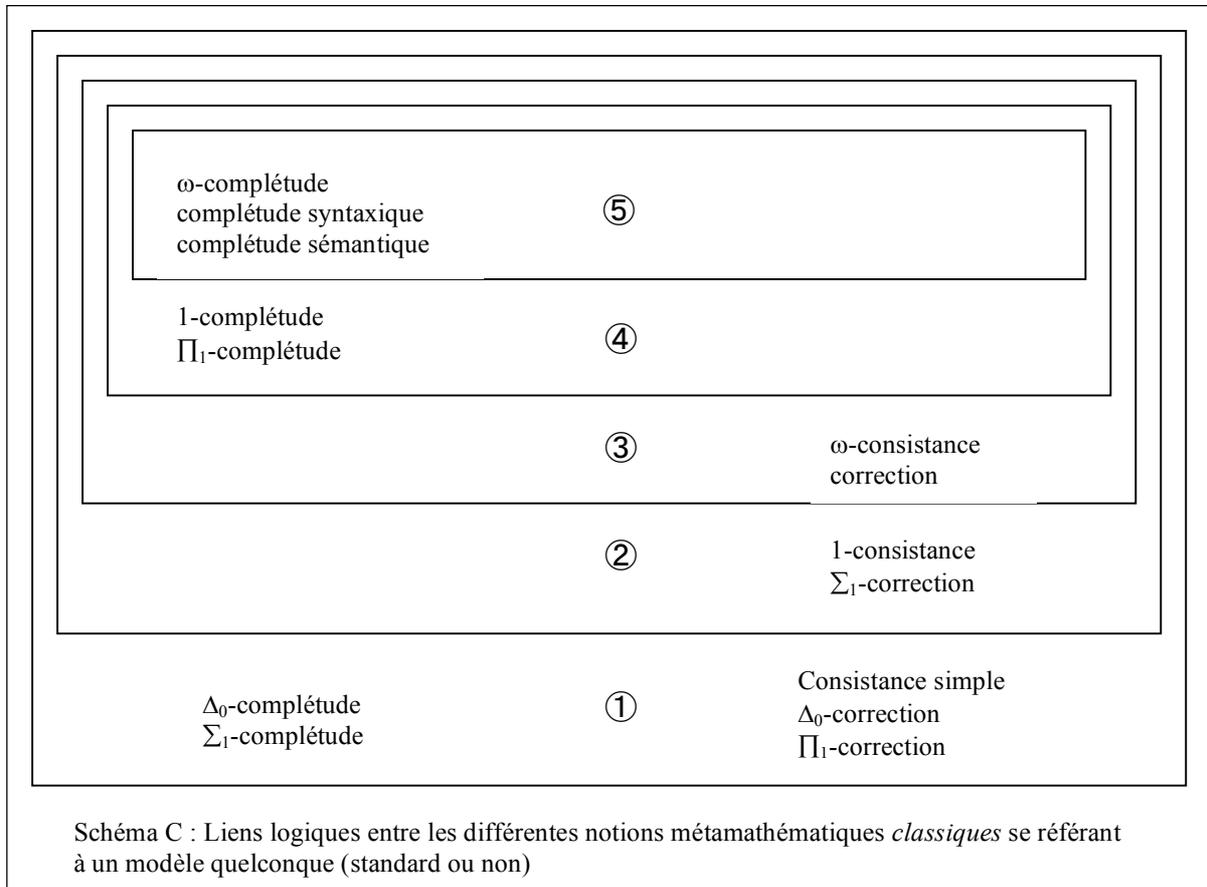
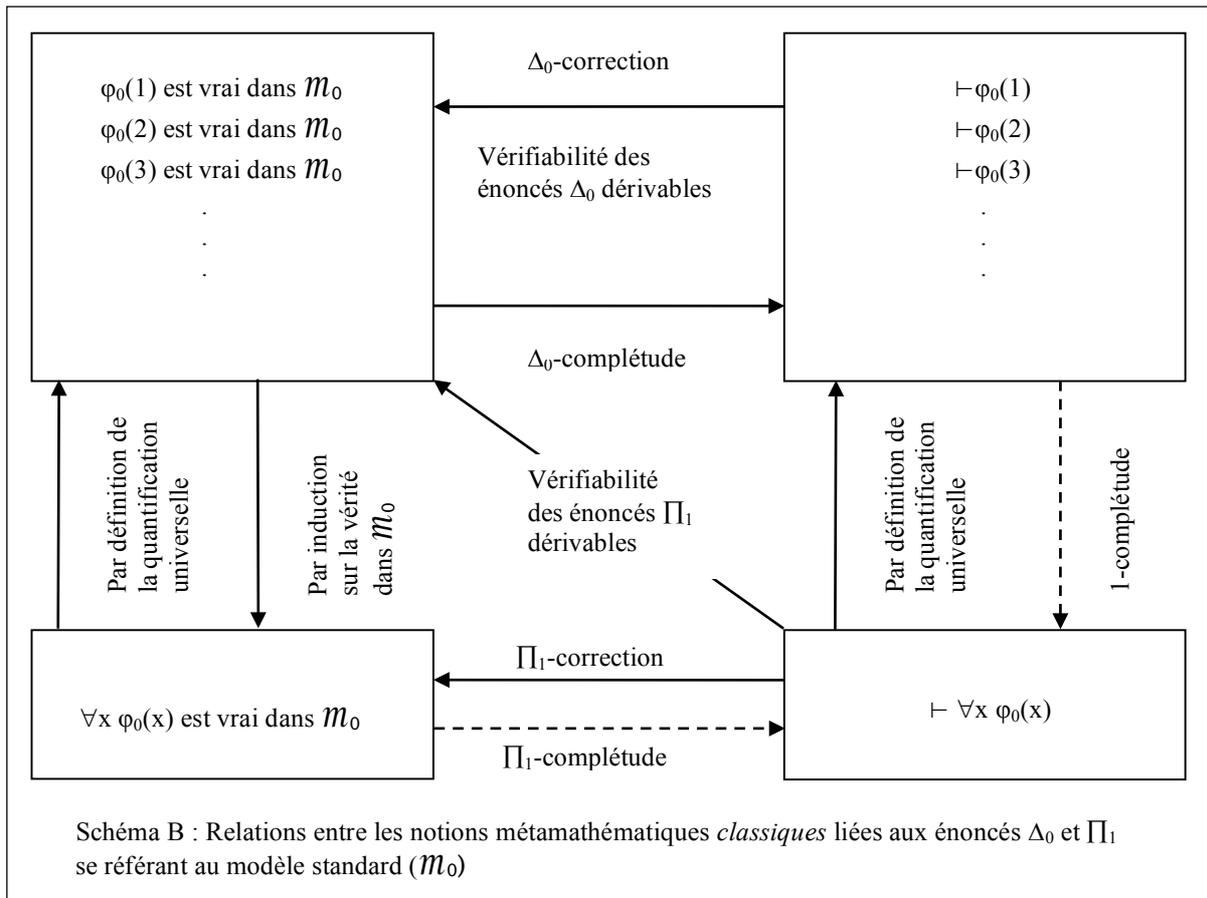
⁹⁸ Qui prendront toute leur importance lorsque nous discuterons des conditions d'application des deux théorèmes de Gödel de 1931.

⁹⁹ Nous rappelons que la 1-consistance, notion purement syntaxique, est l' ω -consistance où les instances de $\varphi(x)$ sont limitées à des énoncés Δ_0 . Le résultat obtenu montre qu'elle correspond sur le plan sémantique à l'union de la Π_1 -correction et de la Σ_1 -correction (l' ω -consistance elle-même étant une exigence plus forte, qui correspond sur le plan sémantique à la correction par rapport à un modèle de tout type d'énoncé du système formel : y compris au-delà de Δ_0 , Π_1 et Σ_1).

pour les énoncés Π_1 et Σ_1 : il correspond à la 1-complétude¹⁰⁰. Pour achever la circonscription de ce paysage métamathématique, il faut encore mentionner la « consistance externe », qui a été introduite à la fois par Bernays et par Gödel, mais *dans deux sens différents*. La consistance externe *au sens de Bernays* a été explicitement introduite par celui-ci pour éliminer la situation « pathologique » que nous avons déjà évoquée de systèmes formels 1-inconsistants tout en étant consistants, pour lesquelles certains théorèmes Σ_1 sont intuitivement *faux*, mais vrais dans un modèle non standard [Hilbert, Bernays 1939 (2001), p. 336-337]. Cela rend clairement incompatible la prise en compte des énoncés Σ_1 avec le finitisme de Hilbert¹⁰¹, comme le fait remarquer Dag Prawitz [Prawitz 1981, p. 258]. La consistance externe au sens de Bernays se ramène à la Σ_1 -correction et est donc équivalente à la 1-consistance, même si l’auteur distingue explicitement les deux notions, possiblement avec une arrière-pensée de prise en compte du contexte finitiste [Hilbert, Bernays 1939 (2001), p. 344, n. 3]. La consistance externe *au sens de Gödel* a été introduite par celui-ci dans le but de rendre compte de la requête du programme de Hilbert ; elle correspond à la vérifiabilité (au sens de Bernays) des énoncés Π_1 dérivables [Gödel 1972b (1990), p. 305]. Au moins dans un cadre classique, elle est donc coextensive avec la consistance simple, résultat mentionné par Gödel et souligné par Solomon Feferman [Feferman 1990, p. 286] ; elle est également coextensive à la Π_1 -correction, comme le signale Richard Zach [Zach 2005a, p. 507-508].

¹⁰⁰ Nous rappelons également que de manière analogue à la 1-consistance, la 1-complétude est l' ω -complétude où les instances de $\varphi(x)$ sont limitées à des énoncés Δ_0 . Le résultat obtenu montre qu'elle correspond sur le plan sémantique, à l'union de la Σ_1 -complétude et de la Π_1 -complétude.

¹⁰¹ Le problème ne disparaît cependant pas en limitant les énoncés « réels » aux seuls énoncés Π_1 . En effet, ainsi que l'indique Kleene, il suffit de considérer la négation d'un théorème Σ_1 faux, qui est à la fois Π_1 (donc « réel »), vrai, et *réfutable* [Kleene 1986a, p. 135]. Or si l'exigence que les énoncés « réels » réfutables soient faux n'est pas une expression du Programme *stricto sensu* (nous allons y revenir), il n'en reste pas moins que de tels énoncés « pathologiques » ne s'accordent pas avec son esprit, ainsi que l'a souligné Jacques Dubucs [Blanché, Dubucs 1996, p. 379].



La question se pose maintenant de savoir ce que deviennent les différentes notions métamathématiques définies dans un cadre classique, lorsque l'on se place dans le cadre finitiste au sens de Hilbert. Pour cela, le cadre finitiste que nous retenons est celui de la thèse de Tait, aussi bien dans sa version standard que selon notre propre interprétation. Comme nous l'avons vu, il en résulte d'emblée trois caractéristiques spécifiques : 1) la vérité prise en compte est la vérité intuitive¹⁰², 2) les énoncés Π_1 qui interviennent dans la Π_1 -correction et la Π_1 -complétude doivent être compris en tant que « jugements hypothétiques », 3) la Σ_1 -correction ainsi que la Σ_1 -complétude ne peuvent même pas y être définies (les énoncés Σ_1 étant « idéaux » n'ont pas de valeur de vérité). Pour ce qui concerne les énoncés Δ_0 , nous avons vu au § 1.4 que ceux-ci sont directement finitistes. Par ailleurs, pour tout système formel de l'arithmétique effectif et consistant, ils sont décidables¹⁰³. Enfin, étant « réels », on sait reconnaître pour chacun d'eux, dans un cadre finitiste, s'il est (intuitivement) vrai ou s'il est (intuitivement) faux. Il en résulte que l'on retrouve directement les relations de coextensivité entre consistance simple, Δ_0 -correction (qui se confond alors avec la vérifiabilité des énoncés Δ_0 dérivables) et Δ_0 -complétude, qui existent dans le cadre classique. Dans le cas des énoncés Π_1 , dans la mesure où chacune de leur instance est un énoncé Δ_0 et où ils sont compris en tant que « jugements hypothétiques », la définition et les relations qui lient les notions de Π_1 -complétude, Π_1 -correction, et vérifiabilité des énoncés Π_1 dérivables, dans un cadre classique, sont également conservées dans un cadre finitiste. Il est par contre clair que ces notions dans un cadre finitiste (c'est-à-dire où les énoncés Π_1 sont compris comme des « jugements hypothétiques »), qui constitue le cadre naturel pour la dernière citée (la consistance externe au sens de Gödel), sont *plus faibles* que leurs pendants classiques (où les énoncés Π_1 sont compris classiquement) et en particulier *plus faibles* que la consistance (simple). Par ailleurs les notions de 1-consistance et de 1-correction n'ont pas de sens dans un cadre finitiste¹⁰⁴, dans la mesure

¹⁰² Les énoncés vrais au sens intuitif selon Hilbert, sont clairement beaucoup plus limités que ceux qui sont vrais dans le modèle standard de n'importe quel système formel raisonnable de l'arithmétique, en incluant par exemple les énoncés Σ_1 vrais, même si ces derniers sont en général également rattachés à l'intuition (évidemment en un sens différent de celui de Hilbert).

¹⁰³ C'est ainsi que dans le système ERA qui ne contient que des énoncés Δ_0 (voir le tableau I du § 1.4), on est en mesure d'établir directement et de manière finitiste, à la fois sa complétude syntaxique et sa consistance (d'une manière analogue à celle préconisée par Hilbert).

¹⁰⁴ Ce qui est cohérent avec le fait que sur un plan sémantique, la 1-consistance est coextensive avec l'union de la Π_1 -correction et de la Σ_1 -correction, cette dernière ne pouvant être définie dans un cadre finitiste (et d'une manière analogue pour la 1-complétude par rapport à la Σ_1 -complétude). Cela ne veut bien sûr pas dire pour autant que le problème posé par la situation « pathologique », qui a conduit Bernays à introduire sa notion de consistance externe, disparaît dans un cadre finitiste (se reporter à la note 101). Nous allons aborder cette situation particulière dans la discussion qui suit.

où leur définition fait intervenir de manière essentielle la quantification universelle comprise classiquement (sachant que les définitions analogues que l'on pourrait introduire dans un cadre finitiste, seraient alors trivialement toujours vérifiées).

Nous abordons enfin la discussion de la pertinence des différentes formulations du programme de Hilbert au sens restreint qui ont été proposées en prolongement de la thèse de Tait (donc en acceptant au moins partiellement celle-ci), au-delà de celle que nous avons retenue. Pour cela, nous choisissons d'aborder le problème de la manière la plus générale (et formelle), en commençant par mettre provisoirement de côté les résultats issus du paragraphe précédent, avec l'objectif de circonscrire l'ensemble des situations possibles. Nous introduisons à cet effet des conventions de notation plus fines que celles que nous avons déjà introduites¹⁰⁵. Sur le plan sémantique, R_V désigne l'ensemble des énoncés réels vrais et R_F désigne l'ensemble des énoncés réels faux¹⁰⁶ (au sens classique ou au sens de Hilbert). Sur le plan syntaxique, pour chacun des deux systèmes formels S et T , l'ensemble des énoncés constitutifs se décompose en ceux dérivables, indécidables et réfutables et bien entendu pour T , il faut encore distinguer les énoncés « réels » et les énoncés « idéaux » sachant que par construction, l'ensemble des énoncés de S se confond avec R . Au sein de S et de T on est alors conduit à une liste exhaustive de neuf classes d'énoncés, trois pour S et six pour T , qui sont rassemblés dans le tableau suivant :

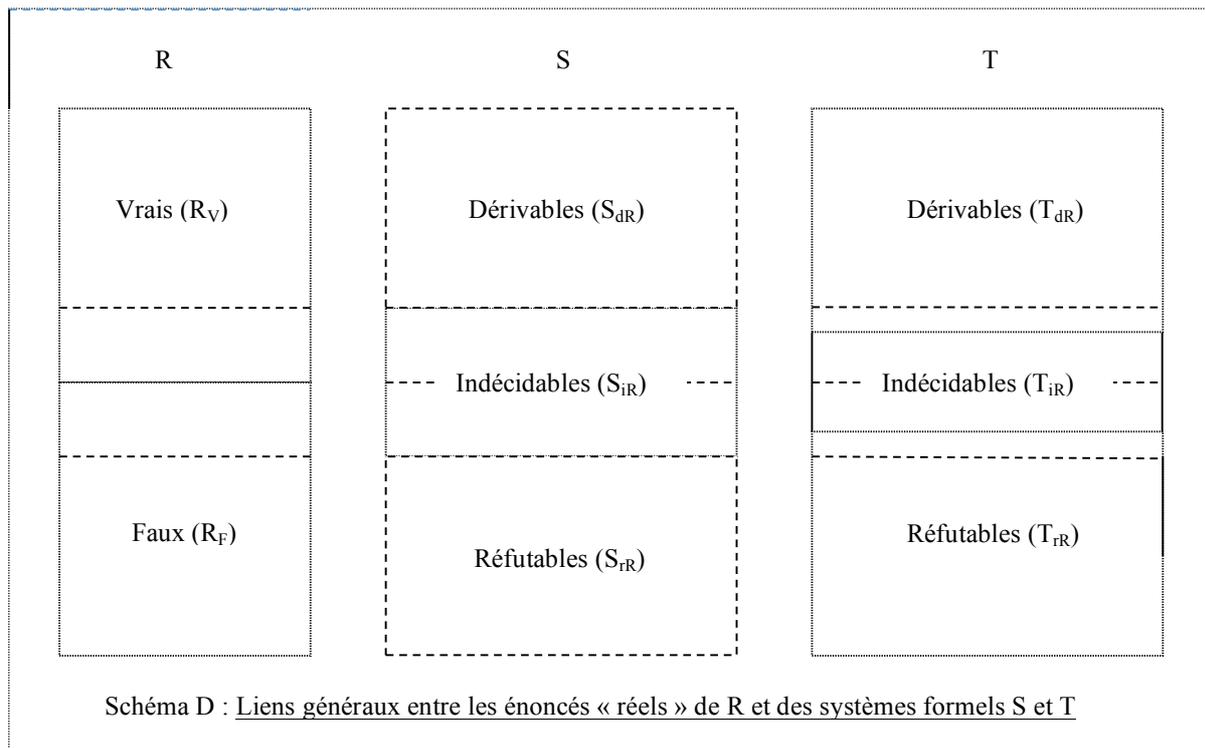
Tableau II : Convention de notation syntaxique concernant les énoncés des systèmes formels S et T

S	« Réels »		T	« Réels »	« Idéaux »
Dérivables	S_{dR}		Dérivables	T_{dR}	T_{dI}
Indécidables	S_{iR}		Indécidables	T_{iR}	T_{iI}
Réfutables	S_{rR}		Réfutables	T_{rR}	T_{rI}

Il en résulte d'emblée les relations : $R = R_V + R_F = S_{dR} + S_{iR} + S_{rR} = T_{dR} + T_{iR} + T_{rR}$ (comme cela apparaît immédiatement sur le schéma D suivant, limité aux énoncés « réels » et dont la représentation suppose implicitement que la 1-consistance est assurée) et : $I = T_{dI} + T_{iI} + T_{rI}$.

¹⁰⁵ Nous désignons par R l'ensemble des énoncés « réels », I l'ensemble des énoncés « idéaux », S le système formel (ou l'ensemble des énoncés de celui-ci) formalisant R avec l'exigence de finitisme et T le système formel (ou l'ensemble des énoncés de celui-ci) formalisant $I + R$ sans l'exigence de finitisme. Ce choix terminologique, essentiellement issu de Gödel en 1930 [Dawson 1984b, p. 125] est aujourd'hui très consensuel.

¹⁰⁶ Sachant que seuls les énoncés « réels » sont concernés par les conditions de vérité, puisque pour Hilbert, les énoncés « idéaux » ne sont ni vrais ni faux.



Ce qui est fondamentalement en jeu dans l'expression du Programme est bien sûr le lien entre R_V , S_{dR} et T_{dR} . La première condition (impérative) de possibilité du Programme que nous avons évoquée, qui correspond à la vérifiabilité de S_{dR} , se traduit par la condition : $S_{dR} \subseteq R_V$. La seconde condition (éventuelle) de possibilité du Programme que nous avons également évoquée, qui correspond à la décidabilité de chacun des énoncés de S, implique la relation symétrique : $R_V \subseteq S_{dR}$. Par construction, il est supposé que T étende S, ce qui se traduit par la relation : $S_{dR} \subseteq T_{dR}$. Enfin, le caractère possiblement conservatif de T par rapport à S pour les énoncés réels se traduit par la relation symétrique : $T_{dR} \subseteq S_{dR}$. Notre point de départ est constitué d'une part de la première condition de possibilité du Programme : $S_{dR} \subseteq R_V$ et d'autre part de l'hypothèse que T étende S par construction, soit : $S_{dR} \subseteq T_{dR}$. Sur cette seule base, il n'y a alors que quatre solutions possibles¹⁰⁷, que nous allons analyser les unes après les autres (par généralité décroissante) : 1) $S_{dR} \subset R_V \subset T_{dR}$, 2) $S_{dR} \subset T_{dR} \subset R_V$, 3) $S_{dR} = T_{dR} \subset R_V$, 4) $S_{dR} = T_{dR} = R_V$. Remarquons au préalable que le Programme sous sa forme générale se déduit directement de nos seules hypothèses initiales. Supposons en effet que T soit (simplement) inconsistant, alors par simple effet de propagation les théorèmes communs avec S vont l'être également, ce qui est incompatible avec l'hypothèse de consistance simple de S. Cette situation est donc interdite.

¹⁰⁷ Qui recourent celles discutées par Peter Smith [Smith 2007, p. 254-255].

La première solution formelle : $S_{dR} \subset R_V \subset T_{dR}$ traduit l'éventualité que certains théorèmes « réels » de T ne soient pas vrais. Comme T est supposé consistant (même si c'est ce que l'on cherche à établir) c'est donc (dans le cadre classique) la situation très particulière que nous avons déjà évoquée, où T est consistant tout en étant 1-inconsistant. Bien que cette situation soit permise par notre formalisme (dans ce paragraphe), on peut raisonnablement considérer que la condition de vérifiabilité qui caractérise S doit s'appliquer également à T , ce qui l'interdit alors. De fait, aucun commentateur n'a envisagé de traduire le Programme de cette manière. Nous n'en dirons donc rien de plus à ce niveau¹⁰⁸.

La deuxième solution formelle : $S_{dR} \subset T_{dR} \subset R_V$ est beaucoup plus intéressante. Elle correspond en effet à l'interprétation raisonnable la plus libérale du Programme, qui n'exige pas le caractère conservatif de T par rapport à S pour les énoncés « réels ». Une telle solution a été défendue avec vigueur par Michael Detlefsen : [Detlefsen 1990], et reprise à sa suite par Marcus Giaquinto : [Giaquinto 2002] et dans une certaine mesure par Panu Raatikainen : [Raatikainen 2003], contre une tradition bien établie jusqu'alors. La question est donc de savoir dans quelle mesure cette interprétation est compatible avec les exigences de Hilbert. Pour intégrer la thèse de Tait (selon une certaine interprétation), Michael Detlefsen est conduit à introduire une double partition parmi les énoncés R , entre ceux qui sont « finitairement vrais » (respectivement « finitairement faux ») qui correspondent strictement à ceux qui sont dérivables dans S (respectivement réfutables dans S), et les autres qui correspondent aux indécidables de S vrais dans R (respectivement faux dans R). En considérant que seuls les premiers sont pertinents pour exprimer les exigences finitistes de Hilbert, il pose la question de savoir si celles-ci imposent alors que les théorèmes « réels » de T soient « finitairement » vrais (donc dérivables dans S) ou bien qu'ils ne soient pas « finitairement » faux (donc réfutables dans S) ? L'existence au moins possible d'énoncés « réels » indécidables dans S mais dérivables dans T , fait que les deux situations ne sont pas identiques¹⁰⁹. Si l'on introduit alors le troisième volet de la thèse de Tait selon l'interprétation standard : les énoncés « réels » vrais sont les théorèmes de PRA, il en résulte dans le premier cas (mais non dans le second) que les théorèmes « réels » de T doivent être des théorèmes de S (qui se confond alors avec PRA), donc que T doit être conservatif par rapport à PRA pour les énoncés Π_1 . En suivant une tendance assez naturelle les commentateurs ont en général choisi la première possibilité (la vérité « finitaire »), tandis que

¹⁰⁸ Nous serons amenés à revenir sur cette situation lorsque nous ferons intervenir les théorèmes de Gödel de 1931.

¹⁰⁹ En anticipant sur la discussion à venir, le fait qu'il y ait effectivement des énoncés Π_1 indécidables dans PRA mais dérivables dans PA (a fortiori dans un système formel formalisant l'arithmétique du deuxième ordre), rend cette alternative cruciale.

[Detlefsen 1990], [Giaquinto 2002] et [Raatikainen 2003] font évidemment état de la seconde (la fausseté « finitaire »), ce qu'ils nomment « la sûreté finitaire » [Giaquinto (2002), p. 185]. Si l'on fait abstraction du fait qu'une telle double partition des énoncés de R est totalement étrangère à Hilbert, si l'on s'en tient aux textes, ceux-ci semblent donner raison à Michael Detlefsen comme cela apparaît sans ambiguïté sur les deux des trois citations que nous avons déjà mentionnées, qui sont explicites sur ce point :

[...] ; nous aurions l'assurance que ces méthodes arithmétiques usuelles ne peuvent jamais conduire à un résultat dont on pourrait démontrer la *fausseté* ; [...].

[Hilbert, Bernays 1934 (2001), p. 98 ; souligné par nous]

Cette vision [du finitisme de Hilbert] suppose [...] qu'aucun théorème *faux* ne devient prouvable à travers l'adjonction des axiomes transfinis [...].

[Gödel en 1930, dans Dawson 1984, p. 125 ; souligné par nous]

Celui-ci peut alors prétendre tout à la fois respecter les exigences de Hilbert et se rattacher (au moins partiellement) à la thèse de Tait, sans que le caractère conservatif de T par rapport à PRA pour les énoncés Π_1 soit imposé. La justification la plus générale et la plus précise en est donnée par Marcus Giaquinto dans [Giaquinto 2002] en trois points :

- 1) le programme est formulé au travers de la nécessité de montrer finitairement qu'aucun énoncé réel faux n'est dérivable dans T [Giaquinto 2002, p. 185] ;
- 2) il est fait appel à la relation classique que nous avons déjà mentionnée selon laquelle tous les énoncés Π_1 faux sont réfutables dans S (consistant par définition) ;
- 3) on en déduit alors directement la formulation proposée du programme :

[...] la conclusion désirée du programme de Hilbert peut être restituée ainsi : pour tout énoncé finitaire [réel] de l'arithmétique α , si α est dérivable dans T, α n'est pas réfutable dans S.

[Giaquinto 2002, p.180 ; le commentaire entre crochets est nôtre]

Cette voie autorise en particulier que certains énoncés réels dérivables dans T soient en fait indécidables dans S, en opposition à l'exigence de conservativité de T par rapport à S pour les énoncés réels. La pierre d'achoppement d'une telle justification concerne le point 2 et porte sur le fait que si la relation en question est bien établie classiquement, elle n'est pas justifiée finitairement¹¹⁰. En effet, tout ce que l'argument technique précis donné par Marcus Giaquinto

¹¹⁰ Ce qui ramène cette solution à une *extension* du programme de Hilbert au sens restreint, à l'instar de l'option instrumentaliste que Michael Detlefsen défend par ailleurs et que nous analyserons au § 3.4.

montre [Giaquinto 2002, p. 179], est que pour tout énoncé $\Pi_1 \alpha$ donné, on est effectivement en mesure de montrer de manière finitiste qu'il est réfutable dans S. L'auteur ne fournit cependant aucune procédure qui soit finitairement acceptable et qui permette d'envisager tous les α , sachant qu'il est évidemment finitairement interdit de les parcourir un à un. C'est précisément pour une raison analogue que nous avons été contraint de limiter la libéralisation que nous avons introduite sur le 3^{ème} volet de la thèse de Tait, en imposant justement la nécessité du caractère conservatif de T par rapport à S pour les énoncés réels.

La troisième solution : $S_{dR} = T_{dR} \subset R_V$, traduit justement le caractère conservatif de T par rapport à S pour les énoncés « réels », ce qui rejoint la position que nous défendons et qui a également été retenue par la plupart des commentateurs. Stephen Simpson en donne la formulation la plus précise :

[...] montrer que la consistance du système formel Z_2 peut être menée dans le système formel PRA. Si cela pouvait être fait, il s'ensuivrait que chaque énoncé Π_1 qui est prouvable dans Z_2 est également prouvable dans PRA. Nous voudrions décrire cette sorte de chose [le programme de Hilbert] en disant que Z_2 est conservatif par rapport à PRA pour les énoncés Π_1 .

[Simpson 1988, p. 352 ; le commentaire entre crochets est nôtre]

Mais l'origine de son introduction, en lien avec la prise en compte de l'intégralité des exigences finitistes de Hilbert, est rarement soulignée par les commentateurs à l'exception de Philip Kitcher [Kitcher 1976, p. 103-104] et dans une moindre mesure de Dag Prawitz [Prawitz 1981, p. 253-254]. Lorsque Michael Potter prétend s'appuyer sur l'unique formulation explicite de Hilbert sur le sujet [Potter 2000 p. 226] que nous avons déjà citée [Hilbert 1926 (1972), p. 235], pour justifier la prise en compte du caractère conservatif de T par rapport S (pour les énoncés Π_1), il commet manifestement un abus d'interprétation¹¹¹. Par ailleurs, l'essentiel de la critique de Michael Detlefsen contre cette solution porte en fait sur l'exigence de complétude syntaxique de S pour les énoncés Π_1 (c'est-à-dire notre quatrième solution, dont l'analyse est à venir) [Detlefsen 1990, p. 357-358]. A sa suite, Panu Raatikainen tombe dans le même travers [Raatikainen 2003, p. 162]. Or une telle critique est immédiatement réfutée sur la base de la simple remarque que l'interprétation contestée est certes une conséquence de la complétude syntaxique de S pour les énoncés Π_1 , lorsque celle-ci est présente, *mais qu'elle ne l'impose nullement*.

¹¹¹ Comme le montre l'interprétation de Hilbert que propose Richard Zach [Zach 2007, p. 430], même si celui-ci a tendance à tomber dans l'excès inverse en niant que le caractère conservatif de T par rapport S soit une exigence du programme de Hilbert au sens restreint, ce qui est tout aussi abusif.

La quatrième solution : $S_{dR} = T_{dR} = R_V$ est défendue par Craig Smorynski dans [Smorynski 1977] et [Smorynski 1988]¹¹². C'est la plus contraignante puisqu'elle impose ainsi que S et partant l'ensemble des énoncés « réels » de $T : T_{dR} + T_{iR} + T_{rR}$ soit 1-complet (dans un cadre classique). Nous avons vu que la 1-complétude est une condition particulièrement forte, qui se ramène à la complétude syntaxique des énoncés de type Π_1 et Σ_1 . Cette solution exprime ainsi l'exigence de décidabilité de chacun des énoncés « réels », ce qui recouvre ce que nous avons appelé la seconde (éventuelle) condition de possibilité du Programme. Une telle interprétation « plus royaliste que le Roi » semble respecter à défaut de la lettre (manquante), l'esprit du Programme. Cette solution s'avère cependant beaucoup trop contraignante, dans la mesure où le seul système formel qui pourrait alors être retenu pour $S : ERA$, est bien trop pauvre pour servir le dessein de Hilbert. Surtout, le fait pour Hilbert de défendre *par ailleurs* une forme d'optimisme rationaliste (le refus de l'*ignorabimus*)¹¹³ n'implique nullement, dans le cadre du programme au sens restreint, une exigence de complétude syntaxique pour les énoncés Π_1 du système formel destiné à rendre compte des énoncés « réels » de manière finitiste. Cette solution est donc inutilement contraignante.

Finalement, sur les quatre interprétations possibles du programme de Hilbert au sens restreint intégrant la thèse de Tait, on peut éliminer d'emblée la (première) solution non raisonnable d'un système formel traduisant l'ensemble des mathématiques « idéales » et « réelles », consistant mais 1-inconsistant dans un cadre classique. En sens inverse, la (quatrième) solution d'une exigence de complétude syntaxique pour les énoncés Π_1 du système formel traduisant de manière finitiste les seules mathématiques « réelles », est quant à elle inutilement contraignante. Enfin, la prise en compte de l'ensemble des contraintes finitistes du programme de Hilbert impose l'exigence du caractère conservatif du système formel traduisant l'ensemble des mathématiques « idéales » et « réelles » par rapport au système formel traduisant de manière finitiste les seules mathématiques « réelles » pour les énoncés « réels » (troisième solution), contre l'option plus libérale de Michael Detlefsen qui ne l'impose pas (deuxième solution). A l'issue de l'ensemble de cette analyse et en conformité avec notre objectif, nous sommes maintenant en mesure de formuler de manière techniquement précise le programme de Hilbert au sens restreint selon notre interprétation. Celui-ci se ramène à la requête suivante : montrer *dans PRA* le caractère conservatif de Z_2 ¹¹⁴ (a fortiori de PA) par rapport à PRA pour

¹¹² Au-delà de la subtilité qu'il introduit, où il en vient à évoquer *les* programmes de Hilbert en distinguant le programme (sans la conséquence du finitisme) du Programme (avec) [Smorynski 1988, p. 43, 52, 54].

¹¹³ Position que nous discuterons au § 5.2.

¹¹⁴ Nous rappelons que Z_2 désigne la théorie de l'arithmétique du *second* ordre.

les énoncés Π_1 ¹¹⁵ (ceux-ci devant être compris comme des « jugements hypothétiques »). Nous allons pouvoir ainsi analyser dans quelle mesure ce Programme a été invalidé par les deux théorèmes de Gödel de 1931.

¹¹⁵ Ce qui recouvre très exactement la formulation de Stephen Simpson que nous avons déjà citée [Simpson 1988, p. 352].

2 Le programme de Hilbert au sens restreint a-t-il été invalidé par les deux théorèmes de Gödel de 1931 ?

Se poser aujourd'hui la question de l'incidence des théorèmes de Gödel de 1931 sur le programme de Hilbert peut sembler curieux tant la cause paraît entendue, comme l'illustre le florilège suivant où une telle position est exprimée par des auteurs reconnus, dans des tonalités variées [Weyl 1946, p. 13], [Ladrière 1957, p. 402], [Kreisel 1958b, p. 472], [Lakatos 1967 (1986), p. 36], [Sieg 1988, p. 338], [Simpson 1988, p. 349], [Smith 2007, p. 258-259] et [Zach 2007, p. 411]. Cette attitude est par ailleurs concordante avec les données historiques connues concernant l'accueil des résultats de Gödel par l'école de Hilbert. D'un point de vue technique, ces résultats ont en effet été acceptés très rapidement à la fois par von Neumann, dès la communication de Gödel lors de la table ronde du 7 septembre 1930 à Königsberg [Dawson 1985 (1990), p. 77], puis par Bernays¹¹⁶ [Bernays 1935a, p. 68] et [Hilbert, Bernays 1939 (2001), p. 314] ainsi qu'Herbrand [Herbrand 1931a, p. 213-214] et [Sieg 2003a, p. 4], sachant que tous les trois¹¹⁷ ont eu un échange épistolaire avec Gödel. Du point de vue de la perception par ces acteurs de l'impact des théorèmes sur le Programme, si l'on met à part le cas de Hilbert lui-même (sur lequel nous reviendrons au § 3.2), les quelques données disponibles vont également dans ce sens [Feferman 1993b (1998), p. 190] et [Carl Hempel, cité dans Dawson 1997, p. 70]. Il est toutefois possible de relever quelques notes discordantes au sein de ce consensus¹¹⁸ [Kleene 1986a, p. 127] et [van Bendegem 1993, p. 343]. L'objection majeure qui a relancé le débat avec un certain retentissement dans la communauté des philosophes de la logique, est cependant venue de Michael Detlefsen : [Detlefsen 1979], [Detlefsen 1986], [Detlefsen 1990] et [Detlefsen 2001], suivi (partiellement) par quelques rares auteurs, principalement Marcus Giaquinto : [Giaquinto 2002] et dans une certaine mesure Panu Raatikainen : [Raatikainen 2003]. C'est donc essentiellement la position défendue par Michael Detlefsen, concernant aussi bien le premier théorème de 1931 que le second, auxquels se heurte le programme de Hilbert d'une manière différente, que nous allons discuter. Au préalable nous répondons aux critiques variées (aussi bien philosophiques que techniques) qui ont été

¹¹⁶ Nous rappelons que c'est dans [Hilbert, Bernays 1939] que l'on trouve la première démonstration complète du second théorème de Gödel de 1931, l'article original se limitant (pour le second théorème) à une esquisse de preuve.

¹¹⁷ Mais non Hilbert [Gödel 1966 (2003), p. 187].

¹¹⁸ Nous abordons au § 7.1 le cas particulier des critiques plus générales de Wittgenstein concernant les théorèmes de Gödel de 1931 ainsi que le programme de Hilbert, dans [Wittgenstein 1956 (1983)].

formulées à l'encontre du Programme, et qui visent à discréditer celui-ci *indépendamment* des objections, essentiellement techniques, apportées par les théorèmes de Gödel de 1931.

2.1 Le programme de Hilbert au sens restreint était-il condamné d'avance ?

La première critique philosophique contre l'idée même ou l'intérêt même du programme de Hilbert consiste dans l'affirmation que les mathématiques, pas plus que les autres sciences, n'ont besoin de fondation. Elle émane d'auteurs variés mais tous d'envergure [Curry 1963 cité par Lakatos 1967 (1986), p. 31], [Putnam 1967 (1983), p. 295], [Kreisel 1968, p. 358] et [Shapiro 1991, p. ix]. A cette critique, au-delà de la remarque classique et consensuelle de la spécificité des mathématiques et de la logique de ne pas dépendre du contenu du réel empirique¹¹⁹ contrairement aux autres disciplines scientifiques, la réponse est qu'il y a bien eu historiquement des contradictions non triviales dans au moins deux systèmes axiomatiques¹²⁰, ce qui illustre le risque encouru comme le rappelle fort justement Quine :

Mais une force supérieure [pour un système formel] s'accompagne du risque croissant d'incohérence cachée. C'est un réel danger. La première théorie des classes développée pleinement et rigoureusement, celle de Frege, était incohérente, comme le montra le paradoxe de Russell. On a montré de la même manière, par le biais de preuves toujours plus subtiles et laborieuses, que différentes théories des classes plus récentes étaient incohérentes ; tel en particulier fut le sort d'une version antérieure de ML lui-même. [Quine 1953a (2003), p. 147 ; le commentaire entre crochets est nôtre]

Au-delà d'une remarque très générale de Putnam sur le fait que la démonstration de consistance d'un système axiomatique n'est pas suffisante pour établir la fondation des mathématiques [Putnam 1967 (1983), p. 303], une deuxième critique philosophique qui a été émise par des auteurs d'origine variée fait valoir l'argument relativiste classique selon lequel il est impossible de fonder quoi que ce soit sur rien, sans point d'ancrage a priori [Gödel 1953, p. 28, n. 30], [Webb 1968, p. 177], [Resnik 1974, p. 116] et [Niebergall, Schirn 1998, p. 300-301, n. 47]. La réponse à l'argument relativiste est bien entendu que le Programme de Hilbert ne repose pas sur rien, mais sur une fondation finitiste précise sur le contenu de laquelle nous avons beaucoup

¹¹⁹ Si l'on met à part la position défendue par Lazlo Kalmar dans [Kalmar 1967].

¹²⁰ Cinq selon Reinhard Kahle, qui n'évoque pas la version initiale du système de Quine, mais ajoute la logique combinatoire de Curry, le λ -calcul de Church, la théorie des constructions de Kreisel et la première théorie des types de Martin-Löf [Kahle 2015, p. 18].

insisté dans le chapitre précédent. La troisième critique philosophique est que les mathématiques sont évidemment (ou intuitivement) cohérentes et qu'une preuve de leur consistance est donc inutile. Elle émane principalement de Georg Kreisel [Kreisel 1958a, p. 474, n. 9], [Kreisel 1968, p. 358-359] et [Kreisel 1971, p. 240], avec un fort impact dans la communauté des philosophes de la logique et des mathématiques. L'angle d'attaque de Georg Kreisel est en trois points : 1) une démonstration de consistance, même constructiviste ou finitiste, n'accroît pas notre confiance dans les mathématiques utilisées (mais présente un intérêt annexe, en relation avec la compréhension de la signification des preuves de consistance obtenues), 2) c'est seulement à la suite d'une pratique et d'une expérience mathématique que nous nous convainquons progressivement de la consistance de systèmes formels au moins aussi forts que PA, 3) l'importance des paradoxes, quant aux risques qu'ils font courir aux mathématiciens, a été manifestement exagérée pour des raisons philosophiques. Sur le dernier point, qui est surtout une affaire d'opinion, on peut lui opposer par exemple le point de vue opposé de Quine que nous avons déjà cité [Quine 1953a (2003), p. 147]. Il convient cependant de prendre en compte deux contre-arguments de Solomon Feferman qui s'opposent à cette réponse :

De plus, dans le passé, toutes les inconsistances résultèrent de défauts de formulation très évidents, et ces défauts ont été éliminés. Bien sûr, il peut encore y avoir des défauts cachés et dont nous n'avons pas conscience, mais aucun n'a émergé, même après la somme massive d'investigations détaillées de tels systèmes. [Feferman 1998, p. 14]

Nommément, dans le cas de PA, nous avons un modèle intuitif absolument clair, constitué par les nombres entiers naturels, [...]. Pour revenir, par exemple à NF [le système de Quine], celui-ci n'a aucun modèle intuitif pour étayer notre croyance directe en sa consistance, [...].

[Feferman 2000b, p. 72 ; le commentaire entre crochets est nôtre]

Concernant le deuxième point, tous les systèmes axiomatiques ne présentent pas le même risque d'incohérence, celui-ci étant d'autant plus grand que le système est plus puissant comme le soulignent Stewart Shapiro (dans le cas particulier du deuxième ordre) et Reinhard Kahle [Shapiro 1991, p. v], [Kahle 2015, p. 18], sachant qu'il n'est bien entendu pas possible de circonscrire a priori la limite supérieure des systèmes formels « intuitivement évidents ». L'affirmation que des systèmes axiomatiques puissants de théories mathématiques générales, à l'instar de ZF, sont « intuitivement » ou « évidemment » consistants n'est donc pas autre chose qu'un acte de foi, particulièrement clair dans l'humour implicite de la formulation de Georg

Kreisel [Kreisel 1968, p. 358-359]¹²¹ ; elle ne saurait donc convaincre un sceptique. Quant au premier point, dont l'aspect technique n'est sans doute pas contestable, il s'adresse en fait au « mathématicien au travail ». Cela nous amène ainsi à suggérer que la position de Georg Kreisel caractérise essentiellement une opposition, entre le point de vue « interne » (à la pratique mathématique) du « mathématicien au travail », et le point de vue « externe » (à la même pratique) du philosophe des mathématiques. Cette opposition est d'ailleurs soulignée par l'auteur lui-même à quelques endroits [Kreisel 1968, p. 324, 359] et [Kreisel 1971, p. 252]. Or, comme le remarque Francesco Berto [Berto 2009b, p. 172] et comme cela ressort intuitivement des propos mêmes de Hilbert que nous avons cités, l'intérêt d'une démarche visant à asseoir les mathématiques sur une base sûre est bien d'ordre *philosophique* : il vise à couper court au *scepticisme* épistémologique. Cette dernière position n'est donc pas contradictoire avec l'idée qu'une telle démarche n'est pas utile pour le « mathématicien au travail » et elle permet ainsi de réconcilier la démarche de Hilbert (entendue comme destinée à répondre au scepticisme), avec la critique de Kreisel à l'encontre du programme au sens restreint (qu'il associe à ce qu'il nomme justement « la tradition critique » [Kreisel 1976, p. 125]).

Les objections techniques contre le Programme de Hilbert qui ont été formulées indépendamment des objections induites par l'irruption des théorèmes de Gödel de 1931 (et des résultats complémentaires établis ultérieurement) portent sur différents points. Tout d'abord à la fois Jean Largeault [Largeault 1972, p. 217, n. 2], Dag Prawitz [Prawitz 1981, p. 268] et Peter Smith [Smith 2007, p. 255-256, n. 3] ont exprimé peu ou prou l'idée générale que l'établissement de la consistance des mathématiques réelles n'était pas suffisant pour garantir celle des mathématiques idéales. Mais, comme nous l'avons déjà indiqué, l'effet de contagion induit par *l'ex falso sequitur quodlibet* qui est associé à la consistance simple (en logique classique) répond à cette objection. Une autre objection plus précise repose sur le simple constat que dans un système formel inconsistant tous les énoncés sont des théorèmes, y compris ceux qui expriment la cohérence du système formel. Cela a été formulé d'une manière lapidaire par Raymond Smullyan : « Croire à la cohérence d'un système sur la base du fait qu'il peut prouver sa propre cohérence est aussi stupide que croire à la sincérité d'une personne parce que cette dernière proclame qu'elle ne ment jamais » [Smullyan 1992 (1993), p. 104]. Dans la mesure toutefois où la démarche de Hilbert visait justement à établir la consistance simple de l'arithmétique en faisant appel à des éléments métamathématiques *externes* au système axiomatique étudié (fondés sur l'intuition des énoncés « réels »), celle-ci ne tombe pas sous le

¹²¹ Qui renvoie manifestement à la formule célèbre attribuée à d'Alembert : « Allez de l'avant et la foi vous viendra ».

coup d'une critique aussi élémentaire¹²². Enfin une objection à la fois précise et forte a également été soulevée par de nombreux auteurs, à commencer par Tarski [Tarski 1933 (1974) p. 20], qui font valoir l'insuffisance du critère de consistance simple qui conduit à autoriser des systèmes « pathologiques » consistants mais ω -inconsistants (et même seulement 1-inconsistants). Nous avons déjà abordé ce point particulier (§ 1.5) et nous y reviendrons (se reporter *supra* au § 3.2) ; nous rappelons qu'il se traduit par le fait que dans un cadre classique, certains théorèmes Σ_1 sont alors faux dans le modèle standard. Il s'avère que dans un cadre finitiste, les énoncés Σ_1 ne sont pas « réels » et ne sont donc en fait, ni vrais ni faux au sens de Hilbert. Toutefois, comme nous l'avons déjà indiqué, le problème ne disparaît pas aussi simplement : il suffit de considérer la négation d'un tel théorème, qui est à la fois Π_1 (donc « réel »), vrai (au sens de Hilbert), et *réfutable*. Là encore, il peut être répondu que l'exigence que les énoncés « réels » réfutables soient faux n'est pas une expression du Programme *stricto sensu* (se reporter à la discussion du § 1.5). Il n'en reste pas moins que l'existence de tels énoncés « pathologiques » ne s'accorde pas avec l'esprit du Programme. Comme nous l'avons indiqué ce point a d'ailleurs bien été perçu *a posteriori* (c'est-à-dire après 1930) par Bernays, qui a justement introduit sa notion de « consistance externe » pour interdire de tels énoncés. Globalement, sans contester la pertinence de la dernière citée, l'argument le plus fort contre ces critiques (qui ont toutes été formulées *a posteriori*) est certainement la réalisation au moins partielle du Programme (sur une partie limitée des mathématiques), sur laquelle nous reviendrons au § 4.

2.2 Le programme de Hilbert au sens restreint a-t-il été invalidé par le premier théorème ?

Le premier théorème de Gödel de 1931 établit l'existence d'un énoncé indécidable au sein d'un système formel T au moins aussi fort que Q ¹²³, effectif et consistant. D'une manière plus précise ses conditions d'application sont au nombre de cinq :

¹²² Bien entendu, si l'on prend en compte le fait que la métamathématique envisagée par Hilbert peut être projetée à l'intérieur de l'arithmétique (par la technique de la codification gödélienne), cette objection resurgit (elle constitue même le ressort de la démonstration du second théorème de Gödel de 1931). Mais l'on sort alors du cadre d'une contestation *d'emblée* du Programme de Hilbert, pour entrer dans celui de l'analyse de sa pertinence *a posteriori*, ce qui est justement l'objet des développements qui vont suivre.

¹²³ Q , l'arithmétique de Robinson, présente la particularité de ne pas disposer du tout de schéma d'induction (tableau I, § 1.4). Bien entendu, ce n'est pas Q (qui n'a été proposé qu'en 1952) qui est utilisé dans l'article de

- 1) le système formel doit être 1-consistant¹²⁴ ;
- 2) l'ensemble de ses expressions bien formées doit être récursif¹²⁵ ;
- 3) l'ensemble de ses dérivations doit être récursif ;
- 4) il doit permettre d'y représenter (au sens formel) toutes les FRP et les propriétés récursives primitives¹²⁶ ;
- 5) il doit vérifier une première condition de dérivabilité¹²⁷ ; si la formule Prov_T ¹²⁸ exprime (au sens formel¹²⁹) dans le langage de T la notion de dérivabilité au sein de T, et si $\ulcorner \xi \urcorner$ désigne d'une manière générale (en suivant les conventions de notation habituelles) le nombre de Gödel d'une formule quelconque ξ :

CD1) pour tout énoncé φ , si $\vdash_T \varphi$, alors $\vdash_T \text{Prov}_T(\ulcorner \varphi \urcorner)$.

Le principe de sa démonstration est rappelé ici (pour un système formel quelconque T satisfaisant les exigences mentionnées).

- 1) On établit d'abord¹³⁰ le *lemme de diagonalisation* : pour toute formule $F(n)$, il existe un énoncé ψ_T tel que : $\vdash_T \psi_T \leftrightarrow F(\ulcorner \psi_T \urcorner)$;

1931, mais un système formel dérivé des *Principia Mathematica*, système formel de force analogue à celle de PA.

¹²⁴ Dans l'article de 1931 il est fait appel à l'hypothèse (inutilement) plus forte de l' ω -consistance. Par ailleurs, le théorème a été ultérieurement renforcé par celui de Rosser de 1936 : [Rosser 1936], en n'exigeant plus alors que la consistance simple (les autres conditions restant inchangées).

¹²⁵ Un ensemble d'entiers est récursif s'il existe un algorithme permettant de décider de manière certaine, à la suite d'un nombre fini de pas, pour tout nombre entier, si celui-ci appartient ou non à cet ensemble. Dans le cas présent, cela suppose un *codage* en nombres entiers des expressions du système formel.

¹²⁶ Une propriété P ou une classe de nombres C_P est récursive primitive, s'il peut lui-être associée une fonction caractéristique f, qui est FRP ; Cette fonction caractéristique est elle-même définie de la manière suivante : 1) $f(n) = 1$ si n vérifie P (appartient à la classe C_P , 2) $f(n) = 0$ si n ne vérifie pas P (n n'appartient pas à la classe C_P).

¹²⁷ En fait cette condition, issue de [Löb 1955], n'est pas indispensable pour le premier théorème [Smorynski 1977, p. 838-839]. Nous l'avons introduite pour la commodité de la démonstration qui est présentée.

¹²⁸ La formule $\text{Prov}_T(n)$ est construite à partir de la propriété $\text{Proof}_T(\varphi)$, qui caractérise le fait d'être une dérivation (formelle) de l'énoncé φ au sein du système formel T. De manière plus précise, à cette propriété il peut être associé la fonction caractéristique suivante (où n est le nombre de Gödel de l'énoncé φ) : $\text{Proof}_T(m, n) = 1$ si m est le nombre de Gödel d'une dérivation de l'énoncé φ (quand elle existe), $\text{Proof}_T(m, n) = 0$ sinon. On montre alors (c'est une part significative du théorème) que $\text{Proof}_T(\varphi)$ est récursive primitive (que $\text{Proof}_T(m, n)$ est FRP). Elle peut donc être capturée dans un système formel répondant aux conditions d'application du théorème. $\text{Prov}_T(n)$ est alors définie de la manière suivante : $\text{Prov}_T(n) = \exists m \text{Proof}_T(m, n)$. L'énoncé $\text{Prov}_T(n)$ (obtenu pour toute assignation de n) est donc de type Σ_1 . Il n'est pas représentable dans un système formel répondant aux conditions d'application du théorème.

¹²⁹ Nous rappelons qu'une propriété P ou une classe de nombres C_P est exprimable (on dit aussi définissable) dans le langage d'un système formel T si et seulement s'il existe dans ce système formel une formule $F(n)$ telle que :

- 1) si n vérifie P (appartient à la classe C_P) alors $F(n)$ est vraie (dans le modèle standard de T) ;
 - 2) si n ne vérifie pas P (n n'appartient pas à la classe C_P) alors $F(n)$ est fausse (dans le modèle standard de T).
- Cette définition s'étend également aux fonctions numériques à une variable $f(m)$, en faisant intervenir une formule à deux variables $F(m, n)$: 1) si $f(m) = n$, alors $F(m, n)$ est vraie (dans le modèle naturel de T), 2) si $f(m) \neq n$, alors $F(m, n)$ est fausse (dans le modèle naturel de T).

¹³⁰ Cela constitue en fait le cœur de la démonstration originale.

2) On applique le lemme de diagonalisation à la négation de Prov_T , soit :

$\vdash_T G_T \leftrightarrow \neg \text{Prov}_T(\ulcorner G_T \urcorner)^{131}$, où G_T est l'énoncé de Gödel (pour le système formel T) qui présente la particularité de refléter sa propre non dérivabilité ;

3) On applique la *première condition de dérivabilité* à l'énoncé G_T :

si $\vdash_T G_T$, alors $\vdash_T \text{Prov}_T(\ulcorner G_T \urcorner)$;

4) Formulons l'hypothèse que l'on puisse dériver G_T : $\vdash_T G_T$; Par (2) on en déduit : $\vdash_T \neg \text{Prov}_T(\ulcorner G_T \urcorner)$ et par (3) : $\vdash_T \text{Prov}_T(\ulcorner G_T \urcorner)$; ces deux résultats sont incompatibles avec la condition de consistance, l'hypothèse est donc fautive et : $\not\vdash_T G_T$ (première partie du théorème) ;

5) A partir de la condition de 1-consistance, on montre¹³² la réciproque de la première condition de dérivabilité : si $\vdash_T \text{Prov}_T(\ulcorner \varphi \urcorner)$, alors $\vdash_T \varphi$; On l'applique à la fonction Prov_T : si $\vdash_T \text{Prov}_T(\ulcorner G_T \urcorner)$, alors $\vdash_T G_T$;

6) Formulons l'hypothèse que l'on puisse dériver $\neg G_T$: $\vdash_T \neg G_T$; Par (2) on en déduit : $\vdash_T \neg \neg \text{Prov}_T(\ulcorner G_T \urcorner)$, soit : $\vdash_T \text{Prov}_T(\ulcorner G_T \urcorner)$, qui devient par (5) : $\vdash_T G_T$; ce résultat et l'hypothèse sont incompatibles avec la condition de consistance, l'hypothèse est donc fautive et : $\not\vdash_T \neg G_T$ (seconde partie du théorème) ;

7) La combinaison des résultats (4) et (6) exprime finalement le fait que l'énoncé G_T est indécidable dans le système formel T.

¹³¹ G_T est un énoncé de type Π_1 (qui peut s'exprimer sous la forme : $\forall m \neg \text{Proof}_T(m, n) = 1$).

¹³² On peut s'en convaincre facilement en prenant en compte le fait que $\text{Prov}_T(\ulcorner \varphi \urcorner)$ se trouve être un énoncé Σ_1 et en utilisant le lien co-extensionnel entre la 1-consistance et la Σ_1 -correction (schéma B du § 2.4). L'argument très simple est alors celui-ci : comme T est Σ_1 -correcte et $\text{Prov}_T(\ulcorner \varphi \urcorner)$ est un énoncé Σ_1 , si $\vdash_T \text{Prov}_T(\ulcorner \varphi \urcorner)$ alors $\text{Prov}_T(\ulcorner \varphi \urcorner)$ est vrai (dans le modèle standard de T) et donc $\vdash_T \varphi$ (par définition de la fonction Prov_T). On peut également faire intervenir le « schéma de réflexion pour les énoncés Π_1 », qui s'exprime par la dérivation : $\vdash_T \text{Prov}_T(\ulcorner \varphi \urcorner) \rightarrow \varphi$ où φ est Π_1 .

Ainsi, sur la base des résultats établis au § 1.3, c'est-à-dire en ne prenant en compte que les deux premiers volets de la thèse de Tait, et indépendamment de la prise en compte du troisième volet, le débat sur le programme de Hilbert au sens restreint semble clos avec une alternative dirimante. Si l'on choisit pour S^{133} un système formel suffisamment faible pour que le premier théorème de Gödel ne s'y applique pas (par exemple ERA mentionné dans le tableau I du § 1.4), la faiblesse du système ne lui permet pas de répondre à l'attente de Hilbert (ne lui permet pas de « gravir les échelons »). Si au contraire on choisit pour S un système ayant la force minimale pour que le premier théorème de Gödel s'y applique (typiquement PRA¹³⁴, selon le troisième volet de la thèse de Tait), celui-ci entraîne l'existence d'un énoncé indécidable, dans S lorsqu'il est appliqué à S et a fortiori dans T lorsqu'il est appliqué à T . Si l'on fait l'hypothèse supplémentaire que l'énoncé indécidable de S est au contraire dérivable dans un système formel suffisamment fort pris pour T (typiquement Z_2), il n'est alors plus possible de vérifier le caractère conservatif de T par rapport à S pour les énoncés Π_1 qui correspond à l'expression technique du programme de Hilbert (avec la thèse de Tait). Le Programme de Hilbert semble bien réfuté dans les deux situations. De fait, la plupart des philosophes des fondements des mathématiques, au moins depuis Georg Kreisel [Kreisel 1968, p. 323], ont conclu en faveur d'une réfutation du programme de Hilbert par le *premier* théorème de Gödel, à l'exception notable de Michael Detlefsen (et Marcus Giaquinto) comme le rappelle Panu Raatikainen [Raatikainen 2003, p. 162]. On constate toutefois à la lecture des documents concernés, que les justifications qui sont apportées en faveur de cette réponse ne sont pas exemptes de reproches et en sens inverse que les critiques soulevées contre ces justifications ne sont pas dénuées de pertinence. Nous allons donc analyser les arguments développés par Michael Detlefsen, Marcus Giaquinto, et Panu Raatikainen, en nous limitant à ceux qui s'appliquent au Programme au sens strict, et en excluant donc ceux qui dépendent de l'engagement philosophique de Michael Detlefsen en faveur d'un instrumentalisme fort, qui seront traités dans un second temps (au § 3.4)¹³⁵.

Il est tout d'abord indéniable que ce n'est pas en soi l'application du premier théorème à PRA pris pour S (l'existence d'un énoncé indécidable dans PRA) qui permet de condamner le Programme : avec Michael Detlefsen contre Craig Smorynski. Nous retrouvons là très

¹³³ Nous rappelons que nous désignons par R l'ensemble des énoncés « réels », I l'ensemble des énoncés « idéaux », S le système formel (ou l'ensemble des énoncés de celui-ci) traduisant R avec l'exigence de finitisme et T le système formel (ou l'ensemble des énoncés de celui-ci) traduisant $I + R$ sans l'exigence de finitisme.

¹³⁴ En fait, comme cela est indiqué dans le tableau I du § 1.4, Q , plus faible que PRA, est le système formel minimal auquel s'applique le premier théorème de Gödel.

¹³⁵ Sachant que cette distinction n'est évidemment pas effectuée dans les textes concernés : [Detlefsen 1986], [Detlefsen 1990] et [Giaquinto 2002].

exactement la quatrième solution du § 1.5 et la raison qui nous a conduit à l'éliminer : la complétude syntaxique de S n'est pas une exigence du Programme¹³⁶. Cet argument concerne tous les énoncés indécidables *communs* à S et à T, sachant que dans les discussions sur la portée du premier théorème contre le Programme, il est important de savoir si l'énoncé de Gödel du premier théorème est supposé être appliqué à S ou bien à T. Ce point reste ambigu en 1986 [Detlefsen 1986, p. 164, 165] et n'est clarifié qu'en 1990 [Detlefsen 1990, p. 357-358]. Par contre, c'est bien le cas particulier de l'existence d'énoncés indécidables dans PRA et *dérivables* dans T qui s'oppose au Programme¹³⁷, en invalidant alors la condition que T soit conservatif par rapport à S pour les énoncés Π_1 . Il faut pour cela s'assurer d'un point technique : que l'énoncé G_S du premier théorème appliqué à S, qui est sans ambiguïté un énoncé Π_1 , soit bien dérivable dans T¹³⁸. Or ce n'est pas nécessairement vrai pour tout système formel (plus puissant que PRA) pris pour T, mais l'important est que c'est le cas pour le système du second ordre Z_2 ¹³⁹, qui est bien conforme à ce que Hilbert avait en vue pour son Programme. A cet argument, la réponse de Michael Detlefsen est la revendication que ce soit S lui-même et non pas R qui serve de référentiel au finitisme [Detlefsen 1986, p. 163, n. 4] et [Detlefsen 1990, p. 360]. Cette réponse n'est évidemment pas acceptable dans le cadre de la présente discussion puisqu'elle revient à en changer les « règles du jeu ». Quant à l'argument purement rhétorique de Marcus Giaquinto qui consiste en un renversement de la charge de la preuve [Giaquinto 2002, p. 180], il ne l'est bien sûr pas d'avantage. Ces deux dernières objections ayant été éliminées, nous retrouvons finalement la position classique d'une opposition du premier théorème de Gödel de 1931 au Programme de Hilbert, mais sur la base d'un argument beaucoup plus précis que ce qui en est habituellement proposé comme justification.

¹³⁶ A fortiori, Michael Detlefsen a raison de contester que le fait que l'énoncé de Gödel du premier théorème appliqué à S soit « vrai », suffise pour s'opposer au Programme [Detlefsen 1986, p. 162-163] : c'est bien entendu un critère finitiste qui doit être retenu (en l'occurrence la vérifiabilité) et non la vérité au sens classique (dans le modèle standard de S).

¹³⁷ Comme par construction, T étend S, la réciproque ne peut pas survenir : il ne peut pas y avoir des énoncés indécidables dans T et décidables dans S.

¹³⁸ On peut en effet concevoir qu'un énoncé indécidable dans S et dérivable dans T soit en fait tributaire des conditions d'application du *second* théorème de Gödel, qui ne sont pas les mêmes que celles du premier. Panu Raatikainen est ainsi conduit, pour cette raison, à revenir sur un argument technique (donc indépendant d'un critère adapté pour le finitisme) qu'il a présenté pour rendre compte de la même exigence que T soit conservatif par rapport à S pour les énoncés Π_1 [Raatikainen 2003, p. 162].

¹³⁹ Même s'il existe aussi des énoncés indécidables dans PRA et décidables dans Z_2 qui échappent au premier théorème, comme celui qui affirme la consistance de PA.

2.3 Le programme de Hilbert au sens restreint a-t-il été invalidé par le second théorème ?

Le second théorème de Gödel de 1931 établit l'impossibilité de dériver dans un système formel T consistant, un énoncé codant dans celui-ci l'expression de sa propre consistance¹⁴⁰. Ses conditions d'applications sont celles du premier théorème, mais où l'exigence de consistance est limitée à celle de la consistance simple (au lieu de la 1-consistance), avec par contre deux nouvelles conditions de dérivabilité. Ces deux conditions de dérivabilités sont les suivantes¹⁴¹, si φ et ψ sont deux énoncés quelconques du système formel T :

$$\text{CD2) } \vdash_T \text{Prov}_T(\ulcorner \varphi \rightarrow \psi \urcorner) \rightarrow (\text{Prov}_T(\ulcorner \varphi \urcorner) \rightarrow \text{Prov}_T(\ulcorner \psi \urcorner)),$$

ce qui implique, par le *Modus Ponens* :

$$\text{si } \vdash_T \text{Prov}_T(\ulcorner \varphi \rightarrow \psi \urcorner), \text{ alors } \vdash_T \text{Prov}_T(\ulcorner \varphi \urcorner) \rightarrow \text{Prov}_T(\ulcorner \psi \urcorner) ;$$

$$\text{CD3) } \vdash_T \text{Prov}_T(\ulcorner \varphi \urcorner) \rightarrow \text{Prov}_T(\ulcorner \text{Prov}_T(\ulcorner \varphi \urcorner) \urcorner),$$

ce qui implique, par le *Modus Ponens* :

$$\text{si } \vdash_T \text{Prov}_T(\ulcorner \varphi \urcorner), \text{ alors } \vdash_T \text{Prov}_T(\ulcorner \text{Prov}_T(\ulcorner \varphi \urcorner) \urcorner).$$

Bien qu'un peu fastidieuse la démonstration intégrale¹⁴² du second théorème est rappelée ici (pour un système formel quelconque T satisfaisant les exigences mentionnées), compte tenu de son importance pour la discussion qui va suivre.

1) On reprend du premier théorème le lemme de diagonalisation appliqué à la négation de la fonction Prov_T :

$$\vdash_T G_T \leftrightarrow \neg \text{Prov}_T(\ulcorner G_T \urcorner) ;$$

2) On retient la première moitié de ce lemme :

¹⁴⁰ La présence de la condition de consistance bloque ainsi l'argument de Smullyan que nous avons déjà évoqué : la pertinence du second théorème n'est pas que le système formel ne permette pas de dériver (le codage de) sa propre consistance : une telle dérivation ne prouverait nullement cette consistance. Elle est dans la réfutation que l'on puisse à la fois, d'une part établir la consistance (simple) du système formel en faisant appel à des éléments *nécessairement extérieurs à lui* (en l'occurrence la vérifiabilité des énoncés « réels » vrais sur une base intuitive), et d'autre part dériver (un codage de) cette consistance au sein de celui-ci.

¹⁴¹ Elles sont exprimées ici sous la forme proposée par [Löb 1955]. Avec la condition CD1 mentionnée au § 2.2, elles ont été introduites pour la première fois dans [Hilbert, Bernays 1939]. En fait, la condition CD2, indispensable pour le théorème de Löb de 1955, s'est avéré ultérieurement n'être pas indispensable pour le second théorème de Gödel de 1931 : [Jerusalem 1973].

¹⁴² Qui suppose tout de même, pour l'appliquer à une théorie donnée (par exemple PA ou PRA), que l'on ait montré en outre que les conditions de dérivabilité s'y appliquent bien.

$$\vdash_{\mathcal{T}} G_{\mathcal{T}} \rightarrow \neg \text{Prov}_{\mathcal{T}}(\ulcorner G_{\mathcal{T}} \urcorner);$$

3) On applique à (2) la première condition de dérivabilité :

$$\vdash_{\mathcal{T}} \text{Prov}_{\mathcal{T}}(\ulcorner G_{\mathcal{T}} \rightarrow \neg \text{Prov}_{\mathcal{T}}(\ulcorner G_{\mathcal{T}} \urcorner) \urcorner);$$

4) On applique à (3) la deuxième condition de dérivabilité :

$$\vdash_{\mathcal{T}} \text{Prov}_{\mathcal{T}}(\ulcorner G_{\mathcal{T}} \urcorner) \rightarrow \text{Prov}_{\mathcal{T}}(\ulcorner \neg \text{Prov}_{\mathcal{T}}(\ulcorner G_{\mathcal{T}} \urcorner) \urcorner);$$

5) On démontre le lemme suivant, pour deux énoncés φ et ψ quelconques :

$$\vdash_{\mathcal{T}} \text{Prov}_{\mathcal{T}}(\ulcorner \neg \varphi \urcorner) \rightarrow \text{Prov}_{\mathcal{T}}(\ulcorner \varphi \rightarrow \psi \urcorner);$$

a) On utilise un théorème classique de la logique des propositions :

$$\vdash_{\mathcal{T}} \neg \varphi \rightarrow (\varphi \rightarrow \psi);$$

b) On applique à (5a) la première condition de dérivabilité :

$$\vdash_{\mathcal{T}} \text{Prov}_{\mathcal{T}}(\ulcorner \neg \varphi \rightarrow (\varphi \rightarrow \psi) \urcorner);$$

c) On applique à (5b) la deuxième condition de dérivabilité :

$$\vdash_{\mathcal{T}} \text{Prov}_{\mathcal{T}}(\ulcorner \neg \varphi \urcorner) \rightarrow \text{Prov}_{\mathcal{T}}(\ulcorner \varphi \rightarrow \psi \urcorner);$$

6) On applique le lemme (5) à l'énoncé $\text{Prov}_{\mathcal{T}}(\ulcorner G_{\mathcal{T}} \urcorner)$ et à l'énoncé \perp traduisant une contradiction (quelconque) :

$$\vdash_{\mathcal{T}} \text{Prov}_{\mathcal{T}}(\ulcorner \neg \text{Prov}_{\mathcal{T}}(\ulcorner G_{\mathcal{T}} \urcorner) \urcorner) \rightarrow \text{Prov}_{\mathcal{T}}(\ulcorner \text{Prov}_{\mathcal{T}}(\ulcorner G_{\mathcal{T}} \urcorner) \rightarrow \perp \urcorner);$$

7) On applique à (4) et (6) (le théorème qui traduit) la transitivité de l'implication matérielle :

$$\vdash_{\mathcal{T}} \text{Prov}_{\mathcal{T}}(\ulcorner G_{\mathcal{T}} \urcorner) \rightarrow \text{Prov}_{\mathcal{T}}(\ulcorner \text{Prov}_{\mathcal{T}}(\ulcorner G_{\mathcal{T}} \urcorner) \rightarrow \perp \urcorner);$$

8) On applique au conséquent de (7) la deuxième condition de dérivabilité :

$$\vdash_{\mathcal{T}} \text{Prov}_{\mathcal{T}}(\ulcorner \text{Prov}_{\mathcal{T}}(\ulcorner G_{\mathcal{T}} \urcorner) \rightarrow \perp \urcorner) \rightarrow (\text{Prov}_{\mathcal{T}}(\ulcorner \text{Prov}_{\mathcal{T}}(\ulcorner G_{\mathcal{T}} \urcorner) \urcorner) \rightarrow \text{Prov}_{\mathcal{T}}(\ulcorner \perp \urcorner));$$

9) On applique à (7) et (8) (le théorème qui traduit) la transitivité de l'implication matérielle :

$\vdash_T \text{Prov}_T(\ulcorner G_T \urcorner) \rightarrow (\text{Prov}_T(\ulcorner \text{Prov}_T(\ulcorner G_T \urcorner) \urcorner) \rightarrow \text{Prov}_T(\ulcorner \perp \urcorner))$;

10) On applique à G_T la troisième condition de dérivabilité :

$\vdash_T \text{Prov}_T(\ulcorner G_T \urcorner) \rightarrow \text{Prov}_T(\ulcorner \text{Prov}_T(\ulcorner G_T \urcorner) \urcorner)$;

11) On applique à (10) et (9) un résultat général du calcul des propositions :

$\vdash_T \text{Prov}_T(\ulcorner G_T \urcorner) \rightarrow \text{Prov}_T(\ulcorner \perp \urcorner)$;

12) On prend la contraposée de (11) :

$\vdash_T \neg \text{Prov}_T(\ulcorner \perp \urcorner) \rightarrow \neg \text{Prov}_T(\ulcorner G_T \urcorner)$;

13) On introduit dans (12) l'énoncé $\text{Con}_T = \neg \text{Prov}_T(\ulcorner \perp \urcorner)$ ¹⁴³ qui traduit dans T la consistance simple de T :

$\vdash_T \text{Con}_T \rightarrow \neg \text{Prov}_T(\ulcorner G_T \urcorner)$ ¹⁴⁴ ; ce résultat est la formalisation du premier théorème au sein de T ;

14) Formulons l'hypothèse que l'on puisse dériver dans T le codage de la consistance simple de T :

$\vdash_T \text{Con}_T$;

15) On applique le *Modus Ponens* à (14) et à (13) :

$\vdash_T \neg \text{Prov}_T(\ulcorner G_T \urcorner)$;

16) On déduit directement de (15) et de (1) :

$\vdash_T G_T$;

¹⁴³ L'énoncé Con_T est un énoncé Π_1 , puisque nous avons vu que $\text{Prov}_T(n)$ est un énoncé Σ_1 .

¹⁴⁴ Même si ce n'est pas utile pour la démonstration, on montre facilement que l'on a en fait la relation plus générale : $\vdash_T \text{Con}_T \leftrightarrow \neg \text{Prov}_T(\ulcorner G_T \urcorner)$ et donc, via 1) : $\vdash_T G_T \leftrightarrow \text{Con}_T$ [Smorynski 1977, p. 828] et [Smith 2007, p. 224].

17) Le résultat (16) est contradictoire avec la première partie du premier théorème¹⁴⁵, donc l'hypothèse (14) est fautive et :

$\not\vdash_T \text{Con}_T$ ¹⁴⁶.

Cette démonstration amène deux remarques concernant la différence entre les deux théorèmes. Tout d'abord, il apparaît crucial dans le second théorème que l'énoncé Con_T traduise bien la signification de la consistance simple au sein de T et donc que la fonction Prov_T traduise bien la signification de la dérivabilité au sein de T , ce qui revient à introduire une notion *intensionnelle*¹⁴⁷ (faisant intervenir le particularisme du codage de Gödel utilisé) au sein même du système formel. A contrario, pour le premier théorème, il est tout à fait secondaire à la démonstration que l'énoncé G_T « dise de lui-même qu'il est indécidable » : ce n'est là que l'artefact d'une certaine sémantique informelle qui peut lui être associée dans un but pédagogique ; la seule chose importante est que G_T soit indécidable, quoi qu'il signifie. Cela conduit bien entendu, pour l'établissement du second théorème, à la nécessité de justifier que la notion de preuve au sein du système formel ait été correctement traduite. Ensuite, pour pouvoir conduire la démonstration du second théorème il est nécessaire d'introduire les conditions de dérivabilité particulières¹⁴⁸ (CD1), CD2, et CD3. En fait, il est apparu depuis [Rosser 1936] et surtout depuis [Feferman 1960] que ces deux remarques sont liées ([Feferman 1990] est particulièrement éclairant sur le sujet). Si les conditions de dérivabilité sont respectées, cela contraint les formes possibles (du codage) de l'expression de la prouvabilité qui les rend compatibles avec la démonstration du second théorème. En sens inverse, si l'une des conditions n'est pas respectée (comme CD2 n'est pas indispensable et CD1 est en général trivialement respectée, c'est CD3 qui est la condition cruciale), alors il est possible d'introduire certaines formulations de prouvabilité et partant certaines formulations de consistance pour lesquelles un système formel répondant aux conditions du premier théorème permet de dériver (le codage de) l'expression de sa propre consistance, en contradiction avec le second théorème. C'est bien entendu par ce biais qu'ont été introduites les objections à l'utilisation du second

¹⁴⁵ Nous avons déjà signalé qu'elle n'est soumise qu'à la condition de consistance simple (c'est seulement la seconde partie du premier théorème qui est soumise à la condition plus forte de 1-consistance).

¹⁴⁶ On rappelle que le second théorème exprime la non dérivabilité mais non l'indécidabilité de l'énoncé exprimant (un codage de) la consistance du système formel.

¹⁴⁷ [Feferman 1960] et [Auerbach 1985] y ont beaucoup insisté.

¹⁴⁸ Ce qui explique également que le second théorème ne s'applique qu'à des théories plus fortes que celles auxquelles s'applique le premier (voir le tableau I du § 1.4), en imposant en particulier la présence du principe d'induction sous une forme minimale. C'est ainsi par exemple que le premier théorème s'applique à l'arithmétique de Robinson : Q , mais non le second.

théorème contre le programme de Hilbert¹⁴⁹, principalement par Michael Detlefsen. Nous allons donc analyser les arguments de celui-ci et les réponses qui lui ont été faites ou qui peuvent lui être faites. Pour cela, outre la technicité inévitable de la discussion, il y a deux difficultés à surmonter. Tout d'abord, la position de Michael Detlefsen est intimement liée à sa propre interprétation du programme de Hilbert, qui n'est pas celle que nous avons retenue et sur laquelle porte notre discussion. Cela nous conduit à introduire artificiellement une distinction qui ne se trouve évidemment pas dans les écrits de celui-ci, en ne retenant dans ce paragraphe que la part de l'argumentation de l'auteur qui est compatible avec notre interprétation, sachant que nous analyserons dans un second temps cette interprétation pour elle-même (au § 3.4). Ensuite, l'argumentation de Michael Detlefsen est d'une nature différente dans [Detlefsen 2001] d'une part et dans les publications précédentes (principalement [Detlefsen 1986]) de l'autre, sachant que [Detlefsen 2001] n'a pas été à ce jour et à notre connaissance, l'objet d'une critique publiée¹⁵⁰, contrairement à [Detlefsen 1986]. Cela nous conduit à séparer notre analyse en distinguant les deux argumentations et en commençant par la seconde citée et première publiée (la plus classique). Nous achèverons cette analyse en discutant un argument fort et direct contre le Programme, utilisant une version dérivée du second théorème de Gödel qui *s'affranchit des conditions de dérivabilité*. Bien qu'ancien et émanant de Gödel lui-même : [Gödel 1972b], il a été signalé par [Feferman 1990] mais n'a été remis au goût du jour que récemment par [Potter 2000] et [Zach 2005a].

La contestation « classique » par Michael Detlefsen de l'invalidation du programme de Hilbert au sens restreint par le second théorème, se trouve principalement dans [Detlefsen 1986] (elle est annoncée dans [Detlefsen 1979] et rappelée dans [Detlefsen 1990]). Elle porte sur la justification philosophique de la condition CD3 (qu'il nomme F-LPC) qu'il estime non établie, ce qui permet donc de s'en affranchir. Il procède pour cela en trois étapes. Tout d'abord, il considère que c'est à celui (qu'il nomme « anti-hilbertien ») qui prétend invalider le Programme à partir du second théorème (G2) de justifier la nécessité d'utiliser la condition CD3 (selon la tactique argumentative classique de renversement de la charge de la preuve). Ensuite il estime que la seule manière d'y parvenir est de montrer que toutes les expressions raisonnables possibles de la consistance impliquent l'utilisation des conditions de dérivabilité (ce qu'il

¹⁴⁹ L'argument classique utilisant le second théorème contre le Programme est le suivant : si une théorie aussi simple que PA (a fortiori PRA) ne peut pas établir par elle-même sa propre consistance, alors elle pourra encore moins établir celle d'une théorie plus forte, interdisant ainsi de « gravir les échelons ».

¹⁵⁰ En effet, si [Roepler 2003] se réfère bien exclusivement à [Detlefsen 2001], il ne s'agit en fait que d'une réflexion très générale sur la représentativité de la dérivabilité au sein d'une théorie, qui ne prend pas du tout en compte la spécificité de l'analyse de Michael Detlefsen.

appelle « le problème de stabilité ») [Detlefsen 1986, p. 100]. Il analyse alors trois justifications de l'utilisation de la condition de dérivabilité CD3, qu'il attribue respectivement à [Mostowski 1966], [Kreisel, Takeuti 1974] et [Prawitz 1981], et en conclut qu'aucune n'est satisfaisante [Detlefsen 1986, p. 81-82]. Il en déduit la justification de la possibilité de s'affranchir de la condition CD3 et d'utiliser des formes de consistance « d'esprit consistant », basées (nous allons y revenir) sur des solutions « à la Rosser 1936 », tout en éliminant les solutions « à la Feferman 1960 » à cause du caractère non effectif de ces dernières [Detlefsen 1986, p. 123]. D'une manière fort subtile mais discrète (dans une note de bas de page), Michael Detlefsen n'affirme pas que la solution qu'il propose valide le Programme, mais seulement qu'il n'est pas prouvé qu'elle ne puisse pas le faire :

Premièrement nous ne sommes pas en train d'affirmer qu'elles [les variantes de Rosser] montrent que le programme de Hilbert *peut* être conduit [...] Deuxièmement, nous ne sommes pas en train d'affirmer que la découverte d'un nouveau type de résultat de type G2 s'étendant aux systèmes de Rosser est impossible. [Detlefsen 1986, p. 140, n. 48 ; le commentaire entre crochets est nôtre]

Le cœur de la critique de Michael Detlefsen, à savoir la contestation des justifications de l'utilisation de la condition de dérivabilité CD3 attribuée à [Mostowski 1966], [Kreisel, Takeuti 1974] et [Prawitz 1981] n'est pas sans intérêt, mais n'est pas non plus exempt de reproches. Nous suivons Michael Detlefsen, contre [Mostowski 1966] (selon [Detlefsen 1986]), sur le fait qu'il est indéniable que la condition CD1 est plus évidente (ou plus intuitive) que CD3. En sens inverse, la proposition de [Kreisel, Takeuti 1974] (toujours selon [Detlefsen 1986]) selon laquelle les conditions CD2 et CD3 sont justifiées par la nécessité que les théories idéales formalisées capturent la pratique mathématique informelle¹⁵¹, est à la fois intéressante et compatible avec l'esprit du Programme au sens classique¹⁵². Il est d'ailleurs à noter que des trois justifications analysées, c'est la seule pour laquelle la critique de Michael Detlefsen repose exclusivement sur la spécificité de l'instrumentalisme qu'il défend, qui sort du cadre de la présente discussion. Quant à l'argument attribué à [Prawitz 1981], celui-ci n'est pas recevable, dans la mesure où il assimile le programme de Hilbert au fait « d'obtenir une preuve finitaire de la correction « réelle » d'un système donné des mathématiques « idéales » à *partir d'une preuve finitaire de consistance* » [Detlefsen 1986, p. 125 ; souligné par nous]. Cet argument

¹⁵¹ Pour CD2, la requête est celle de la clôture sous le *Modus Ponens*. Pour CD3, elle est que la pratique idéale, informelle, englobe toutes les mathématiques élémentaires dans son entièreté et de ce fait inclut toutes les vérités exprimées par les énoncés Σ_1 .

¹⁵² Même si le lien introduit pour cela entre la condition CD3 et la Σ_1 -complétude reste contestable, à la fois par sa non évidence et par le fait que les énoncés Σ_1 sont « idéaux ».

n'est en effet clairement pas compatible avec l'expression générale du Programme qui résulte de la prise en compte de la condition de finitisme¹⁵³, où c'est la correction de la partie « réelle » des mathématiques « idéales » qui doit être établie *de manière finitiste*, ce que l'on peut à la rigueur exprimer sous la forme d'une exigence de correction « réelle », mais dont le contenu reste alors à préciser (il doit faire intervenir la vérifiabilité au sens de Bernays). Une telle position permet ainsi à Michael Detlefsen de défendre à bon compte une atténuation de la force du second théorème contre le Programme [Detlefsen 1986, p. 129]. Cette remarque conduit à l'objection majeure¹⁵⁴ que l'on peut opposer à Michael Detlefsen : pour défendre le Programme contre le second théorème, il ne suffit pas de contester la faiblesse de la justification de la condition de dérivabilité CD3, il faut encore montrer que l'alternative proposée, en l'occurrence une solution « à la Rosser », reste compatible avec le Programme et en particulier avec ses exigences finitistes. C'est ce que Michael Detlefsen n'a pas fait, ce qui peut se comprendre dans la mesure où il se limite à sa propre interprétation (instrumentaliste), mais ce qui atténue alors fortement la portée de sa contestation, au-delà de l'ambiguïté de la formulation de celle-ci [Detlefsen 1990, p. 345]. C'est bien sûr sur ce point que les coups des différents commentateurs de [Detlefsen 1986] se sont portés, avec une belle unanimité¹⁵⁵.

Les critiques majeures contre [Detlefsen 1986], et plus généralement contre l'idée d'utiliser des systèmes formels qui s'affranchissent de la condition de dérivabilité CD3 en vue de sauver le programme de Hilbert au sens restreint, sont récentes¹⁵⁶ : [Giaquinto 2002], [Raatikainen 2003], [Zach 2005a], [Smith 2007] et [Zach 2007]. Elles accordent toutes à Michael Detlefsen, que les justifications philosophiques de l'utilisation de la condition de dérivabilité CD3 qui ont été proposées ne sont pas dirimantes. Panu Raatikainen va jusqu'à proposer d'accepter cette condition a priori, sous la forme d'une thèse « à la Church » qui resterait à justifier [Raatikainen 2003, p. 174, n. 8] en reprenant d'ailleurs ainsi la suggestion plus ancienne de Michael Resnik [Resnik 1974, p. 127]. Ces critiques, nombreuses et variées, contestent toutes le fait que les solutions techniques proposées respectent l'esprit du Programme ou sont pertinentes par rapport à lui. Comme nous l'avons annoncé, les solutions techniques s'affranchissant de CD3¹⁵⁷

¹⁵³ Au-delà du fait que cette correction réelle y est encore traduite par la Σ_1 -complétude.

¹⁵⁴ Au-delà du fait que toute sa critique est formulée contre de ce qu'il appelle : « l'argument standard » (en faveur d'une invalidation du programme par le second théorème) et qui est en fait [Detlefsen 1986, p. 78-79], une restitution particulièrement laborieuse et compliquée de l'argument très simple qui est habituellement proposé en ce sens et que nous avons déjà mentionné [n. 148].

¹⁵⁵ Si, comme nous l'avons indiqué, Marcus Giaquinto et dans une moindre mesure Panu Raatikainen suivent Michael Detlefsen dans sa critique de l'effet du premier théorème sur le Programme, aucun de ces deux auteurs ne le suit dans son explication contre l'effet du second théorème.

¹⁵⁶ Au-delà de [Resnik 1974] qui est sans doute la plus ancienne.

¹⁵⁷ Solomon Feferman en présente un certain nombre d'exemples [Feferman 1990, p. 282-283].

constituent deux grandes familles : les solutions « à la Rosser » (issues de [Rosser 1936]) et les solutions « à la Feferman » (issues de [Feferman 1960]). Par rapport à un système formel ordinaire, la première consiste en une modification des règles de génération des théorèmes, tandis que la seconde se traduit par un changement dans la description des axiomes du système formel. Une caractéristique majeure des solutions « à la Feferman » est que celles-ci ne sont pas effectives (l'ensemble des axiomes du système formel obtenu ainsi n'est pas récursif), ce qui est rédhibitoire dans le cadre de la présente discussion. Elles ont de fait été éliminées par Michael Detlefsen pour cette raison, comme nous l'avons déjà indiqué. Les solutions « à la Rosser » ne présentent pas ce défaut et permettent de construire, pour toute théorie ordinaire T , une théorie modifiée T^R qui en conserve le langage, les axiomes, et les règles d'inférence, mais permet de dériver une certaine expression (du codage) de sa consistance. Les critiques qui ont été formulées contre l'utilisation de solutions « à la Rosser » dans le but de sauver le Programme de l'effet du second théorème, sont essentiellement au nombre de cinq. Nous les présentons selon un ordre de force croissante (selon notre appréciation).

1) Peter Smith fait valoir le caractère arbitraire du choix d'une théorie de Rosser particulière T^R (à partir d'une même théorie classique donnée T), sur des critères nécessairement externes à la théorie et qui ne sont pas accompagnés d'une justification finitiste [Smith 2007, p. 244]. La réponse de Michael Detlefsen serait sans doute qu'il propose justement une justification philosophique du choix effectué, bien sûr dans le cadre instrumentaliste qui est le sien.

2) Peter Smith dénonce encore la spécificité de l'expression de consistance obtenue et la faiblesse de son lien avec la consistance au sens classique [Smith 2007, p. 245]. Il mentionne en particulier le fait qu'une preuve « d'esprit consistant » laisse ouverte la possibilité que l'ensemble des théorèmes de la théorie T soit sémantiquement inconsistant, donc n'ait pas d'interprétation formelle (de modèle). La limite de portée de cette critique est qu'elle sort du domaine du finitisme de Hilbert.

3) Dans un registre proche, Panu Raatikainen souligne la possibilité technique, à partir d'une théorie T *inconsistante* (mais non absolument inconsistante), d'obtenir une théorie T^R dont on pourrait montrer au sein de PRA qu'elle est « d'esprit consistant » [Raatikainen 2003, p. 165]. La limite de portée de cette critique est cette fois qu'on sort du domaine d'application du second théorème (ainsi d'ailleurs que du premier).

4) Marcus Giaquinto fait remarquer que le défaut majeur des solutions « à la Rosser » est qu'elles ne respectent pas la règle de coupure¹⁵⁸, ce qui ne leur permet pas de rendre compte

¹⁵⁸ Qui permet d'accepter un énoncé en tant que théorème sur la seule base du fait d'être dérivé à partir de prémisses qui ont été dérivées auparavant (et qui ne sont pas des axiomes).

d'une manière satisfaisante de l'arithmétique ordinaire [Giaquinto 2002, p. 188]. Une telle contrainte s'oppose indéniablement à l'esprit du Programme, dont l'objectif est de rendre compte de l'ensemble des mathématiques ordinaires.

5) Marcus Giaquinto (ainsi que Peter Smith de manière plus indirecte) souligne enfin que le caractère crucial pour la théorie T^R d'être un reflet formel fidèle de la théorie T à partir de laquelle elle est construite, ne peut pas être lui-même établi au sein de la théorie T donc a fortiori dans PRA [Giaquinto 2002, p. 188] et [Smith 2007, p. 258]. Il en résulte une contrainte majeure pour le Programme, dans la mesure où il n'est alors pas possible de justifier finitairement (dans PRA¹⁵⁹) le passage de T à T^R , donc l'utilisation même de T^R .

De l'ensemble de ces critiques et surtout de la cinquième citée, il ressort que l'utilisation d'une solution « à la Rosser » n'apparaît pas compatible avec les exigences du Programme selon l'interprétation que nous avons retenue pour celui-ci. Comme les solutions « à la Feferman » ont déjà été éliminées pour des raisons analogues, cela condamne la possibilité de principe de contourner le second théorème en s'affranchissant de la condition de dérivabilité CD3, au moins pour ce qui concerne les solutions connues à ce jour pour cela. Bien sûr, comme le remarque Marcus Giaquinto [Giaquinto 2002, p. 189], cette conclusion laisse encore ouverte la possibilité de principe d'une solution nouvelle hypothétique, s'affranchissant de la condition de dérivabilité CD3, mais différente des familles « à la Rosser » et « à la Feferman » et qui échapperait aux contraintes qui viennent d'être mentionnées.

La contestation de l'invalidation du programme de Hilbert au sens restreint par le second théorème que l'on trouve dans [Detlefsen 2001], est à la fois extrêmement spécifique, très technique, et particulièrement subtile. Son point de départ est la remarque non contestable, que si l'expression technique du Programme de Hilbert distingue soigneusement les deux systèmes formels que nous avons appelés respectivement S et T , le second théorème de Gödel et les conditions de dérivabilité associées sont classiquement exprimés au sein d'un système formel unique¹⁶⁰. L'idée est alors de reformuler les conditions de dérivabilité et la démonstration du second théorème, en distinguant S (le système dans lequel on effectue la démonstration : typiquement PRA) et T (le système dont on cherche à établir la consistance : Z_2 ou PA). Selon Michael Detlefsen, il en résulte de nouvelles formulations des conditions de dérivabilité, ainsi qu'une quatrième condition spécifique (qui se ramène à une tautologie dans le cas limite où les

¹⁵⁹ Même sans retenir l'interprétation standard du troisième volet de la thèse de Tait, la seule justification finitaire acceptable *pour cette condition* du passage de T à T^R est bien qu'elle puisse être établie dans PRA.

¹⁶⁰ Il est à noter que [Smorynski 1977] introduit cependant cette distinction dans sa présentation des deux théorèmes de Gödel, mais c'est une situation rare.

deux systèmes formels sont identiques). En utilisant alors trois arguments de nature différente, celui-ci entend montrer que dans le cas le plus général des caractéristiques de S et de T, les formulations des conditions de dérivabilité obtenues et en particulier la quatrième introduite, conduisent à des contraintes techniques et philosophiques qui militent contre leur prise en compte et donc contre la portée du second théorème contre le Programme. Avant même une discussion de l'argumentation de [Detlefsen 2001] une objection (faible) vient naturellement à l'esprit : une contrainte sur l'utilisation du second théorème qui ne s'applique *que* dans le cas particulier où l'on distingue S et T, ne peut pas être d'un grand secours pour le programme de Hilbert au moins selon l'interprétation standard. Elle ne peut en effet empêcher que dans le cas extrême où T est confondu avec S, l'impossibilité d'établir la consistance de PRA avec les moyens de PRA (donc de manière finitiste), ne permet déjà pas de répondre au pré-requis du Programme. La distinction de S et de T conduit assez naturellement aux nouvelles conditions de dérivabilité suivantes, que propose par exemple Craig Smorynski [Smorynski 1977, p. 827] :

CD1') pour tout énoncé φ , si $\vdash_T \varphi$, alors $\vdash_S \text{Prov}_T(\ulcorner \varphi \urcorner)$;

CD2') $\vdash_S \text{Prov}_T(\ulcorner \varphi \rightarrow \psi \urcorner) \rightarrow (\text{Prov}_T(\ulcorner \varphi \urcorner) \rightarrow \text{Prov}_T(\ulcorner \psi \urcorner))$;

CD3') $\vdash_S \text{Prov}_T(\ulcorner \varphi \urcorner) \rightarrow \text{Prov}_T(\ulcorner \text{Prov}_T(\ulcorner \varphi \urcorner) \urcorner)$.

Avec ces nouvelles conditions, la démonstration du second théorème que nous avons présentée peut-être reconduite, en remplaçant simplement presque partout la dérivation dans T par la dérivation dans S au prix d'une contrainte faible. Par « presque partout » il est entendu toutes les étapes à l'exception de 1), 2) et 16) pour lesquelles la dérivation dans T est maintenue. Quant à la contrainte, elle se traduit par le rajout après l'étape 15) : $\vdash_S \neg \text{Prov}_T(\ulcorner G_T \urcorner)$, d'une étape 15a) : $\vdash_T \neg \text{Prov}_T(\ulcorner G_T \urcorner)$ qui traduit justement le passage de S à T pour l'énoncé : $\neg \text{Prov}_T(\ulcorner G_T \urcorner)$. Cette contrainte est toujours vérifiée dans le cadre de notre interprétation du Programme (a fortiori de l'interprétation standard), par le fait que par hypothèse T étend S. Il apparaît ainsi que dans le cadre de cette interprétation, la prise en compte de la distinction entre S et T n'introduit aucune contrainte spécifique pour l'utilisation du second théorème, qui s'y applique de la même façon. Dans [Detlefsen 2001], Michael Detlefsen introduit quant à lui des conditions de dérivabilité très différentes, faisant intervenir une fonction $\text{Prov}_S(x)$ au-delà de $\text{Prov}_T(x)$. De manière plus précise, s'il conserve les conditions CD1' et CD2' (en suivant notre notation), il introduit une condition CD3'_D plus complexe que CD3' (faisant intervenir Prov_S) ainsi qu'une nouvelle condition CD4'_D qui introduit justement un lien entre Prov_S et Prov_T :

CD3'_D) $\vdash_S \text{Prov}_T(\ulcorner \varphi \urcorner) \rightarrow \text{Prov}_S(\ulcorner \text{Prov}_T(\ulcorner \varphi \urcorner) \urcorner)$;

$$CD4'_D \vdash_S \text{Prov}_S(\ulcorner \phi \urcorner) \rightarrow \text{Prov}_T(\ulcorner \phi \urcorner).$$

A la suite d'une longue discussion, il justifie pour l'essentiel l'introduction de la condition $CD3'_D$ par un argument épistémologique. Il considère qu'au sein d'un système formel unique, la condition $CD3$ est et doit être le reflet dans le système formel de la condition $CD1$. Le maintien de cette interprétation au cas où S et T sont distingués conduit naturellement à la condition $CD3'_D$ [Detlefsen 2001, p. 47-48]. Quant à la condition $CD4'_D$, son origine est double : d'une part elle est nécessaire à l'établissement du second théorème une fois la condition $CD3'$ remplacée par $CD3'_D$, d'autre part elle traduit pour l'auteur la condition que « S est une sous-théorie de T », sans avoir à prendre en compte l'extension de S par T , condition qu'il refuse [Detlefsen 2001, p. 47, n. 11]. Michael Detlefsen doit encore faire intervenir le lemme de diagonalisation sous une forme mêlant l'énoncé G_S et la fonction $\text{Prov}_T : \vdash_S G_S \leftrightarrow \neg \text{Prov}_T(\ulcorner G_S \urcorner)$ ¹⁶¹. Nanti de ces trois relations, il peut alors établir le second théorème en distinguant S et T , d'une manière différente de celle que nous avons exposée, sur la base du premier théorème appliqué à $S : \nvdash_S G_S$ et non à $T : \nvdash_T G_T$ [Detlefsen 2001, p. 48]. Voici alors le résumé des trois arguments détaillés qu'il avance pour justifier le caractère peu plausible de l'utilisation des conditions $CD3'_D$ et $CD4'_D$ et donc, selon lui, le caractère peu plausible de l'utilisation des conditions de dérivabilité en général lorsque S est distingué de T :

- 1) dans le cas où S n'est pas récursivement énumérable, tout en respectant une certaine forme raisonnable de finitisme, alors que T est récursivement énumérable on va au-devant de problèmes ;
- 2) on n'est pas obligé d'utiliser la condition $CD4'_D$ pour traduire de manière satisfaisante et intensionnelle la notion informelle de prouvabilité ;
- 3) si l'on prend en compte les conditions du finitisme, techniquement, $CD4'_D$ n'est pas vérifié en général.

Notre réponse à chacun de ces arguments est la suivante :

- 1) il s'agit d'une situation à la fois hypothétique et exotique, pour laquelle l'auteur n'apporte aucun élément en sa faveur ;
- 2) cela revient à demander au partisan de l'interprétation standard ou de notre interprétation de justifier la condition $CD4'_D$, alors que celle-ci a été introduite pour prendre en compte des

¹⁶¹ Il est bien entendu nécessaire de justifier un tel résultat, qui ne ressort naturellement ni du lemme de diagonalisation sur $S : \vdash_S G_S \leftrightarrow \neg \text{Prov}_S(\ulcorner G_S \urcorner)$, ni de celui sur $T : \vdash_T G_T \leftrightarrow \neg \text{Prov}_T(\ulcorner G_T \urcorner)$. Or si la dérivation de l'implication matérielle dans une direction s'établit facilement à partir de la condition $CD4'_D$, il n'en est pas de même de celle dans l'autre direction. Mais l'auteur ne donne aucune précision sur ce point, en se limitant à une justification de nature philosophique visant à montrer qu'il doit en être ainsi.

contraintes spécifiques de l'instrumentalisme de Detlefsen, qui ne s'appliquent pas à cette interprétation ;

3) l'auteur suggère lui-même la réponse [Detlefsen 2001, p. 68] ; l'argument ne s'applique plus lorsque T étend S, ce qui est le cas de l'interprétation standard et de la nôtre, mais ce qui est refusé par l'instrumentalisme de Detlefsen.

De toute cette explication, il ressort que les conditions de dérivabilité particulières introduites par Michael Detlefsen ne sont pas imposées par la seule condition de devoir distinguer S et T, mais par des contraintes supplémentaires introduites par l'auteur et liées à sa propre interprétation du Programme. Comme les arguments contre l'application du second théorème au Programme introduits dans [Detlefsen 2001] sont directement liés à ces conditions de dérivabilité particulières, ils sont simplement non pertinents dans le cadre de notre interprétation du Programme¹⁶².

Nous achevons notre discussion de l'impact du second théorème sur le Programme par la prise en compte du résultat mentionné pour la première fois par Gödel en 1967¹⁶³ et repris dans [Gödel 1972b, remarque 1]. Solomon Feferman a fourni une présentation éclairante de ce résultat [Feferman 1990, p. 284-287], mais son importance dans le cadre de la discussion du programme de Hilbert au sens restreint n'a été soulignée que récemment par Michael Potter¹⁶⁴ [Potter 2000, p. 258], puis a été reprise ensuite [Giaquinto 2002, p. 189-190], [Zach 2005a, p. 507-510] et [Smith 2007, p. 248]. Il s'agit d'une extension du second théorème qui conduit à la non dérivabilité de la « consistance externe » au sens de Gödel (au lieu de la consistance simple), mais surtout qui s'établit sur la base des seules exigences requises par le premier théorème, donc *sans les conditions de dérivabilité*¹⁶⁵ (en particulier la condition CD3). Les trois points remarquables sont alors que :

¹⁶² Sachant que ces arguments ne s'appliquent pas d'avantage dans le cadre de l'interprétation instrumentaliste de Michael Detlefsen, tout simplement parce que celui-ci y refuse justement les conditions de dérivabilité en général.

¹⁶³ Lors de la publication par Jean van Heijenoort de la traduction en Anglais de son article de 1932 : « Über Vollständigkeit und Widerspruchsfreiheit ».

¹⁶⁴ Au-delà du fait que sur un plan technique, la démonstration de ce résultat proposée par Michael Potter est incorrecte [Zach 2005a, p. 508].

¹⁶⁵ D'un point de vue technique il n'y a là aucune contradiction avec le second théorème, dans la mesure où l'on montre que la *dérivation* dans la théorie du fait que (le codage de) la consistance externe se déduit (du codage) de la consistance simple, n'est quant à elle possible qu'en introduisant les conditions de dérivation.

1) la consistance externe au sens de Gödel recouvre la vérifiabilité au sens de Bernays pour les énoncés Π_1 (§ 1.5) et correspond donc strictement à l'exigence finitiste du Programme¹⁶⁶, en étant alors un critère plus faible que la consistance simple ;

2) ce résultat peut être étendu à PRA¹⁶⁷ ;

3) en s'affranchissant des conditions de dérivabilité, il coupe court à la possibilité de sauver le Programme¹⁶⁸ en utilisant en particulier une solution « à la Rosser ».

C'est donc un argument extrêmement fort (d'une extension) du second théorème contre le Programme, qui n'a été remis en cause¹⁶⁹, à notre connaissance, par aucun commentateur du programme de Hilbert. Michael Detlefsen en particulier n'a pas publié de réponse à cet argument, mais il nous a transmis une telle réponse¹⁷⁰ à la suite d'une question de notre part. Son argument repose sur le résultat technique que nous venons de mentionner (note 164) : pour établir (dérivé) l'équivalence entre la consistance externe (au sens de Gödel) et la consistance simple, les conditions de dérivabilité (CD3 en particulier) sont nécessaires. Bien évidemment cette réponse ne tient que si la référence retenue pour le finitisme et pour le Programme est elle-même la consistance simple. Or dans le cadre de l'interprétation que nous avons retenue, la situation est inversée : la référence est la consistance externe (au sens de Gödel), qui est considérée comme l'expression finitiste de la consistance simple. Dans ce cadre, [Gödel 1972b, remarque 1] constitue bien une objection directe au Programme, sans nécessiter le passage par les conditions de dérivabilité (CD3 en particulier).

En conclusion, il ressort d'abord de cette longue discussion sur l'impact du second théorème de Gödel sur le programme de Hilbert au sens restreint, que la situation n'est pas aussi simple que ce que beaucoup de commentateurs en ont exprimé. En particulier les arguments présentés par Michael Detlefsen sont subtils, soulèvent des questions légitimes, et ne peuvent donc pas

¹⁶⁶ Par rapport aux exigences finitistes : 1) la démonstration est bien effectuée en faisant intervenir la vérité intuitive des instances des énoncés Π_1 , 2) telle que la démonstration est effectuée, l'on peut raisonnablement admettre que les énoncés Π_1 y sont compris comme des « jugements hypothétiques ».

¹⁶⁷ Gödel n'a donné de son résultat qu'une esquisse de preuve dans [Gödel 1972b, remarque 1]. Le principe de la démonstration est présenté dans [Feferman 1990], mais la démonstration complète, en particulier pour l'extension à PRA, se trouve dans [Jeroslow 1973].

¹⁶⁸ C'est explicitement à l'égard de telles perspectives que Gödel l'a introduit, sous le titre : « La meilleure et la plus générale version de l'improuvabilité de la consistance dans le même système ».

¹⁶⁹ Avec une réserve concernant Richard Zach. Celui-ci conteste en effet la pertinence finitiste de la solution formelle proposée dans [Feferman 1990] (en fait reprise de [Jeroslow 1973]) pour l'expression de la consistance externe sous la forme d'un schéma de réflexion. Il propose à la place une expression unique mais faisant intervenir (le codage d') une fonction de vérité [Zach 2005a, p. 509]. Cette réserve nous paraît doublement contestable. Tout d'abord la vérifiabilité de Bernays pour les énoncés Π_1 exige justement un schéma (et non une infinité d'instances) ; il est donc naturel de la traduire par un schéma. Ensuite et en sens inverse, il est totalement contraire à l'esprit du Programme de faire intervenir au sein du système formel un prédicat de vérité (même de portée limitée).

¹⁷⁰ Communication personnelle de Michael Detlefsen du 27/02/2012.

être balayés d'un revers de la main. Il apparaît toutefois à l'issue de notre analyse, que dans le cadre de notre interprétation du Programme, qui *n'est pas* celle de Michael Detlefsen, les objections soulevées par celui-ci ne sont pas recevables. Il en résulte qu'aucune échappatoire au fait que le second théorème s'oppose au Programme n'apparaît raisonnablement possible dans ce cadre, en retrouvant ainsi la réponse classique, mais sur la base d'une argumentation beaucoup plus précise. Ajouté au fait que nous avons déjà obtenu une réponse analogue pour le premier théorème, il en ressort que le seul espoir de sauver le programme de Hilbert au sens restreint consiste à « changer les règles du jeu », c'est-à-dire à revenir sur son contenu avec deux voies possibles (non exclusives) : étendre le finitisme, ou réduire le domaine des mathématiques concernées¹⁷¹. Nous allons analyser successivement ces deux voies dans les deux chapitres qui suivent.

¹⁷¹ Marcus Giaquinto exprime d'une manière différente une idée analogue [Giaquinto 1983, p. 129] : on ne peut pas conserver *à la fois* les deux parties du Programme qui sont, d'une part son socle finitaire unanimement accepté, et d'autre part la capacité de rendre compte de la totalité des mathématiques sur cette base en « gravissant les échelons ».

3 Peut-on sauver le programme de Hilbert au sens restreint en élargissant les critères du finitisme ?

Comme on le sait, la voie qui consiste à élargir le finitisme pour sauver le programme de Hilbert au sens restreint et à laquelle invite Gödel à la fin de l'article de 1931 [Gödel 1931 (1989), p. 141-142], a été évoquée par Hilbert lui-même (dans le seul commentaire qu'on lui connaisse sur le sujet), après l'irruption des théorèmes de Gödel de 1931 :

[...] je voudrais souligner que l'opinion, apparue durant un certain temps, selon laquelle certains résultats nouveaux de Gödel empêcheraient ma théorie de la preuve d'aboutir s'est montrée erronée. Car en fait, ce résultat montre tout juste que, pour les preuves futures de non-contradiction, le point de vue finitiste doit être exploité avec plus d'acuité que ne le requiert le traitement des formalismes élémentaires.

[Hilbert, préface à Hilbert, Bernays 1934 (2001), p. 43]

Une telle position amène naturellement deux questions, d'abord sur la circonscription du nouveau finitisme proposé, ensuite et surtout sur la justification épistémologique de celui-ci : que reste-il alors des caractéristiques du Programme initial ? Pour asseoir et faciliter la discussion qui va suivre, nous commençons par proposer une typologie des différentes formes possibles de finitisme et de constructivisme, sachant qu'il n'y a pas de consensus sur la terminologie. Nous nous appuyons pour cela sur la hiérarchie précise de l'ordinal limite des différents systèmes formels de l'arithmétique du 1^{er} ordre, que l'on peut associer naturellement à chacune des positions envisagées¹⁷².

Tableau III : Hiérarchie des formes possibles de finitisme et de constructivisme¹⁷³

Ordinal limite	$n_{\text{sup}} < \omega$	ω	ω^ω	ϵ_0	Γ_0 ¹⁷⁴
	borné	de Hilbert	large	Intuitionnisme de Brouwer	
	Finitismes				
	Constructivismes				
	Prédicativisme au sens de Feferman				

¹⁷² Nous avons déjà présenté dans le tableau I (§ 1.4), les principaux systèmes formels de l'arithmétique du 1^{er} ordre et les ordinaux limites correspondants.

¹⁷³ Une telle partition recouvre à la fois celle de Michael Potter [Potter 2000, p. 254] et la hiérarchie de Simpson [Simpson 2009, p. 43], en les élargissant à la situation du finitisme borné. Elle s'apparente également à la circonscription, plus vague, de Hao Wang [Wang 1954, p. 263] et [Wang 1958, p. 472]. La position selon laquelle l'ordinal limite ω^ω constitue la limite supérieure du finitisme est défendue par Gödel [Gödel 1938b (1995), p. 105], même s'il est ensuite revenu sur cette position [van Atten 2014, p. 208].

¹⁷⁴ Γ_0 est le premier ordinal imprédictif au sens de Feferman [Smorynski 1982, p. 185-186].

Il convient ensuite de définir les exigences minimales que devrait satisfaire toute extension raisonnable du finitisme de Hilbert. Dans la mesure où il s'agit d'essayer de sauver le programme de Hilbert au sens restreint, il convient évidemment que l'ensemble des *objectifs* définis par Hilbert pour celui-ci soient conservés. Comme nous l'avons vu, si Hilbert autorise une conception extrêmement libérale en mathématiques (le « paradis cantorien »), il n'en est pas de même en métamathématique, en particulier pour les démonstrations de consistance. Ces dernières sont en effet fortement contraintes par la double exigence (voir infra § 1.1) : 1) que la base intuitive en soit acceptée par tout mathématicien quelles que soient ses inclinations philosophiques, 2) que l'extension de la partie « idéale » (non intuitive) des mathématiques soit progressive, en prenant en compte des systèmes formels de plus en plus forts (selon l'image de « gravir les échelons »), ce qui, pour ces démonstrations, impose une démarche *effective* (au sens moderne) et interdit *de facto* l'utilisation d'une démarche *imprédictive* au sens de Poincaré¹⁷⁵.

La tentative la plus détaillée de définition d'un finitisme élargi par l'école de Hilbert se trouve sans aucun doute dans [Hilbert, Bernays 1939 (2001), § 5.3], en relation en particulier avec l'introduction par Hilbert de la « règle- ω » en 1930 que nous avons déjà évoquée (§ 1.4). Par ailleurs, de nombreuses autres tentatives ont été proposées sur une base finitiste élargie visant implicitement une sorte de « programme de Hilbert » modifié, en prenant acte des théorèmes de Gödel de 1931. Les plus connues et les plus anciennes sont au nombre de trois¹⁷⁶, à savoir, un résultat de Gödel lui-même : [Gödel 1933b], le théorème de Gentzen de 1936 et encore un résultat de Gödel : [Gödel 1958]¹⁷⁷. Une proposition beaucoup plus récente de Torkel Franzen : [Franzen 2004] semble répondre implicitement à l'objectif même du Programme en utilisant pour cela le second théorème de Gödel de 1931, ce qui est évidemment paradoxal. Une solution également récente, se réclamant explicitement d'un « programme de Hilbert » modifié, est l'instrumentalisme de Michael Detlefsen : [Detlefsen 1986] que nous avons déjà évoqué, mais que nous n'avons pas encore analysé en tant que tel. Enfin, les tenants de la logique

¹⁷⁵ Une définition *imprédictive* au sens de Poincaré est une définition qui mentionne ou présuppose une collection infinie comprenant elle-même l'entité définie. Une telle définition correspond alors à un cercle vicieux, sauf si la collection infinie peut être définie indépendamment par un autre moyen, ou bien si elle est considérée exister « en elle-même », par exemple dans le cadre d'une ontologie réaliste.

¹⁷⁶ Les trois voies correspondantes ont été comparées dès 1938 par Gödel (Conférence de Zilsel : [Gödel 1938b]) et surtout en 1941 (conférence de Yale : [Gödel 1941]), en anticipant alors le résultat de 1958.

¹⁷⁷ Le lecteur pourrait s'étonner de ne voir mentionner ni Georg Kreisel ni Solomon Feferman, deux auteurs dont les travaux ont porté sur des extensions du programme de Hilbert, le premier s'étant souvent exprimé sur le sujet au point d'y consacrer un long article ([Kreisel 1976]). En fait, nous avons déjà abordé la position de Georg Kreisel au § 2.1, quant à Solomon Feferman, au-delà de ce que nous en avons déjà dit au § 2.3, nous l'aborderons en analysant les réalisations partielles du Programme au § 4.

paraconsistante ont proposé de manière plus radicale d'accepter l'inconsistance. Nous allons donc étudier successivement ces différentes solutions, en nous attachant à en déterminer à la fois la portée, ainsi que la mesure dans laquelle elles constituent une extension raisonnable du finitisme de Hilbert selon les exigences que nous venons de définir.

3.1 Les résultats « positifs » obtenus postérieurement aux théorèmes de Gödel de 1931

Deux de ces résultats : [Gödel 1933b] et [Gödel 1958] faisant intervenir le système formel de l'arithmétique du 1^{er} ordre dans le cadre de la logique intuitionniste : HA (formalisée par Heyting), il est utile de commencer par une présentation succincte de la démarche intuitionniste. Comme cela est bien connu, celle-ci se caractérise, d'une manière générale, par le refus de dissocier l'existence d'un objet mathématique (ou la vérité d'un énoncé) de sa mise en évidence de manière constructive (en un sens qui reste à définir). La conséquence la plus emblématique en est le refus d'appliquer la loi du tiers exclu à des collections infinies. Ainsi, contrairement à la position classique, si l'hypothèse de non-existence¹⁷⁸ d'un objet donné parmi une collection infinie se traduit par une contradiction, il est refusé d'en inférer l'existence de cet objet tant qu'un moyen permettant de construire cet objet n'a pas été proposé¹⁷⁹. Selon son fondateur : Brouwer, cette démarche est associée à un engagement philosophique, selon lequel les mathématiques résultent fondamentalement d'une construction sans langage [Brouwer 1947 (1992), p. 417], qui tire son origine de la perception consciente du temps, le langage permettant seulement de communiquer ces constructions à d'autres personnes [Brouwer 1952 (1992), p. 449]. Dans ce cadre, une entité mathématique n'est pas figée, mais peut toujours acquérir une propriété qu'elle ne possédait pas auparavant [Brouwer 1955 (1992), p. 461]. Comme on peut s'en douter, une vision aussi radicale a été fortement critiquée comme l'illustre la remarque ironique de Raymond Wilder : « On peut avoir l'impression que l'intuitionniste a pris le *dictum* très souvent cité de Kronecker, selon lequel "Les entiers ont été créés par Dieu, mais tout ce qui suit est l'ouvrage de l'homme", et substitué "intuition" à "Dieu" » [Wilder 1950 (1986), p. 195]. D'une manière plus précise (ou plus sérieuse), le caractère subjectif de la notion de preuve qui

¹⁷⁸ Par contre, en conformité avec la position classique, si l'hypothèse *d'existence* d'un objet donné conduit à une contradiction, ce résultat est bien considéré comme la preuve que l'objet n'existe pas.

¹⁷⁹ C'est ainsi, par exemple, que l'énoncé : « La conjecture de Goldbach est vraie ou bien est fausse », trivialement vrai en logique classique (en quelque sens de vérité que ce soit), est intuitionnistiquement faux.

en découle a été relevé d'une manière particulièrement claire par Gödel : « Cette notion de preuve intuitionnistiquement correcte ou de preuve constructive, manque de la précision désirée. En fait, on peut dire qu'elle fournit elle-même un contre-exemple contre sa propre admissibilité, dans la mesure où il est douteux de savoir si une preuve utilisant cette notion de preuve constructive est constructive ou non » [Gödel 1941 (1995), p. 190-191]. Comme le souligne Marcus Giaquinto, c'est justement le fait que la notion de preuve intuitionniste ne soit pas constructivement définie, en ne portant pas seulement sur des configurations d'objets préalablement définis, qui explique que l'intuitionnisme aille au-delà du finitisme de Hilbert [Giaquinto 2002, p. 194-195]. Gödel a indiqué de manière très précise où se situait le franchissement du finitisme lorsque l'intuitionnisme est représenté par HA : il s'agit en l'occurrence des axiomes de HA concernant l'absurdité, qui font intervenir non seulement des nombres mais également des preuves, sachant qu'il n'y a pas de procédure finie pour générer toutes les preuves possibles [Gödel 1933c, p. 52-53].

En 1933, dans [Gödel 1933b], Gödel¹⁸⁰ établit que la non-contradiction de l'arithmétique intuitionniste entraîne la non-contradiction de la « théorie axiomatique des nombres » selon la terminologie de Bernays [Bernays 1935a, p. 69]. La consistance de la seconde s'en déduit alors directement si l'on accepte la consistance de la première. Pour l'arithmétique intuitionniste c'est le système formel HA qui est utilisé, plus exactement un système formel partiel de HA [Hilbert, Bernays 1939 (2001), p. 427]. Quant à ce que recouvre exactement l'expression de « théorie axiomatique des nombres », Anne Troelstra précise qu'il s'agit d'un système semi-formel basé sur celui de Herbrand de 1931, mais où ne sont utilisés de manière essentielle que les deux premiers groupes d'axiomes A et B, sachant que l'on peut admettre que la démonstration s'étend naturellement à PA [Troelstra 1986, p. 282-283]. La démonstration elle-même établit le caractère conservateur de cette théorie axiomatique des nombres par rapport à HA, pour les énoncés Π_1 (l'extension aux énoncés Π_2 ayant été établie ultérieurement par Georg Kreisel dans [Kreisel 1958a]). Elle présente sans ambiguïté un caractère constructif en un sens plus fort que HA lui-même. Le fait que le système HA aille bien au-delà des exigences finitistes de Hilbert est alors apparu clairement dès la publication de Gödel. Ce point a d'ailleurs été indiqué par Gödel de manière non-équivoque [Gödel 1933c, p. 52-53] et il est particulièrement souligné par Anne Troelstra [Troelstra 1990, p. 220]. En reposant sur HA qui enfreint les deux critères

¹⁸⁰ Une précision historique doit être apportée ici. Lorsqu'il fait état de ce résultat, Bernays mentionne les travaux de « Gentzen et de Gödel ». En fait, comme Bernays l'indique brièvement [Bernays 1935a, p. 69] et comme cela est surtout précisé par Anne Troelstra [Troelstra 1986, p. 282, 284], le même résultat a été obtenu indépendamment par Gödel et par Gentzen en 1932 mais n'a été publié que par Gödel en 1933 (Gentzen ayant alors retiré son propre article).

que nous avons introduit pour caractériser le finitisme, le résultat de 1933 ne répond clairement pas aux exigences d'une extension raisonnable du finitisme. Pour clore cette présentation, mentionnons un point d'histoire. Il est classiquement considéré que c'est à la suite de la publication de [Gödel 1933b] que l'école de Hilbert a cessé d'assimiler finitisme et intuitionnisme [Troelstra 1986, p. 284]. Une publication récente de William Tait : [Tait 2006b], fait état d'une situation plus complexe, en évoquant, pour l'école de Hilbert dans les années 1920, la référence à deux formes distinctes d'intuitionnisme : 1) la première en un sens large, qui recouvre l'intuitionnisme de Brouwer (que l'on retrouve dans HA), 2) la seconde en un sens plus restrictif, qui serait dû à Weyl, et qui serait alors très proche du finitisme de Hilbert [Tait 2006b, p. 84-85]. Pour William Tait, l'ambiguïté provient du fait qu'à cette époque, von Neumann et Herbrand utilisaient le même terme d'intuitionnisme pour désigner (selon les circonstances) l'une ou l'autre de ces deux positions sensiblement différentes.

La première démonstration directe de la consistance de PA date de 1936 et émane de Gerhard Gentzen : [Gentzen 1936a], sachant qu'une version plus simple en a été publiée par le même auteur deux ans plus tard : [Gentzen 1938b]. Elle a ensuite été complétée et enrichie ultérieurement par différents auteurs. Bien entendu, en conformité avec le second théorème de Gödel de 1931, cette démonstration utilise des moyens qui vont au-delà de PA¹⁸¹. La question fondamentale qui se pose à l'égard de la démonstration de Gentzen est de savoir dans quelle mesure ces moyens peuvent être considérés comme une extension acceptable du finitisme initialement envisagé par Hilbert, et plus généralement de définir la portée épistémologique du résultat obtenu. Ce dernier point est à la fois controversé et difficile¹⁸², comme l'ont fait remarquer les auteurs majeurs qui l'ont évoqué [Kleene 1967 (1971), p. 265] et [Feferman 1988b, p. 366]. Les informations très précises, données principalement par Gentzen lui-même, permettent d'abord de mesurer l'ampleur de l'écart de la démonstration par rapport au finitisme de Hilbert : 1) la démarche n'est pas effective (au sens moderne), dans la mesure où le nombre total d'étapes n'est pas spécifié à l'avance [Gentzen 1936a (1969), p. 196], 2) il est fait usage de manière essentielle d'un énoncé Σ_1 , sans que l'on ne soit en mesure de montrer qu'il y a bien au moins une instance qui le vérifie, en contradiction avec le critère proposé par Bernays pour qu'un tel type d'énoncé puisse être considéré comme finitiste (en anticipant la discussion à

¹⁸¹ C'est ainsi que la théorie minimale permettant de démontrer la consistance de PA (ERA à laquelle on ajoute le moyen de traiter l'induction transfinie pour les formules sans quantificateur) est plus faible que PA sous certains aspects, tout en étant plus forte sous d'autres.

¹⁸² Le commentaire de Hilbert déjà cité [préface à Hilbert, Bernays 1934 (2001), p. 43], est manifestement insuffisant pour en déduire que celui-ci acceptait l'extension du finitisme associée au résultat de Gentzen, contrairement à ce qu'affirme par exemple Imre Lakatos [Lakatos 1967 (1986), p. 36-37].

venir du § 3.2) [Gentzen 1936a (1969), p. 201], 3) la démonstration nécessite de faire intervenir de manière essentielle une séquence infinie de nombre supposée donnée, elle est donc clairement imprédicative au sens de Poincaré [Gentzen 1936b (1969), p. 230] et [Gödel 1938b (1995), p. 107]. Voici alors les principaux éléments que Gentzen a proposé pour justifier le caractère « finitiste » (en un sens évidemment élargi) de sa démonstration :

Somme toute, je suis porté à croire qu'en termes de la distinction fondamentale entre les méthodes de preuve discutables et indiscutables, *la preuve du caractère finitaire* de la procédure de réduction peut encore être considérée comme indiscutable, [...]. [Gentzen 1936a (1969), p. 197]

Je ne vois pas, cependant, à partir de quel « point » ce qui est constructivement indiscutable est supposé finir, et où une extension ultérieure de l'induction transfinitie est alors considérée comme devenant discutable. Je pense plutôt que la confiance dans les nombres transfinis requise pour la preuve de consistance, est comparable à celle des premiers segments ordinaux, disons jusqu'à ω^2 , de la même manière que la confiance en un calcul numérique étendu à une centaine de pages, par rapport au calcul de quelques lignes : c'est seulement une entreprise considérablement *plus vaste* d'entreprendre de convaincre quelqu'un de sa certitude du début à la fin. [Gentzen 1938b (1969), p. 286]

Le premier élément n'est évidemment pas autre chose qu'un acte de foi, qui ne saurait convaincre un sceptique. Quant à l'argument défendu dans le second, Gödel y a répondu ultérieurement de manière explicite, en précisant qu'il y a en fait un saut qualitatif entre ω^2 et ε_0 [Gödel 1958 (1992), p. 503] et [Gödel 1972a (1990), p. 273]. Il apparaît ainsi qu'en s'opposant au second critère que nous avons introduit, la démonstration de Gentzen ne peut pas être considérée comme finitiste en un sens raisonnablement étendu de ce terme. Au-delà de Gentzen lui-même, le principal auteur qui ait défendu que la démonstration de 1936 constitue une extension naturelle du finitisme initial de Hilbert est Jean Ladrière : [Ladrière 1951] et [Ladrière 1957]. La justification que celui-ci propose repose cependant sur une assimilation des démarches constructive et finitiste, en considérant d'une part que la prise en compte d'une collection infinie d'objets correspond à un procédé « non constructif » [Ladrière 1957, p. 213], et d'autre part que l'aptitude à pouvoir toujours aller au-delà d'une suite donnée correspond au point de vue « constructif » [Ladrière 1957, p. 219]. L'auteur argumente alors que le théorème de Gentzen de 1936 respecte ainsi l'idéal constructif dans le sens défini, tout en dépassant l'idéal hilbertien qui est limité à des considérations finies [Ladrière 1957, p. 223]. Mais une telle considération du point de vue « constructif », si elle est bien cohérente avec la distinction classique et reprise par Hilbert entre infini en acte et infini potentiel, ne rend pas compte d'un caractère essentiel du finitisme hilbertien selon lequel le socle accepté a priori doit être

intuitivement acceptable. Elle néglige en particulier la limitation forte introduite par Hilbert dans la prise en compte de l'infini potentiel, implicite dans l'utilisation des énoncés Π_1 , selon laquelle de tels énoncés ne doivent être compris que sous la forme de jugements *hypothétiques*, ce qui les ramène *in fine* à des énoncés finis [Hilbert, Bernays 1934 (2001), p. 87]. Par ailleurs, le fait que la manipulation des ordinaux par Gentzen dans son théorème soit constructive (en un certain sens), ne change pas le fait que le seul usage des transfinis revient implicitement à prendre en compte l'infini en acte, ce qui est contradictoire avec le sens « constructif » défendu par Jean Ladrière. L'auteur est conscient du problème posé et y répond par l'appel à une sorte de dialectique entre l'intuitif et le formel, qui constitue la signature de sa propre interprétation plus générale des théorèmes de Gödel de 1931 (sur laquelle nous reviendrons au § 7.4) [Ladrière 1957, p. 223]. Le recours à une telle position se traduit surtout par un engagement philosophique fort, ce qui s'oppose cette fois au premier critère que nous avons introduit. Nous contestons donc à Jean Ladrière que ses arguments permettent de considérer le théorème de Gentzen de 1936 comme une extension raisonnable du finitisme de Hilbert¹⁸³.

Il est généralement admis que la démonstration de Gentzen présente cependant un caractère « constructif » [Zach 2007, p. 435-436], mais il reste à préciser en quel sens, avec la difficulté de l'absence de consensus sur la définition de cette notion [Fefermann 1988b, p. 366] et [Boniface 2004, p. 257-258]. La difficulté porte bien entendu sur la justification de l'induction transfinie, même si celle-ci n'y est utilisée que sur des énoncés sans quantificateurs¹⁸⁴. Alors que Rosser dénonce l'absence d'une preuve *constructive* [Rosser 1939b, p. 812], Gödel considère que celle-ci possède « un haut degré d'intuitivité » [Gödel 1938b (1995), p. 107]. Jean-Yves Girard fait cependant remarquer de manière pertinente que l'induction jusqu'à ω sur un énoncé quelconque (caractéristique de l'arithmétique) est certainement plus proche de l'intuition que l'induction (même limitée à des énoncés sans quantificateurs) jusqu'à ε_0 (dans le cas de PA) [Girard 1983, p. 45]. Sur un plan technique, deux éléments de la démonstration peuvent être évoqués en faveur d'une certaine forme de constructivité : 1) Gentzen donne un procédé constructif pour « éliminer les coupures », 2) les ordinaux utilisés sont constructibles (par récurrence) à partir des entiers naturels, sans que pour autant, et comme nous l'avons déjà indiqué, la démonstration ne soit ni effective (au sens moderne) ni prédicative (au sens de Poincaré). C'est donc seulement en ce sens faible qu'elle peut être considérée comme

¹⁸³ Cela nous conduit au passage à souligner qu'indépendamment de la question de son caractère finitiste, la démonstration de Gentzen de 1936 ne répond certainement pas à l'objectif du programme de Hilbert au sens restreint. En effet, elle conduit à établir la consistance de PA mais non celle de Z_2 et aucune de ses extensions n'a permis à ce jour de dépasser significativement PA.

¹⁸⁴ Techniquement, jusqu'à l'ordinal limite : ε_0 dans le cas de PA.

constructive, en constituant alors un terme intermédiaire entre le point de vue finitiste et l'intuitionnisme, comme le souligne Bernays [Bernays 1941, p. 147-148].

Le résultat obtenu par Gentzen permet également de répondre à une question restée en suspens concernant le Programme de Hilbert selon l'interprétation standard, à savoir la capacité d'établir (de manière finitiste) la consistance de PRA lui-même, puisque nous avons vu que celui-ci est sujet au second théorème de Gödel de 1931. Du point de vue historique, les affirmations de Bernays à ce sujet sont particulièrement ambiguës, en laissant entendre l'obtention d'une démonstration finitiste de PRA¹⁸⁵ [Hilbert, Bernays 1934 (2001), p. 406-407] et ce même, de manière plus surprenante, après 1936 [Hilbert, Bernays 1939 (2001), p. 100]. Sur ce point, [Sieg, Ravaglia 2005] fournit une explication claire de la situation : en fait, les preuves de consistance « finitaires » de PRA données dans [Hilbert, Bernays 1934] et [Hilbert, Bernays 1939] font appel à des moyens appartenant au système formel Z_μ , qui va lui-même au-delà de PA [Sieg, Ravaglia 2005, p. 995]. La justification proposée pour expliquer une telle attitude est alors que pour l'école de Hilbert dans les années 1920, PRA était considéré comme ayant une justification finitiste directe et n'avait donc pas besoin de démonstration de consistance ; les démonstrations de consistance évoquées ne doivent alors être considérées que comme des « galops d'essai » de démonstrations de consistances nécessaires pour les systèmes formels « idéaux » [Sieg, Ravaglia 2005, p. 997]. En 1936, après la parution du résultat de Gentzen, il est apparu clairement qu'une démonstration de la consistance de PRA dans PRA était impossible, en nécessitant de faire intervenir des ordinaux inférieurs dont la limite est ω^ω (Jean-Yves Girard est particulièrement éclairant sur l'application du théorème de Gentzen à PRA [Girard 1983, p. 38-39]). C'est également à ce moment qu'il a été réalisé que le second théorème de Gödel de 1931 pouvait être généralisé à PRA¹⁸⁶. Une explication à l'incohérence apparente entre l'affirmation de l'existence d'une démonstration finitiste et ces deux résultats était alors envisageable, en supposant que si une démonstration de *consistance* de PRA nécessitait d'aller au-delà de PRA, une démonstration de l'exigence finitiste plus faible de *vérifiabilité des énoncés dérivables de PRA* était peut-être possible au sein de PRA. Comme nous l'avons vu à la fin du § 2, [Gödel 1972b (1990), remarque 1] et [Jerolow 1973] sont venu éliminer cette possibilité. Il n'y a donc pas d'échappatoire au fait qu'une démonstration finitiste

¹⁸⁵ En 1934 il apparaît également une ambiguïté pour Gödel lui-même sur le contenu du système formel A, pour lequel il affirme, lors de la conférence de Cambridge : [Gödel 1933c (1995)], que c'est sur lui que Hilbert entendait fonder le finitisme : pour William Tait et Solomon Feferman il s'agit de PRA [Tait 2006b, p. 103-104] et [Feferman 2011, p. 117-118], tandis que pour Wilfried Sieg il s'agit d'un système formel qui va au-delà [Sieg 2003a, p. 9, n. 1]. Cette ambiguïté disparaît en 1938, lors de la conférence de Zilsel : [Gödel 1938b (1995)], en faveur de PRA.

¹⁸⁶ Comme cela est mentionné explicitement par Bernays [Hilbert, Bernays 1939 (2001), p. 389].

de non-contradiction de PRA nécessite des moyens qui vont au-delà de PRA ; il n'est donc pas possible d'assurer la condition de possibilité du Programme, dans le cadre de l'interprétation standard de la thèse de Tait. Une proposition à ce sujet a cependant été suggérée à la fois par Michael Potter [Potter 2000, p. 251-252] et Marcus Giaquinto [Giaquinto 2002, p. 192-193] sur la base de deux arguments sensiblement différents. Il s'agit de l'affirmation que l'application du théorème de Gentzen de 1936 à PRA (mais non à PA) est compatible avec le prolongement le plus étendu raisonnablement envisageable du finitisme de Hilbert. Michael Potter s'appuie pour cela sur le résultat technique que tout ordinal de ce type peut s'exprimer sous la forme d'une suite *finie* telle que : $\alpha = \omega^{n_1} + \omega^{n_2} + \dots + \omega^{n_k}$. On peut cependant lui répliquer qu'utiliser une opération même élémentaire (l'addition finie) sur des transfinis, fait tout de même implicitement intervenir l'infini en acte. Marcus Giaquinto propose un raisonnement différent. Il commence par remarquer que le finitisme originel d'Hilbert repose sur une suite finie non bornée (par exemple représentable sur un ruban), dont la limite supérieure est : ω . On peut alors imaginer non plus un ruban mais un tableau¹⁸⁷ avec deux suites analogues dans deux dimensions, dont la limite supérieure devient : ω^2 . Par extrapolation sur le nombre de dimensions on peut envisager un nombre quelconque de celui-ci : m , éventuellement non borné mais fini, dont la limite supérieure est : ω^m . Un tel nombre correspond justement à l'ordinal en jeu dans la démonstration, dont la limite supérieure est alors : ω^ω . On aboutit ainsi, via le théorème de Gentzen de 1936, à une démonstration « finitiste » en un certain sens¹⁸⁸ de la consistance de PRA, qui vient conforter notre interprétation du programme de Hilbert, sur la base d'une extension du finitisme initial qui semble raisonnable¹⁸⁹. Bien entendu l'argument interdit de dépasser ω^ω qui reste très en deçà de ε_0 , et ne s'applique donc pas à PA. Aussi, il ne

¹⁸⁷ Au § 1.2, nous avons utilisé un raisonnement analogue pour justifier que toute FRP peut être considérée comme une fonction réelle.

¹⁸⁸ Qui rejoint, d'une part la notion de « presque finitiste » d'Aleksandar Ignjatovic [Ignjatovic 1994, p. 331] et d'autre part le « finitisme large » de Michael Potter [Potter 2000, p. 254], que celui-ci oppose au « finitisme étroit » (celui de Hilbert) dont l'ordinal requis est seulement ω (voir le tableau III, § 3).

¹⁸⁹ Cette extension respecte-t-elle les deux critères que nous avons introduit à cet effet ? L'argument de [Giaquinto 2002] que nous venons d'évoquer pour justifier l'extension de l'ordinal limite de : ω à ω^ω , associé à la remarque de Jean-Yves Girard selon laquelle le seul point de la démonstration qui aille au-delà de PRA concerne la partie purement équationnelle (sans logique) de PRA, pour laquelle nous n'avons pas de doute sérieux [Girard 1983, p. 38-39], permet de répondre au premier critère. Une fois que l'on a admis l'extension de l'ordinal limite de : ω à ω^ω comme base intuitive, en remarquant que la réduction de la démonstration de consistance à la seule part équationnelle de PRA est elle-même effectuée de manière finitiste au sens de Hilbert [Girard 1983, p. 38-39], on peut admettre que la démonstration ne fait pas intervenir l'imprédictivité au sens de Poincaré et que le nouveau système formel (PRA augmenté de la nouvelle base intuitive) est bien effectif ; cela permet de répondre au second critère.

permet évidemment pas de répondre au programme de Hilbert au sens restreint, mais il constitue une proposition intéressante de la limite supérieure d'une extension raisonnable du finitisme¹⁹⁰.

En 1958 Gödel a proposé une troisième solution, explicitement comme alternative, d'une part à son résultat de 1933 et d'autre part à celui de Gentzen de 1936, avec l'objectif que les bases en soient plus intuitives (plus proches du finitisme de Hilbert) [Gödel 1958 (1992), p. 504]. Ce résultat (en allemand), également connu sous le nom d'interprétation *Dialectica*, sera ensuite revu lors de la publication de sa traduction en anglais en 1972, sachant que la matière en avait déjà été présentée dès 1941, lors de la conférence de Yale : [Gödel 1941]. Gödel s'appuie sur un certain système formel qu'il nomme T, qui est une extension de PRA (Feferman le nomme d'ailleurs PR^ω) : ses axiomes sont les mêmes que ceux de PRA, les variables sur lesquelles s'applique l'induction sont inchangées, tandis que les autres variables, ainsi que les constantes, peuvent avoir un type fini arbitraire sur les nombres naturels. Les théorèmes de la logique propositionnelle intuitionniste doivent être par ailleurs des théorèmes de T¹⁹¹. La clef de la démonstration est l'utilisation de fonctionnelles calculables de type fini sur les entiers naturels, ce qui implique une imprédictivité au sens de Poincaré, dans la mesure où les définitions des fonctions de type le plus bas impliquent essentiellement des fonctions de type plus élevé. Une autre limitation au caractère constructif et intuitif de la démarche est la suivante : pour justifier sa règle (\rightarrow^D)¹⁹², il convient de recourir à la généralisation du principe de Markov, qui fait intervenir des instances de propositions de Markov dont le caractère constructif est sujet à discussion [Troelstra 1990, p. 232] et [Chiffi 2012, p. 76]. Sur un plan philosophique, le moins que l'on puisse dire est que l'affirmation de Gödel selon laquelle la notion de fonctionnelle calculable de type fini est mieux définie et plus intuitive que la notion de preuve dans HA, n'a pas emporté l'adhésion de l'ensemble des auteurs majeurs qui se sont exprimés sur la question. Si Solomon Feferman et Richard Zach se rangent à l'avis de Gödel [Feferman 1993c, p. 215] et [Zach 2007, p. 437], Wilfried Sieg et William Tait sont d'un avis contraire [Sieg 1994, p. 99] et [Tait 2006b, p. 107-108]. La situation est particulièrement bien résumée par Anne Troelstra : « Une conséquence est qu'il reste objet de débat de savoir si une réduction réelle a été obtenue, dépendant de la vision de chacun sur ce qui est intuitivement évident » [Troelstra 1977, p. 1036]. On peut d'ailleurs noter que selon Solomon Feferman, Gödel lui-même n'était pas satisfait de sa justification de l'utilisation des fonctionnelles

¹⁹⁰ On peut noter au passage qu'une telle limite reste inférieure à $\omega^{\omega^{\omega}}$ utilisé dans la démonstration d'Ackermann de 1924 (se reporter supra au § 4.1).

¹⁹¹ Il s'agit en effet d'une démonstration de consistance de HA, qui s'applique à fortiori à PA via [Gödel 1933b].

¹⁹² Avec les notations d'Anne Troelstra dans [Troelstra 1990], qui introduit l'article de Gödel et en donne une description technique beaucoup plus détaillée que l'article lui-même.

calculables de type fini [Feferman 1993c, p. 215]. Cela se traduit d'ailleurs par un flottement du vocabulaire de Gödel, relevé par Mark van Atten, à propos de la définition de « la fonction calculable de type tx » : jugée « démontrable intuitionnistiquement » en 1958, « reconnaissable constructivement » en 1970 (dans une révision envisagée mais non publiée), et « constructivement évident » en 1972 [van Atten 2014, p. 197-198]. En tout état de cause, la limitation majeure que nous avons mentionnée : le caractère imprédicatif au sens de Poincaré des fonctionnelles calculables de type fini, s'oppose radicalement au second critère que nous avons introduit. Ajouté au constat empirique que nous venons d'évoquer d'un non-respect du premier critère, l'interprétation de 1958 ne peut certainement pas être considérée comme une extension raisonnable du finitisme de Hilbert.

3.2 L'introduction de la règle- ω et la position de l'école de Hilbert sur l'extension du finitisme

C'est à la fin de sa vie, dans ses deux derniers articles publiés, que Hilbert a introduit la « règle- ω » permettant de dépasser les limites imposées par les théorèmes de Gödel de 1931, en relation avec une extension du finitisme originel de son programme. Cette introduction a cependant été proposée sous deux formes sensiblement différentes¹⁹³, correspondant à deux extensions différentes du finitisme :

1) la première a été exposée par Hilbert lors de la conférence de Hambourg de décembre 1930 ([Hilbert 1931]), elle a été reprise dans [Herbrand 1931b] et est évoquée plus qualitativement en 1939 [Hilbert, Bernays 1939 (2001), p. 414 - 415] ;

2) la seconde a été exposée par Hilbert lors de la conférence de Göttingen de juillet 1931¹⁹⁴ et est évoquée plus qualitativement en 1939 [Hilbert, Bernays 1939 (2001), p. 415]¹⁹⁵.

¹⁹³ Il convient de souligner que le même nom de « règle- ω » recouvre en fait différentes notions apparentées mais différentes, au-delà des deux propositions de Hilbert. D'après Roman Murawski, la règle- ω aurait été évoquée publiquement pour la première fois par Ajdukiewicz en 1928 sous le nom de « règle de Tarski » [Murawski 1999, p. 288]. Selon Solomon Feferman, elle aurait été effectivement proposée par Tarski en 1933, reprise par Carnap en 1934 et 1935, puis par Rosser en 1937 (qui la nomme « règle de Carnap »), et enfin par lui-même en 1962 [Feferman 1986, p. 213].

¹⁹⁴ « Beweis des *tertium non datur* », Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalische Klasse, p. 120-125.

¹⁹⁵ Selon Wilfried Sieg, la première vise à assurer la complétude syntaxique du système formel de l'arithmétique (du 1^{er} ou du 2^{ème} ordre), dont la consistance est alors (erronément) supposée assurée par les travaux antérieurs d'Ackermann et von Neumann (nous reviendrons sur ce point au § 4.1) [Sieg 2013c, p. 171]. Il existe une controverse historique sur la question de savoir si cette première version de la règle- ω a été introduite ou non par Hilbert en réponse au premier théorème de Gödel de 1931 (l'introduction de cette règle étant postérieure à la

Voici tout d'abord la formulation initiale de Hilbert, à la fois informelle et schématique :

[...] cette nouvelle règle d'inférence finitiste : lorsqu'on a déjà démontré que la formule $A(z)$ devient une formule numérique vraie chaque fois que z est un chiffre donné, on peut prendre pour formule initiale la formule $\forall x A(x)$. On rappelle que la proposition $\forall x A(x)$ a une portée bien plus grande que la formule $A(z)$ où z est un chiffre donné quelconque. Car dans le premier cas on peut *remplacer* le x de $A(x)$ non seulement par un chiffre, mais par n'importe quelle expression constructible dans notre formalisme et qui a le caractère d'un chiffre ; [...]. Or notre procédé transforme aussi les formules que notre nouvelle règle nous conduit à écrire, en des formules numériques. Notamment $\forall x A(x)$ se transforme en une formule $A(z)$ où z est un chiffre déterminé.

[Hilbert 1931 (1992), p. 193-194 ; en italique, traduction revue par nous]

Voici maintenant la formulation par Herbrand de la même règle, sans doute un peu plus claire, mais toujours informelle :

Introduisons enfin, d'après Hilbert [1931] un quatrième groupe d'hypothèses, qui n'appartient pas à l'arithmétique classique. *Ax étant une proposition sans variables apparentes* [dont les instances sont des énoncés de type Δ_0], *si, quand on la considère intuitionistiquement* [de manière finitiste], *on peut démontrer par des procédés intuitionnistes* [finitistes] *qu'elle est vraie pour tout x, nous ajoutons (x)Ax aux hypothèses.* [Herbrand 1931b (1968), p. 227 ; les commentaires entre crochets sont nôtres]

Il apparaît ainsi, sans ambiguïté à la lecture de la formulation de Herbrand, que dans cette version la règle ne porte que sur des énoncés Π_1 , tandis que dans la seconde version (de juillet 1931) elle porte sur tous les énoncés arithmétiques, à l'instar de son utilisation par Tarski en 1933. Mais dans les deux cas, Hilbert revendique explicitement un caractère finitiste (contrairement à Tarski), sans donner pour autant d'argument justificatif en ce sens. Voici ce qu'il précise pour la *seconde* formulation :

Finalement, le fait important, et pour notre interrogation, décisif, doit être mis en évidence, à savoir, que tous les axiomes et schèmes d'inférence VI, que j'ai appelés « transfinis », ont un caractère *strictement fini* : les propositions qu'ils contiennent peuvent être exécutées dans le fini.

[Hilbert, juillet 1931, cité dans Sieg 2013c, p. 191 et souligné par nous]

conférence de Königsberg du 7 septembre 1930). Après une analyse détaillée, Solomon Feferman conclut qu'aucune réponse tranchée ne peut être apportée à cette question [Feferman 1986, p. 210]. Toujours selon Wilfried Sieg, la seconde vise à assurer de manière finitiste (élargie) la consistance de l'arithmétique du 1^{er} ordre, malgré le (ou au-delà du) second théorème de Gödel de 1931 (dont la connaissance par l'Ecole de Hilbert en juillet 1931 est par contre indubitable) [Sieg 2013c, p. 171].

On se retrouve alors dans une situation analogue à celle du finitisme initial du programme de Hilbert au sens restreint : pour pouvoir discuter du caractère finitiste associé à la règle- ω , il faut d'abord donner une interprétation formelle de la présentation informelle proposée de cette règle. La présentation originale de Hilbert est en effet associée à un système *semi-formel*, comme cela a été généralement souligné [Feferman 1986, p. 210-211], [Tait 2002, p. 418] et [Sieg 2013c, p. 172].

Le point commun à toutes les interprétations qui ont été proposées des deux versions de la règle- ω introduites par Hilbert, respectivement en 1930 et 1931, est que les prémisses sont de nature syntaxique, associées à un système formel donné, tandis que le conséquent est de nature sémantique, pouvant être formel (associé à un modèle) ou informel. Comme cela a été également généralement souligné [Webb 1980, p. 155], [Ignatovic 1994, p. 325], [Potter 2000, p. 247-248] et [Tait 2010, p. 102], il faut alors supposer que l'on dispose d'un *processus* permettant d'établir que si l'on peut montrer : $\vdash \varphi_0(x)$ pour tout x donné, alors on est en droit de considérer que : $\forall x \varphi_0(x)$ est vrai (en un sens évidemment à préciser). A la suite de Solomon Feferman, un consensus s'est dégagé pour considérer que le système formel initial auquel devaient être attachées les prémisses de la règle- ω est PA (Z)¹⁹⁶, ce système devenant Z^* une fois muni de la règle- ω [Feferman 1986, p. 210-211]. Une manière simple de balayer les différents cas simples possibles, est alors de prendre en compte les différents niveaux métamathématiques de notre schéma C du § 1.5. Cela permet de retrouver naturellement les principales interprétations qui ont été proposées. Pour être exhaustif, il convient toutefois d'ajouter la solution particulière et subtile proposée par Daniel Isaacson dans [Isaacson 1992], qui fait intervenir le codage gödelien des prémisses de la règle- ω en exploitant pour cela les résultats de [Rosser 1937] et [Feferman 1962]. Cela le conduit alors à une solution de force intermédiaire par rapport aux solutions précédentes. Dans tous les cas, il convient de souligner l'apparition de deux contraintes majeures lorsque la règle- ω est interprétée formellement, qui en limitent considérablement la portée. La première a été généralement relevée : en utilisant une infinité de prémisses constituant un ensemble non récursif, il résulte que le système formel associé n'est plus effectif (ce qui ne pouvait évidemment être perçu en 1930 et 1931 : se reporter

¹⁹⁶ La seule exception étant Michael Potter qui propose un système formel plus faible, qu'il nomme EA, avec une ambiguïté, que nous avons déjà signalée, pour savoir ce que recouvre exactement celui-là [Potter 2000, p. 246-247].

supra au § 5.3) [Webb 1980, p. 153], [Dubucs 1986, p. 305] et [Potter 2000, p. 247-248]¹⁹⁷. La seconde, plus subtile, est plus rarement soulignée : le système Z^* reste syntaxiquement incomplet (nous y reviendrons en abordant la solution proposée par Daniel Isaacson), comme cela a été montré par Rosser dans [Rosser 1937] et anticipé par Gödel dans sa lettre à Bernays du 2 avril 1931 [Feferman 2003, p. 44]. Si l'on prend un peu de recul, la traduction formelle la plus simple de la règle- ω admet théoriquement quatre interprétations possibles, qui se ramènent en pratique à deux dans le cadre du programme de Hilbert au sens restreint. En effet, les prémisses peuvent être théoriquement, soit de nature syntaxique ($\vdash \varphi_0(1), \vdash \varphi_0(2), \vdash \varphi_0(3), \dots$), soit de nature sémantique par rapport au modèle standard ($\varphi_0(1)$ est vrai dans \mathcal{M}_0 , $\varphi_0(2)$ est vrai dans \mathcal{M}_0 , $\varphi_0(3)$ est vrai dans \mathcal{M}_0 , \dots), et de même pour la conclusion ($\vdash \forall x \varphi_0(x)$) ou bien ($\forall x \varphi_0(x)$ est vrai dans \mathcal{M}_0). En appliquant les critères les plus élémentaires du finitisme de Hilbert on montre facilement que les deux ensembles de prémisses sont équivalents (voir le schéma B du § 1.5), ce qui permet de considérer que les prémisses sont de nature syntaxique, en cohérence avec le consensus dont nous avons parlé. Il ne reste alors plus que deux solutions, fixées par la nature de la conclusion, qui donnent respectivement le sens « sémantique » ($\forall x \varphi_0(x)$ est vrai dans \mathcal{M}_0) et le sens « syntaxique » ($\vdash \forall x \varphi_0(x)$). Ces deux sens se traduisent par des relations qui n'ont pas du tout la même force : très faible (et même triviale) pour le sens sémantique et très forte pour le sens syntaxique (qui se ramène à la 1-complétude). Pour aller plus loin et établir graduellement les interprétations possibles de la règle- ω lorsque celle-ci est interprétée formellement, il est pertinent de gravir progressivement les cinq niveaux représentés sur le schéma C du § 1.5, tout en précisant dans chaque cas les contraintes induites, en particulier sur la nature du finitisme impliqué.

L'interprétation minimale de la règle- ω correspond au niveau 1 du schéma C. Ne sont alors pris en compte que les énoncés Π_1 (soit la formulation de Hilbert de 1930). Le finitisme est fixé par le critère de vérifiabilité de Bernays. Il est tout de même nécessaire de faire l'hypothèse de la consistance simple du système formel de départ (que nous avons appelé S), sachant que l'établissement finitiste de cette consistance est l'une des conditions de possibilité du programme de Hilbert au sens restreint (§ 1.1). Dans ce cas, l'interprétation « sémantique » de la règle- ω est obtenue d'emblée : si $\vdash \varphi_0(1), \vdash \varphi_0(2), \vdash \varphi_0(3), \dots$, alors $\forall x \varphi_0(x)$ est vrai dans

¹⁹⁷ Ce qui conduit d'ailleurs Michael Potter à proposer de ne pas interpréter formellement du tout la règle- ω pour éviter ce travers. Mais bien entendu, l'on n'aboutit alors à une solution entièrement informelle encore plus éloignée du programme de Hilbert originel.

m_0 . Si une telle interprétation n'introduit ainsi aucune contrainte sur le finitisme standard (en remplaçant évidemment la vérité dans m_0 par la vérité intuitive), elle n'apporte rien de plus et ne permet en particulier pas de dépasser les limitations introduites par le premier théorème de Gödel de 1931.

Bien qu'elle ne puisse pas être directement rattachée à une interprétation particulière de la règle- ω , l'étape supérieure consiste à imposer la 1-consistance (niveau 2 du schéma C). Elle présente l'intérêt indéniable d'éviter le cas pathologique d'un système formel consistant mais 1-inconsistant, qui s'oppose clairement à l'esprit du Programme¹⁹⁸. Ne sont toujours pris en compte que les énoncés Π_1 . Par contre, au-delà de la consistance simple, la Σ_1 -correction est nécessaire, ce qui pose un problème évident de rattachement au finitisme standard, dans la mesure où les énoncés Σ_1 n'y sont pas considérés comme réels. Bernays a cependant ouvert la possibilité d'appliquer la vérifiabilité aux énoncés Σ_1 , à la condition de disposer d'un moyen *finitiste* pour trouver au moins une instance n qui vérifie l'énoncé : $\exists x \varphi_0(x)$, sans avoir à parcourir la suite infinie des valeurs de x . Il faudrait donc disposer d'une solution de finitisme qui permette simultanément d'assurer la Σ_1 -correction et en plus de trouver à chaque fois la valeur d'au moins une instance n (quand elle existe), donc de pouvoir affirmer qu'il n'y en a pas quand c'est le cas, ce qui reviendrait à disposer d'un critère finitiste de *décidabilité* des énoncés Σ_1 .

Le pas suivant consiste à introduire l'interprétation (syntaxique) sans doute la plus naturelle de la règle- ω selon la formulation de Hilbert *de 1930*, qui revient à imposer la 1-complétude (niveau 4 du schéma C). La règle- ω est encore limitée aux seuls énoncés Π_1 . Par contre, il est nécessaire d'étendre le critère finitiste permettant d'affirmer qu'un énoncé élémentaire $\varphi_0(x)$ est vrai, au-delà du critère de vérifiabilité de Bernays. La justification de son caractère finitiste devient alors problématique, dans la mesure où cela revient à considérer comme finitistes les énoncés Π_1 « en tant que tels », en opposition frontale avec l'exclusion dont ils avaient justement été l'objet dans la cadre du finitisme originel (§ 1.3). Gödel avait clairement perçu ce problème dès 1931 (lettre à Bernays du 2 avril 1931 et échange avec Carnap de mai 1931), en voyant l'utilisation d'un tel expédient comme une étape compromettant le Programme de Hilbert [Feferman 2003, p. 45]. Bernays le reconnaît lui-même sous une formulation plus ambiguë :

¹⁹⁸ Nous avons déjà signalé que c'est dans ce but que Bernays a introduit sa notion de « consistance externe » (§ 1.5).

Dans notre preuve de non-contradiction, on utilise parfois une forme d'induction complète qui ne se formalise pas par le schéma d'induction de la théorie réursive des nombres. Cette espèce de dépassement de la théorie réursive des nombres repose sur le fait que, dans nos formulations de preuves, interviennent quelquefois des hypothèses relatives à la validité de propositions universelles [...] un énoncé qui contient une telle hypothèse *présente aussi une difficulté pour l'attitude finitiste*. Il semble donc approprié que des points, dans lesquels un théorème général intervient comme prémisse, soient traités, dans le cadre de la réflexion finitiste, non en théorèmes, mais en règles d'inférence comme nous l'avons déjà fait [§ 3.4] pour les critères de réfutabilité des formules du calcul des prédicats. [...] Alors nous ne pouvons pas représenter la deuxième prémisse du schéma d'induction par une formule $A(n) \rightarrow A(n')$ dans le cadre de la théorie réursive des nombres.

[Hilbert, Bernays 1939 (2001), p. 411 ; souligné par nous ; le commentaire entre crochets est nôtre]

Mais pour formaliser certains résultats généraux de la théorie de la preuve, il est souhaitable d'admettre aussi comme théorèmes [...] par exemple des propositions dans lesquelles la vérifiabilité de formules où la calculabilité de fonctions constitue une prémisse. Et *un tel procédé ne semble pas non plus un manquement à l'égard des idées fondamentales de la théorie finitiste de la preuve*.

[Hilbert, Bernays 1939 (2001), p. 414-415 ; souligné par nous]

L'étape ultime consiste évidemment à généraliser la précédente à l'ensemble des énoncés¹⁹⁹, ce qui revient à imposer cette fois la ω -complétude (niveau 5 du schéma C). C'est l'interprétation (syntaxique) la plus naturelle de la règle- ω selon la formulation de Hilbert *de 1931*. Elle nécessite bien entendu une nouvelle extension du finitisme. Mais on ne voit alors plus du tout comment en justifier le caractère finitiste, malgré la déclaration opposée de Hilbert que nous avons déjà citée. Wilfried Sieg suggère que l'idée de Hilbert était alors de se raccrocher au constructivisme tel qu'il est proposé par Gentzen dans son théorème de 1936 [Sieg 2013c, p. 172, n. 30]. Nous avons cependant montré (§ 3.2) qu'un tel constructivisme ne peut raisonnablement pas être considéré comme finitiste, même en un sens étendu. Wilfried Sieg ne dit lui-même pas autre chose, en soulignant que ce point est explicitement reconnu par Bernays dans une correspondance *privée* adressée à Hilbert le 11 octobre 1931 [Sieg 2013c, p. 174]. Le même point sera formulé publiquement, mais de manière plus ambiguë, quelques années plus tard :

¹⁹⁹ Daniel Isaacson : [Isaacson 1992] et à sa suite Roman Murawski : [Murawski 2002], ont proposé une évolution plus graduelle caractérisée par une hiérarchie de systèmes formels auxquels la règle- ω a été appliquée : PA (Σ_1 -complétude), PA¹ (Π_1 -complétude, Σ_2 -complétude), PA² (Π_2 -complétude, Σ_3 -complétude), ... PA ^{ω} (ω -complétude).

Mais une fois qu'on a franchi cette étape, on est conduit à aller encore plus loin dans la même direction ; une telle autre étape [...] consiste en ce que des constatations, dans lesquelles des propositions universelles interviennent comme prémisses, soient elles-mêmes utilisées comme prémisses d'inférences d'induction.

Or on ne voit pas clairement où prendre appui pour admettre encore des extensions de méthode dans la direction indiquée, si nous voulons conserver les tendances fondamentales de la théorie de la preuve.
[Hilbert, Bernays 1939 (2001), p. 415 ; souligné par nous]

Signalons enfin le témoignage d'Olga Taussky – Todd de 1987, selon lequel Gödel se répandit en invectives contre Hilbert à la publication de celui-ci de juillet 1931, disant quelque chose comme : « comment peut-il écrire un article pareil après ce que j'ai fait ? » [Feferman 2011, p. 129, n. 12].

Comme nous l'avons indiqué, au-delà des solutions simples que nous venons d'évoquer, Daniel Isaacson a proposé dans [Isaacson 1992] une interprétation syntaxique de la règle- ω beaucoup plus subtile, que l'auteur considère comme l'extension la plus naturelle du finitisme originel [Isaacson 1992, p. 119]. Cette interprétation repose sur [Rosser 1937] et [Feferman 1962] et correspond au codage dans un système formel, de la dérivabilité de chaque $\varphi_0(x)$ qui en exprime les prémisses. Elle passe par différentes formulations, selon le système formel dans lequel la dérivation est formellement prise en compte et le système formel (possiblement différent) dans lequel ce fait est lui-même dérivable. De manière plus précise, Daniel Isaacson est conduit successivement à trois formulations :

1) PA' ; le codage de chaque $\varphi_0(x)$ est effectué dans PA , ce fait étant lui-même dérivable dans PA ; Il montre alors que G_{PA} (l'énoncé indécidable du 1^{er} théorème de Gödel de 1931 appliqué à PA) est dérivable dans PA' , Con_{PA} est dérivable dans PA' , mais certains énoncés \prod_1 de PA' vrais dans \mathcal{M}_0 ne sont pas dérivables dans PA' , ce qui fait que PA' n'est pas syntaxiquement complet ;

2) PRA' ; le codage de chaque $\varphi_0(x)$ est effectué dans PRA (en fait dans le système formel $I\Sigma_1$, de même force que PRA : dont l'ordinal limite est ω^ω , mais avec le langage de PA : voir le tableau I du § 1.4) ce fait étant lui-même dérivable dans PRA (donc finitiste selon la thèse de Tait) ; Con_{PRA} est dérivable dans PRA' , mais Con_{PA} n'est évidemment pas dérivable dans PRA' ;

3) PA'' ; le codage de chaque $\varphi_0(x)$ est effectué dans PA , ce fait étant lui-même dérivable dans PRA (donc finitiste) ; L'auteur considère que cette solution est la traduction formelle la plus naturelle de la règle- ω , en portant sur PA et en répondant aux exigences finitistes de Hilbert

(via la thèse de Tait) ; Il montre alors que bien que les prémisses de PA'' soient plus fortes que celles de PA' , PA'' est équivalent à PA' et possède donc les mêmes limitations [Isaacson 1992, p. 126]. Par contre PA'' n'implique pas PRA' .

Si l'interprétation d'Isaacson présente ainsi l'avantage de ne pas exiger d'extension du finitisme original (interprété en suivant la thèse de Tait), elle bute sur les mêmes pierres d'achoppement que les interprétations plus simples et plus classiques que nous avons évoquées : elle ne permet pas de passer outre les théorèmes de Gödel de 1931 appliqués à PA , ni-même d'établir la consistance de PRA (condition de possibilité du Programme selon l'interprétation standard de la thèse de Tait) et le nouveau système formel obtenu (PA'' ou PA') reste non effectif.

L'impression générale qui ressort finalement de cette analyse est que l'introduction par Hilbert de la règle- ω sous la forme de décembre 1930 et a fortiori sous celle de juillet 1931, le conduit en fait à perdre quasiment sur tous les plans : 1) la complétude syntaxique du système formel obtenu (lorsque la règle est interprétée formellement) n'est pas obtenue, 2) le système formel obtenu n'est même pas effectif (toujours lorsque la règle est interprétée formellement), 3) si l'intérêt épistémologique attaché au finitisme initial (son pouvoir de conviction lié à son évidence) est conservé avec l'interprétation complexe proposée par Daniel Isaacson, il disparaît sans solution de remplacement crédible avec les interprétations les plus simples associées à la 1-complétude et la ω -complétude. Il est alors intéressant de revenir plus généralement à la position défendue par l'école de Hilbert après les différentes solutions techniques proposées après l'irruption des théorèmes de Gödel de 1931, pour tenter de répondre aux deux questions que nous avons évoquées : quel contenu de finitisme élargi celle-ci a-t-elle proposé et surtout avec quelle justification philosophique compte tenu des attendus du Programme ? Pour cela, nous ne disposons d'aucun écrit de Hilbert (au-delà de sa préface à [Hilbert, Bernays 1934 (2001), p. 43] que nous avons déjà citée) et seulement des commentaires de Bernays (principalement : [Hilbert, Bernays 1939 (2001), § 5.3] ainsi que les textes rassemblés dans [Bernays 1976a (2003)]). De ces éléments il ressort sans ambiguïté l'affirmation par Hilbert et Bernays de la nécessité d'élargir le point de vue finitiste initial. Par contre, le contenu et surtout l'ampleur de cet élargissement ne sont pas clairement définis, les auteurs donnant l'impression de faire feu de tout bois à partir de tous les éléments techniques disponibles sur le moment, c'est-à-dire, au-delà de la règle- ω , d'une part le résultat de Gödel de 1933 et d'autre part celui de Gentzen de 1936²⁰⁰. Concernant la question de savoir si les présupposés de l'intuitionnisme sont acceptables comme finitisme étendu, on constate une évolution indéniable de la pensée de

²⁰⁰ Nous n'avons pas trouvé d'allusion explicite de Bernays à l'article de Gödel de 1958 analysé dans ce sens.

l'école de Hilbert entre 1934, où l'intuitionnisme est considéré comme une extension du finitisme et 1939, où le système formel HA est considéré comme finitiste :

Le point de vue de méthode de « l'intuitionnisme », que Brouwer admet par principe, constitue une certaine *extension de l'attitude finitiste*, [...]. Une telle version élargie du point de vue finitiste, [...], s'avère nécessaire quand on veut parvenir, au moyen de traitements finitistes, au-delà d'un certain domaine élémentaire. [Hilbert, Bernays 1934 (2001), p. 97-98]

Le formalisme (**Z**) constitue un système partiel du formalisme que Heyting a établi pour formaliser l'arithmétique intuitionniste [...] En fait, le cours de la preuve dont nous avons tiré qu'une contradiction dans (Z) [PA] a pour conséquence une contradiction dans (**Z**) aussi, *est bien tout à fait finitiste*. [Hilbert, Bernays 1939 (2001), p. 427 ; souligné par nous]

Cependant aucune justification philosophique n'est alors donnée à l'appui de cette position, en particulier de la dernière affirmation citée. Quant à la position de l'école de Hilbert à l'égard de la démarche de Gentzen de 1936, elle semble limitée au rattachement de cette démarche à la seule utilisation de la règle- ω (sans en préciser d'ailleurs la version) :

[...] l'écart [de la démarche de Gentzen 1936], avec les procédés de conduite finitiste de preuve que nous avons utilisés jusqu'ici, réside purement et simplement dans l'admission de propositions universelles comme prémisses de propositions. [Hilbert, Bernays 1939 (2001), p. 435 ; le commentaire entre crochets est nôtre]

C'est seulement dans [Hilbert, Bernays 1939, § 5.3] que l'on trouve une tentative par Bernays d'explicitier et de justifier l'extension du finitisme. L'explication reste toutefois très générale et elle est un peu énigmatique :

Une hypothèse finitiste se rapporte toujours à une situation caractérisée de façon intuitive. Mais on ne peut pas accorder la même valeur à l'idée qu'une proposition universelle est tenue pour valable. Certes, au lieu de l'hypothèse que la proposition universelle en question est tenue pour valable, on peut prendre l'hypothèse que la proposition est vérifiée [reconnue comme exacte]. Donc on suppose qu'une preuve existe. Mais en l'occurrence, c'est *en contenu* que la preuve doit être supposée valide ; et, pour une preuve, la forme probante en contenu n'a pas non plus de caractéristique intuitive. [Hilbert, Bernays 1939 (2001), p. 411 ; dans cette traduction, pour en faciliter la compréhension, nous avons remplacé le terme « tient » par « est tenue pour valable »²⁰¹]

²⁰¹ Suivant en cela le sens proposé par Marcel Guillaume (le principal traducteur de [Hilbert, Bernays 1934] et [Hilbert, Bernays 1939]) dans une communication personnelle du 30 octobre 2011.

L'auteur précise alors sa pensée, en en se rattachant à la vérifiabilité (au sens technique) :

Mais en fait, même dans le cas où nous utilisons une hypothèse comme celle de la vérifiabilité d'une formule A , on ne se réfère pas du tout à la nature d'une éventuelle preuve, en termes de contenu, de la vérifiabilité de A ; ce dont il s'agit dans cette hypothèse, c'est plutôt que nous sommes en mesure, à chaque endroit du raisonnement où on parle d'une formule provenant de A par substitution, de faire usage du fait que la formule considérée est une formule vraie. [Hilbert, Bernays 1939 (2001), p. 411]

L'idée suggérée par Bernays, est donc que la possession de la preuve finitiste d'un énoncé Π_1 va en fait déjà bien au-delà de la seule exigence de vérifiabilité de cet énoncé, qui se borne à la requête que l'on puisse établir la vérité (au sens intuitif) de chacune de ses instances²⁰², en ne limitant pas le nombre de ces instances. Une telle remarque permet effectivement de rendre compte de l'impossibilité d'établir la consistance simple de PRA à partir des seuls moyens de PRA. Mais comme nous l'avons montré, il est parfaitement possible d'imaginer une extension du finitisme raisonnable mais limitée qui y réponde, sans justifier pour autant l'extension beaucoup plus large introduite par la règle- ω dans sa première version. Celle-ci constitue en effet une rupture profonde par rapport au fondement de la position finitiste, selon lequel les énoncés Π_1 ne doivent être considérés que comme un « jugement hypothétique ». La première conséquence en est la perte induite du caractère précis de ce qu'est l'exigence finitiste, ce dont l'auteur est conscient :

Assurément, telle qu'elle vient d'être formulée, cette question n'est pas précise ; car nous avons introduit l'expression « finitiste », non pas comme un terme strictement délimité, mais seulement pour désigner la ligne directrice d'une méthode, grâce à laquelle certaines catégories de formation de concepts et de raisonnements peuvent être reconnues comme sûrement finitistes, et certaines autres comme non-finitistes, mais qui cependant ne fournit *aucune ligne de séparation rigoureuse* entre les catégories qui satisfont aux exigences de la méthode finitiste, et celles qui n'y satisfont plus.

[Hilbert, Bernays 1939 (2001), p. 414 ; souligné par nous]

Le problème qui se pose alors est que l'introduction de flou sur quelque chose d'aussi fondamental, du point de vue finitiste, que la délimitation même du raisonnement finitiste²⁰³, s'oppose frontalement à la démarche revendiquée par Hilbert d'assurer la connaissance mathématique sur des bases indiscutables ([Hilbert 1931 (1992), p. 188] déjà cité). Par ailleurs,

²⁰² Nous remercions ici Marcel Guillaume d'avoir attiré notre attention sur ce point.

²⁰³ Nonobstant l'ambiguïté de cette délimitation dans les textes de l'école de Hilbert (d'avant 1931), sur laquelle nous avons beaucoup insisté au § 1.

à partir du moment où ce premier pas est franchi, en ayant ouvert la « boîte de Pandore » de l'infini en acte, l'acceptation de la règle- ω dans sa seconde version s'ensuit naturellement. Il en résulte un nouveau problème de devoir décider où et sur la base de quels critères, s'arrêter. Bernays est là aussi conscient du problème, sans qu'il n'y apporte pour autant de réponse :

Mais une fois que l'on a franchi cette étape, on est conduit à aller encore plus loin dans la même direction. Une telle autre étape, déjà mentionnée, consiste en ce que des constatations, dans lesquelles des propositions universelles interviennent comme prémisses, soient elles-mêmes utilisées comme prémisses d'inférences d'induction. *On ne voit pas clairement où prendre appui pour admettre encore des extensions de méthode dans la direction indiquée, si nous voulons conserver les tendances fondamentales de la théorie de la preuve.* [Hilbert, Bernays 1939 (2001), p. 415 ; souligné par nous]

Pour tenter de comprendre les raisons d'une telle attitude de l'école de Hilbert, il convient sans doute de dissocier celles de Hilbert de celles de Bernays (au-delà de celles de Gentzen). Il est vraisemblable que jusqu'à sa mort Hilbert n'ait en fait pas accepté que son programme puisse être remis en cause. En témoigne, au-delà de sa réaction (privée) de colère à l'annonce du résultat de Gödel de 1931 (lettre de Bernays à Constance Reid du 3 août 1966 : [Feferman, 1986, p. 210]), la formulation publique de l'introduction à [Hilbert, Bernays 1934] que nous avons déjà citée (§ 3). Il s'est alors manifestement enfermé dans un combat d'arrière-garde, au travers de l'introduction des deux versions successives de la règle- ω , sans accepter de reconnaître qu'alors que l'objectif du Programme n'était pas atteint pour autant, tout l'intérêt épistémologique du finitisme initial sur lequel il avait tant insisté dans les années 1920 était perdu. Cette analyse rejoint celle de Constance Reid :

La confiance sans bornes dans le pouvoir de la pensée humaine, qui l'a conduit inexorablement à ce dernier grand travail de sa carrière, rendait maintenant presque impensable pour lui d'accepter émotionnellement le résultat de Gödel. Il y avait peut-être aussi, le rejet tout à fait humain du fait que la découverte de Gödel était la vérification de certaines indications, dont il avait lui-même jusqu'ici refusé de reconnaître la signification, que le cadre du formalisme n'était pas suffisant pour la charge qu'il voulait lui faire porter. [Reid 1970, p. 198]

Le cas de Bernays est très différent. Comme le mentionnent aussi bien Hourya Sinaceur que Solomon Feferman, sa vision philosophique personnelle le conduisait à une conception beaucoup plus libérale que celle de Hilbert, quant à l'intérêt même des finalités du Programme, et partant quant à la nécessité de maintenir les caractéristiques épistémologiques du finitisme initial [Sinaceur 2003, p. 10] et [Feferman 2011, p. 113]. Son exégèse nécessite alors de bien

distinguer ses écrits selon qu'ils sont antérieurs ou postérieurs à la mort de Hilbert en 1943. Dans les premiers, qui incluent en particulier [Hilbert, Bernays 1934] et [Hilbert, Bernays 1939], il est clair qu'il s'est astreint à tenter de restituer la position de l'école de Hilbert (et donc celle de Hilbert) à laquelle il était rattaché²⁰⁴ et ce *même lorsqu'il n'en partageait manifestement pas (ou plus) tous les éléments*. On peut signaler à ce sujet la remarque *privée* sur la règle- ω (dans la seconde version) que nous avons déjà évoquée, adressée par Bernays à Hilbert dans une lettre du 11 octobre 1931, dans laquelle il souligne que l'évolution proposée n'est pas compatible avec le finitisme défini dans [Hilbert, Bernays 1934]. Concernant ses écrits postérieurs à la mort de Hilbert, Bernays s'est peu exprimé, et surtout tardivement, sur sa vision personnelle. C'est ainsi que c'est principalement dans les années 1970, donc à la fin de sa vie, qu'il introduit des réserves sur le fait que de telles extensions du finitisme, au-delà du fait de s'écarter indéniablement du finitisme original, soient même acceptables :

Cependant, la notion de conséquence matérielle *nous éloigne totalement des idées de méthode de la théorie de la preuve de Hilbert*, et le problème est de savoir si, dans la démonstration de la non-contradiction du système (Z), nous ne pouvons pas éviter cette notion de conséquence matérielle, quoique nous allions au-delà du raisonnement finitiste, dans la mesure où nous admettons des propositions universelles comme prémisses.

[Hilbert, Bernays 1939/1970 (2001), p. 429, ajout de la seconde édition ; souligné par nous]

Cette réserve se transforme ensuite en la reconnaissance que la justification philosophique attendue reste encore à établir et ne l'a donc pas été [Bernays 1971/1976 (2003), p. 213], [Bernays 1976b (2003), p. 82]. Il en vient enfin à contester la raison d'être du Programme, dont on rappelle que celui-ci vise, pour Hilbert, à résoudre *définitivement* le problème des fondements ainsi que le fondement du finitisme initial, à savoir, les signes concrets représentant *les nombres naturels finis* [Bernays 1970/1976 (2003), p. 209, 210].

Au-delà de l'école de Hilbert, on peut noter que Gödel, au moins jusqu'au début des années 1940, était parfaitement lucide quant à l'incompatibilité des différentes solutions proposées d'extension du finitisme avec la spécificité des exigences finitistes. Nous l'avons déjà évoqué à propos de [Gödel 1933] et [Gentzen 1936]. On peut également mentionner sa réaction dans

²⁰⁴ Selon Marcel Guillaume, dans deux communications personnelles : du 24 décembre 2011 et du 19 janvier 2012, la raison pour laquelle Bernays n'a pas au minimum introduit des réserves philosophiques aux extensions évoquées du finitisme, tient aussi au contexte historique et social de l'époque (la prise de pouvoir par les Nazis, Bernays étant lui-même considéré comme juif, ce qui le conduira à quitter Göttingen en 1934 [Reid 1970, p. 206], la précarité de son statut universitaire à Göttingen et enfin, pour la première édition du second tome, l'imminence de la guerre). Il considère également que c'est par devoir moral auprès de Hilbert décédé, que Bernays n'a pas voulu introduire de telles réserves dans la seconde édition des deux tomes.

la lettre à Bernays du 16 janvier 1942, à l'article de Bernays de 1941, dans lequel celui-ci ne propose rien de moins que de « ne pas concevoir l'évidence et la sûreté de façon trop absolue » [Bernays 1941, p. 152] :

[...] ; seul ce que vous dites à la page 152, lignes 8 à 11 n'est pas compréhensible pour moi. Ne serait-ce pas équivalent à abandonner le point de vue formaliste ? [Gödel 1942 (2003), p. 133]

3.3 L'inépuisabilité des mathématiques selon Franzen

Le terme « d'inépuisabilité des mathématiques » a été introduit par Torkel Franzen²⁰⁵, qui lui a consacré un ouvrage [Franzen 2004] pour traduire une conséquence positive des théorèmes de Gödel de 1931, à savoir : l'héritabilité de la correction (au sens formel) de théories de plus en plus puissantes issues de PA. Voici la présentation qu'en fait l'auteur :

Ainsi dans ce cas, le second théorème a la conséquence positive que nous pouvons toujours étendre tout système [formel] que nous reconnaissons comme correct [par rapport au modèle standard] en un système [formel] plus puissant que nous reconnaissons également comme correct, en ajoutant comme nouvel axiome l'affirmation que le système [formel] original est consistant. [...] Chaque fois que nous décidons de définir une théorie comme PA_{ω} qui est démontrablement correcte étant donné PA correct, nous sommes également capables de l'étendre en une théorie plus forte, qui est également correcte. [Franzen 2005, p. 113 ; les commentaires entre crochets sont nôtres]

Bien qu'il n'en fasse pas état, une telle propriété semble pour le moins paradoxale dans le contexte du Programme de Hilbert, puisqu'elle revient à permettre indirectement de « gravir les échelons » en matière de correction, donc de consistance, pour des théories de plus en plus fortes issues de PA. Cela justifie que nous en analysions le contenu, en commençant par remarquer que l'auteur affirme s'appuyer sur une remarque de Gödel mentionnée dans [Gödel 1951 (1995)] sur *l'incomplétabilité* des mathématiques, pour en déduire sa propre proposition sur *l'inépuisabilité* des mathématiques. Nous rappelons la remarque en question de Gödel :

Dans la mesure où il [le présent argument] *rend impossible que quelqu'un puisse mettre en place un certain système d'axiomes et de règles bien défini et émettre l'assertion suivante à son sujet : je perçois*

²⁰⁵ L'argument ayant été quelque peu anticipé (mais non développé) par Michael Dummett [Dummett 1994, p. 337].

(avec certitude mathématique) tous ces axiomes et ces règles comme valables²⁰⁶ et de plus, je crois qu'ils contiennent toutes les mathématiques. Si quelqu'un propose une telle formulation, il se contredit lui-même. Dans la mesure où s'il perçoit les axiomes considérés comme valables, il perçoit aussi (avec la même certitude) qu'ils sont consistants.

[Gödel 1951 (1995), p. 309 ; le commentaire entre crochets est nôtre]

Le fond de cette remarque, qui repose sur l'interprétation par Gödel du caractère « complet » (au sens informel) des mathématiques à partir d'un système formel donné, comme devant inclure le caractère consistant de ce système formel, découle directement du second théorème de 1931. Dans ce sens particulier de l'incomplétabilité des mathématiques, cette remarque n'est ni contestable ni problématique. Il en est tout autrement de la proposition de Torkel Franzen, ce que celui-ci reconnaît d'ailleurs en précisant que sa proposition n'est « ni non controversée ni non problématique » [Franzen 2004, p. 2].

Précisons ensuite la portée et la limite de l'héritabilité que défend l'auteur. Celui-ci fait référence à la hiérarchie classique entre les trois principales formes de consistance se rapportant à l'arithmétique²⁰⁷ (se reporter au schéma C du § 1.5) : la consistance simple, la Σ_1 -correction (co extensive à la 1-consistance), et la correction (co extensive à la ω -consistance). Il précise alors que si l'on suppose la correction de PA, on peut en déduire la correction (et partant la consistance) de $PA_1 = PA +$ l'axiome « PA est consistant », puis de $PA_2 = PA_1 +$ l'axiome « PA_1 est consistant », Ce résultat est étendu (sous la même hypothèse) avec la série : $PA_1' = PA +$ l'axiome « PA est 1-consistant », $PA_2' = PA_1' +$ l'axiome « PA_1' est 1-consistant », Par contre il ne s'applique *pas* à la série : $PA_1'' = PA +$ l'axiome « PA est correct », $PA_2'' = PA_1'' +$ l'axiome « PA_1'' est correct », Avec les mêmes notations, si l'on suppose seulement la 1-consistance de PA (au lieu de sa correction), le résultat s'applique encore à la série : PA, PA_1 , PA_2 , ..., mais non à la série : PA, PA_1' , PA_2' , ... (et évidemment encore moins à la série : PA, PA_1'' , PA_2'' , ...). Enfin, si l'on se limite à l'hypothèse la plus faible : PA est consistant, alors le résultat ne s'applique même pas à la série : PA, PA_1 , PA_2 , ... (en cohérence avec le second théorème de 1931). Pour justifier son résultat Torkel Franzen utilise un raisonnement (non formel) particulièrement simple [Franzen 2004, p. 186] :

1) PA étant supposé correct, tous les axiomes de PA sont vrais et les règles de dérivation préservent la vérité ;

²⁰⁶ Nous avons traduit ici et dans l'analyse qui suit par « valable », le terme anglais « correct » utilisé par Gödel et par Torkel Franzen, pour bien marquer la distinction avec la propriété métamathématique de « correction » (« soundness » en Anglais), également utilisée par Torkel Franzen dans son ouvrage.

²⁰⁷ Ce qui revient implicitement à affecter un sens technique précis au terme général : « valable ».

2) comme PA est correct, PA est consistant et l'énoncé (exprimable dans PA) : « PA est consistant » est vrai ; les axiomes de PA_1 sont donc vrais ;

3) les règles de dérivation de PA_1 étant les mêmes que celles de PA, elles préservent la vérité, donc les théorèmes de PA_1 sont vrais et PA_1 est correct.

On reconnaît là la « démonstration » classique de consistance de PA sans portée épistémologique, dénoncée en particulier par Jacques Dubucs [Blanché, Dubucs 1996, p. 368] et qui est issue de [Turing 1939] et [Feferman 1962] auxquels Torkel Franzen se réfère explicitement. Il est clair qu'elle fait appel à des éléments ensemblistes qui vont au-delà de l'arithmétique, ce qui se traduit par le fait que les deux résultats cités sont non effectifs, ce que l'auteur reconnaît [Franzen 2004, p. 208]. Solomon Feferman est particulièrement éclairant sur la nature précise de cette non-effectivité :

Il y a cependant une entourloupette à ce résultat [Turing 1939], nommément qu'il est très dépendant de *comment* le passage dans le transfini est effectué et il n'y a aucune voie évidente à l'avance pour préférer une telle voie plutôt qu'une autre. [...] Mais, sur le fond, l'entourloupette est que *nous avons échangé l'incomplétude par l'incertitude sur la voie à suivre* [pour effectuer le passage dans le transfini] ; en d'autres mots, vous ne pouvez pas obtenir quelque chose à partir de rien.
[Feferman 1998a, p. 17 ; les commentaires entre crochets sont nôtres]

Dans les années 1960, j'ai obtenu un renforcement considérable du résultat de complétude de Turing, [...]. Cependant, ce théorème de complétude souffre de la même entourloupette que le précédent résultat de Turing, dans la mesure où il est dépendant des voies qui sont prises dans le transfini, dont le choix n'est pas justifié à l'avance. [Feferman 1998a, p. 18]

La subtilité de l'argument de Torkel Franzen repose en fait sur l'opposition entre les caractères, non exprimable de la vérité et exprimable de la dérivabilité, dans PA, associées au second théorème de Gödel de 1931 [Franzen 2004, p. 11]. Un tel résultat n'est donc pas inintéressant par rapport à la question de l'aptitude et des limites pour un système formel, à rendre compte des propriétés métamathématiques qui peuvent lui être associées, *en supposant que ce système formel est correct* (par rapport au modèle standard) et a fortiori consistant. C'est bien là le cadre de réflexion au sein duquel se situe Torkel Franzen : « L'inépuisabilité, le sujet principal de ce livre, est centrée sur le cas où nous acceptons les axiomes d'une théorie comme vrais » [Franzen 2004, p. 218]. Mais ce résultat ne peut être d'aucun secours si c'est justement la justification de la consistance du système formel qui est en jeu, comme c'est le cas du programme de Hilbert.

3.4 L'instrumentalisme de Detlefsen

Nous avons déjà discuté les positions défendues par Michael Detlefsen sur le programme de Hilbert (aux § 1.5, 2.2, et 2.3), mais en nous limitant alors au contexte du programme de Hilbert au sens restreint selon *notre* interprétation (et également selon l'interprétation standard). Nous allons maintenant aborder la propre interprétation de Michael Detlefsen du programme de Hilbert au sens restreint, avec l'objectif d'en analyser la pertinence épistémologique et d'examiner dans quelle limite cette interprétation résiste quant à elle aux deux théorèmes de Gödel de 1931. Il convient d'emblée de souligner que si la démarche précise proposée par Michael Detlefsen est sans aucun doute inspirée par le programme de Hilbert originel, elle s'éloigne cependant de celui-ci sur des points essentiels, à la fois quant à ses présupposés et quant à son objectif. L'auteur a d'ailleurs l'honnêteté de le reconnaître lui-même [Detlefsen 1986, p. 40-41, n. 26] et [Detlefsen 1986, p. 57]. Le trait majeur de l'interprétation de Michael Detlefsen est son caractère *instrumentaliste* (en mathématique) mais sous une forme à la fois très particulière et sophistiquée, comme l'auteur le reconnaît [Detlefsen 1986, p. x]. Il ne s'agit plus ici de chercher à fonder l'ensemble des mathématiques d'une manière « absolue » mais de rechercher simplement la plus grande efficacité possible dans les démonstrations d'un corpus *particulier* des mathématiques (preuves idéales « qui sont de quelque valeur comme instrument de l'acquisition épistémique humaine » [Detlefsen 1986, p. 84]), dont on se sera assuré au préalable de sa sûreté relative [Detlefsen 1990, p. 358]. L'assise intuitive retenue est également différente de celle de Hilbert, en se fondant en l'occurrence sur un *système formel* particulier (nommé S_D ²⁰⁸), au lieu d'un certain corps de vérités intuitives, dont les théorèmes *fixent* alors la vérité intuitive et dont la consistance est bien sûr admise a priori [Detlefsen 1990, p. 361]. Il en résulte un changement radical de la signification du vocabulaire hilbertien selon l'interprétation de Michael Detlefsen, par rapport à l'interprétation que nous avons retenue²⁰⁹. Pour Michael Detlefsen, sont à la fois et indifféremment : « réels », « finitaires », « contentuels », et « pleins de sens », tous les énoncés exprimables (au sens formel) dans S_D ; par extension, sont appelées : méthodes « réelles » ou « contentuelles » et raisonnements « réels » ou « finitaires », les dérivations dans S_D . Sont appelés : « idéaux » les énoncés d'un

²⁰⁸ Pour se prémunir contre un risque de confusion avec les notations retenues dans le cadre de notre interprétation du programme de Hilbert, nous faisons suivre les notations retenues par Michael Detlefsen de l'indice : D.

²⁰⁹ En perdant au passage le bénéfice de la distinction méthodologique que nous avons introduite, lorsque nous avons soigneusement distingué les couples que nous avons appelés respectivement : R versus I et S versus T (tableau II, § 1.5).

certain système formel (nommé T_D) qui sont exprimables dans S_D ²¹⁰, mais dont la preuve, lorsqu'il s'agit de théorèmes, peut avoir été obtenue à l'aide de dérivation de T_D , soit des moyens plus puissants que ceux de S_D (dérivations qui sont appelées raisonnements « idéaux » ou méthodes « idéales ») [Detlefsen 1986, p. 38, n. 14]. L'auteur introduit enfin une partition au sein des théorèmes de S_D , entre les théorèmes « problématiques » (nommés R_D) dont on veut remplacer les dérivations de S_D par celles de T_D , et les théorèmes « non problématiques » pour lesquels il est estimé qu'il n'y a aucun intérêt pragmatique à remplacer les dérivations dans S_D : [Detlefsen 1986, p. 18-19]. Si l'on tente alors de rassembler les indications détaillées mais assez confuses, présentes dans [Detlefsen 1986] et surtout dans [Detlefsen 1990], on aboutit à quatre relations caractéristiques²¹¹ concernant les liens métamathématiques existants entre les systèmes formels S_D et T_D :

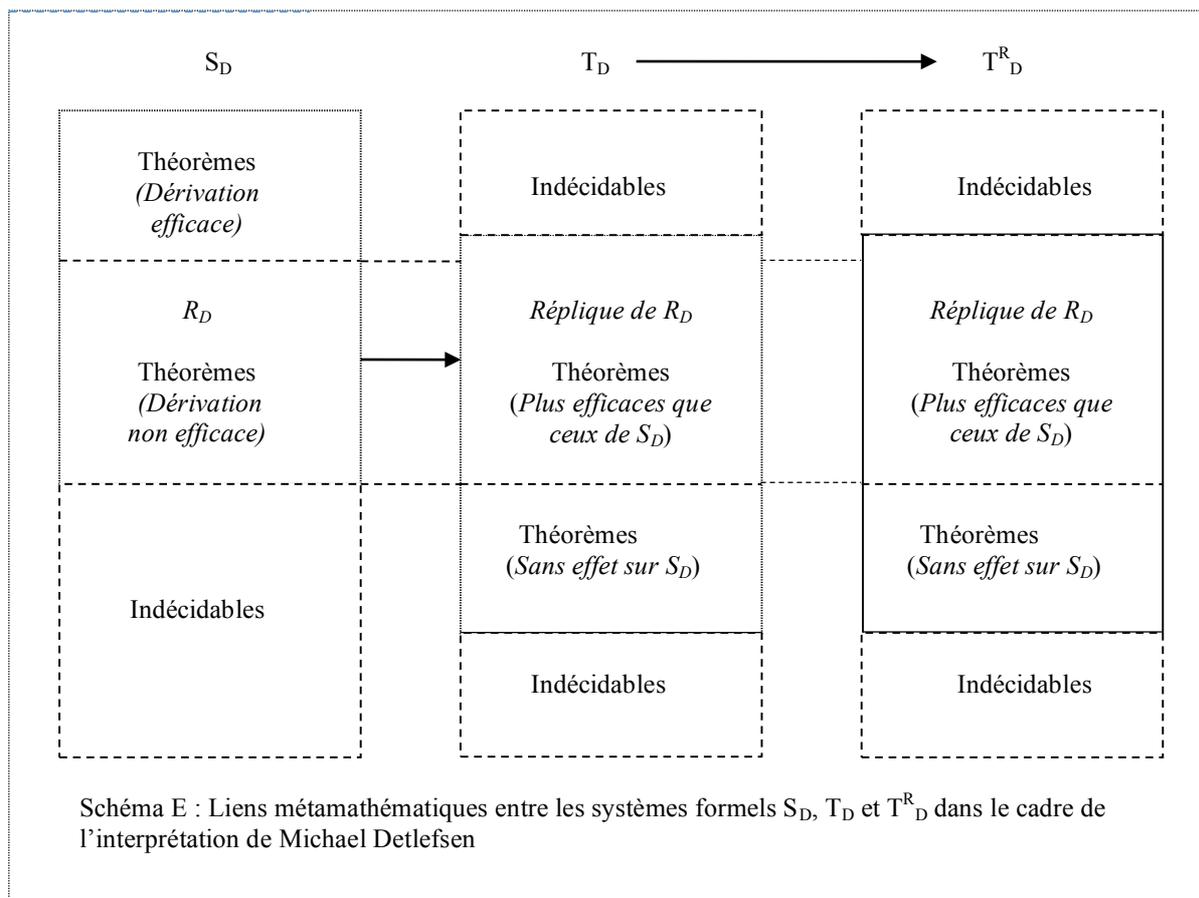
- 1) T_D peut ne pas décider certains théorèmes décidés par S_D ; T_D n'a pas à être une extension de S_D ²¹² [Detlefsen 1990, n. 21, p. 375-376] ;
- 2) T_D peut décider certains théorèmes qui ne sont pas décidés par S_D ; T_D n'a pas à être conservatif par rapport à S_D [Detlefsen 1990, p. 361] ;
- 3) le caractère d'extension de T_D par rapport à S_D se limite à la partie R_D de S_D , soit la condition que tout théorème de S_D appartenant à R_D doit être un théorème de T_D [Detlefsen 1990, p. 369] ;
- 4) la seule chose qui est interdite par l'autorité épistémique de S_D est qu'un énoncé soit à la fois dérivé par T_D et *réfuté* par S_D [Detlefsen 1990, p. 358, 361-362].

Pour plus de clarté, nous avons représenté ces relations sur le schéma E ci-après (limité aux énoncés vrais dans le modèle standard de S_D ou de T_D pour une raison de lisibilité), auquel a été ajouté le système formel T_D^R qui intervient dans ce qui suit.

²¹⁰ Soit parce que l'ensemble des énoncés exprimables dans T_D ne va pas au-delà de ceux exprimables dans S_D , soit parce que l'on ne s'intéresse qu'à ces derniers.

²¹¹ Il s'agit seulement des relations métamathématiques générales entre S_D et T_D . Doivent s'y ajouter des conditions spécifiques liées à la dérivabilité dans T_D , qui seront introduites dans le cadre de la discussion de l'impact du second théorème de Gödel de 1931 qui suit.

²¹² Les théorèmes de S_D qui n'appartiennent pas à R_D et qui ne sont pas destinés à être remplacés par des théorèmes de T_D , peuvent donc librement être des indécidables de T_D .



Il reste bien entendu à définir les *contenus* de S_D , R_D et T_D . S_D n'est pas strictement défini, même si l'auteur fait explicitement référence à la thèse de Tait, en présentant PRA comme « une borne du raisonnement finitaire » [Detlefsen 1986, p. 58] tout en laissant entendre la possibilité de son dépassement²¹³, en renvoyant en fait à un critère empirique (et vague) en relation avec le « comportement arithmétique des objets physiques » pour la fixation de son contenu : [Detlefsen 1986, p. 33, 58-59]. Il va jusqu'à évoquer le cas d'un désaccord possible entre ce qui nous paraît intuitivement vrai et ce qui peut être dérivé dans S_D , pour conclure que dans ce cas, il nous faut soit modifier S_D , soit adapter notre intuition à ce que permet S_D [Detlefsen 1986, p. 36, n. 3]. Concernant R_D , Michael Detlefsen précise que la délimitation correspondante ne répond qu'à un critère d'efficacité, en se référant à Georg Kreisel sur ce point [Detlefsen 1986, p. 36, n. 3] et [Detlefsen 1990, p. 370]. Il en résulte qu'il n'y a aucune exigence à démontrer « finitairement » (donc dans S_D) la troisième relation métamathématique mentionnée, selon laquelle tout théorème de S_D appartenant à R_D doit être un théorème de T_D [Detlefsen 1990,

²¹³ Michael Detlefsen défend actuellement (communication personnelle du 27/02/2012) la position que le point de vue finitaire ne peut pas être formalisé dans PRA, ni dans n'importe quel système formel (récurivement axiomatisable).

p. 370]. La situation est analogue pour T_D , pour lequel il est précisé que celui-ci ne peut pas être défini a priori et qu'à l'instar de R_D , sa fixation ne peut être qu'empirique, en relation avec la prise en compte de considérations « en pratique » (en opposition à celles « en principe » qui caractérisent les interprétations habituelles) [Detlefsen 1986, p. 83, 168]. Se pose alors bien entendu le problème de la définition d'un critère objectif de complexité, fondamental dans le cadre d'une démarche instrumentaliste, pour pouvoir circonscrire à la fois les preuves réelles qui seront conservées et les preuves idéales qui seront introduites, dont on imagine la difficulté. L'auteur en est conscient, tout en reconnaissant que le travail reste à faire [Detlefsen 1990, p. 376, n. 24]. Il n'y a par contre aucune ambiguïté sur le fait que la démonstration que tout théorème de la part de T_D qui est la réplique de R_D (quatrième relation métamathématique mentionnée), qui traduit la sûreté de l'instrumentalisme de Detlefsen, doit être finitaire (donc dans S_D). Il apparaît ainsi que l'interprétation du programme de Hilbert au sens restreint par Michael Detlefsen n'a qu'un lointain rapport avec l'interprétation standard (ou la nôtre), ce que celui-ci reconnaît [Detlefsen 1986, p. 75, n. 22]. L'objectif de Hilbert, d'assurer la sûreté de l'intégralité des mathématiques, est abandonné au profit d'un instrumentalisme appliqué à une portion seulement des mathématiques. La force de l'évidence intuitive de la base finitiste retenue par Hilbert est perdue par la référence abstraite à un système formel particulier comme point d'ancrage. Mais la faiblesse la plus grande de l'interprétation de Michael Detlefsen réside sans aucun doute dans l'absence d'une définition précise du contenu et des limites de ses constituants essentiels : non seulement le système formel « idéal » recherché (ce pourrait être un programme de travail), mais également l'ensemble des théorèmes « réels » que l'on souhaite pouvoir remplacer par des théorèmes « idéaux », et surtout la base « réelle » même sur laquelle tout cela est fondé. La justification pragmatique qui est donnée par l'auteur pour justifier cette absence de précision, nous paraît un bien faible attrait en regard du relativisme auquel celle-ci conduit. Il reste à s'assurer de la résistance d'une telle interprétation aux deux théorèmes de Gödel de 1931 pour déterminer dans quelle mesure ce dernier objectif au moins est atteint.

L'analyse de l'impact du premier théorème de Gödel de 1931 sur cette interprétation est immédiate, sur la base de l'analyse similaire déjà effectuée sur le programme de Hilbert selon notre interprétation (§ 2.2). Il apparaît en effet que le choix retenu par Michael Detlefsen est encore plus libéral que notre deuxième solution du § 1.5, puisque non seulement certains indécidables de S_D peuvent être décidés par T_D mais également, de manière réciproque, certains indécidables de T_D peuvent être décidés par S_D . Dans ces conditions, il est clair qu'une telle interprétation est immunisée contre l'effet du premier théorème de Gödel de 1931, sous la condition faible qu'à la fois G_S (l'énoncé de Gödel du premier théorème de 1931 appliqué à S_D)

et G_T (le même énoncé appliqué à T_D) n'appartiennent pas à R_D . On retrouve ainsi la position défendue par Michael Detlefsen et contestée par nous au § 2.2, selon laquelle le « programme de Hilbert » n'est pas affecté par le premier théorème de Gödel de 1931. La contradiction apparente vient simplement du fait que ce résultat s'applique bien à l'interprétation de Michael Detlefsen du programme de Hilbert, mais non au Programme selon notre interprétation (ou l'interprétation standard).

L'analyse de l'impact du second théorème de Gödel de 1931 sur la même interprétation est un peu plus subtile. Comme nous l'avons déjà indiqué (§ 2.3), le principe de la solution utilisée par Michael Detlefsen est de s'affranchir des conditions de dérivabilité (CD3 en particulier), qui sont nécessaires à l'établissement du second théorème, en utilisant pour cela une solution « à la Rosser ». Techniquement, cela revient à associer à une théorie donnée T , munie des prédicats standard de dérivabilité et de consistance, une théorie associée T^R , munie de prédicats spécifiques de dérivabilité et de consistance dans laquelle on peut dériver la consistance de T^R . Le point est alors que tous les axiomes et les théorèmes de T se retrouvent dans T^R , ce qui permet ainsi d'affirmer que l'on est bien assuré, en un certain sens, de la sûreté de T . De manière plus précise, la prouvabilité dans T^R (et partant la consistance associée) est construite à partir de la prouvabilité (classique) dans T en rajoutant la contrainte qu'aucune dérivation obtenue dans T en ayant un nombre de Gödel plus petit que celui de la dérivation correspondante dans T^R , n'est contradictoire avec celle-ci. On conçoit qu'un tel procédé soit effectif (c'est une caractéristique des solutions « à la Rosser »), mais il est indéniablement complexe et non intuitif. Cette complexité n'est pas niée par Michael Detlefsen, mais celui-ci considère qu'elle n'en est pas moins raisonnablement acceptable, avec un caractère particulier « dynamique » (opposé au caractère « statique » de la démarche classique) : [Detlefsen 1990, p 345]. Disons tout de suite que l'auteur prévient ne pas être en mesure d'assurer la seule véritable condition à vérifier qu'il impose à sa démarche (le point 4 de notre liste) : établir « finitairement » (dans S_D), que tout théorème de T_D qui est la réplique de R_D n'est pas réfutable dans S_D [Detlefsen 1986, p. 140-141, n. 48] et [Detlefsen 1990, p. 370]. Un tel aveu constitue évidemment une limitation très forte de cette interprétation. Mais au-delà de cette limitation, quelle est la réponse que l'instrumentalisme de Detlefsen apporte aux objections que nous avons déjà évoquées (§ 2.3), concernant les solutions « à la Rosser » qui prétendent s'affranchir du second théorème de Gödel de 1931 ? L'objection classique majeure est celle de l'impossibilité de justifier finitairement l'équivalence extensionnelle entre T et T^R , sans introduire la condition de dérivabilité CD3. La réponse est ici que l'instrumentalisme de Detlefsen est associé à la double précaution, d'une part de limiter le domaine d'intérêt à R_D (et non à S_D tout entier) et d'autre

part de ne pas exiger une démonstration finitaire (dans S_D) pour la fixation de l'étendue de R_D . Cela lui permet ainsi de contourner cette objection [Detlefsen 1990, p. 370]. On notera cependant que le mérite de ce résultat doit être tempéré par le fait que toute l'interprétation de Michael Detlefsen a été justement conçue *en vue de l'obtenir*, selon l'aveu même de l'auteur (dont on peut souligner l'honnêteté) :

Les parallèles entre le modèle instrumentaliste de construction d'une théorie présenté au-dessus et la variante [de formalisation] de Rosser ne sont pas un accident ; en fait, le premier survient comme *le résultat direct d'une tentative de donner un sens philosophique à la variante de Rosser* de la conception de la formalisation. [Detlefsen 1986, p. 122 ; souligné par nous]

Concernant l'autre objection majeure qui est l'argument de [Gödel 1972b, remarque 1], la réponse de l'auteur, que nous avons déjà mentionnée dans le cadre de notre interprétation du programme de Hilbert (§ 2.3), fournit directement la réponse appropriée dans le présent contexte. En effet, pour passer de la consistance externe (au sens de Gödel) obtenue sans les conditions de dérivabilité, à la consistance simple seule pertinente dans le cadre de l'instrumentalisme de Detlefsen, la condition CD3 est nécessaire, ce qui ramène alors le résultat de 1972 à celui du second théorème de 1931. On remarque toutefois que si cette réponse est techniquement non contestable, elle met l'accent sur l'écart entre les notions intuitives de preuve et de consistance dans un contexte finitiste (habituel) d'une part et le caractère particulièrement abstrait et artificiel de leur pendant instrumentaliste « à la Detlefsen » d'autre part. Pour terminer avec les objections possibles²¹⁴, signalons au passage que la propre analyse de [Detlefsen 2001] ne concerne évidemment pas l'instrumentalisme de Detlefsen, puisque les conditions de dérivation ne s'y appliquent pas. Globalement, il apparaît ainsi que l'interprétation de Michael Detlefsen du programme de Hilbert permet bien de s'affranchir du second théorème de Gödel de 1931, contre [Smith 2007] qui n'a peut-être pas pris en compte avec suffisamment d'acuité toute la subtilité de son contenu²¹⁵.

Si l'on essaye de dresser un bilan des résultats obtenus, il s'avère que sur le plan technique, au-delà de la limitation forte de l'incapacité d'établir le critère fondamental à respecter dans le cadre de cette interprétation (la « correction » de T_D par rapport à R_D à établir dans S_D), cette dernière permet effectivement de passer outre à la fois des deux théorèmes de Gödel de 1931.

²¹⁴ Nous passons sur les objections de moindre importance (les quatre premières abordées au § 2.3), pour lesquelles nous avons déjà en grande partie répondu.

²¹⁵ Tout en reprochant à Michael Detlefsen d'avoir « monté une défense d'arrière garde de Hilbert » [Smith 2007, p. 258].

Il reste la question de savoir quel est le bénéfice d'une telle interprétation sur un plan philosophique. Par rapport à notre interprétation (et l'interprétation standard) du programme de Hilbert, trois choses ont été indubitablement perdues :

- 1) l'objectif d'assurer la sûreté de l'intégralité des mathématiques est réduit à une portion seulement des mathématiques ;
- 2) la force de l'évidence intuitive de la base finitiste retenue par Hilbert pour y parvenir est remplacée par la référence abstraite à un système formel particulier, non précisément défini ;
- 3) la contrainte d'utiliser une démarche « à la Rosser » conduit à l'introduction d'un concept de prouvabilité à la fois contraignant et non intuitif.

En comparaison, le seul gain obtenu est celui de la satisfaction de disposer d'un système formel de l'arithmétique dans lequel une certaine forme de consistance peut être établie. Nous avons le sentiment qu'il s'agit là d'une victoire à la Pyrrhus, et nous pouvons raisonnablement nous demander si cela conduit vraiment à gagner quelque chose, en retrouvant alors le commentaire de David Auerbach [Auerbach 1992, p. 87]. Pour achever cette analyse, il n'est peut-être pas sans intérêt de signaler une remarque de Georg Kreisel et Gaisi Takeuti, qui ont évoqué dès 1974 la possibilité d'une telle démarche en l'associant alors à l'esprit de Wittgenstein₂ [Kreisel, Takeuti, 1974, p. 47-48].

3.5 Le recours à la logique paraconsistante

La proposition d'utiliser la logique paraconsistante sous l'une de ses formes, en vue de s'opposer aux théorèmes de Gödel de 1931²¹⁶, a été suggérée par les tenants de cette logique en suivant trois voies différentes. 1) Tout d'abord, et c'est certainement la proposition la plus ancienne, Graham Priest a proposé un argument en faveur de la logique paraconsistante basée sur l'utilisation du premier théorème de Gödel de 1931 : [Priest 1979], [Priest 1984] et [Priest 1987]. Après être resté longtemps en sommeil, cet argument a été réactivé récemment sans changement : [Berto 2009b]. Reposant sur l'idée d'appliquer ce théorème au cas d'un système formel hypothétique qui formaliserait l'ensemble de la vérité mathématique intuitive,

²¹⁶ La caractéristique fondamentale d'un système formel basé sur la logique paraconsistante est que celui-ci est simplement *in*consistant tout en étant absolument consistant (alors qu'en logique classique les deux caractéristiques sont équivalentes sous la requête faible que le système formel contienne le calcul des propositions). Pour y parvenir, deux voies ont été proposées : syntaxique, par Newton da Costa et sémantique, par Graham Priest. Bien entendu, comme le système formel est simplement inconsistant, les deux théorèmes de Gödel de 1931 ne peuvent s'y appliquer.

l'argument est en fait très proche de celui proposé par John Lucas à propos du mécanisme appliqué à l'esprit humain : [Lucas 1961]. Nous les analyserons donc conjointement (§ 6.5). 2) Ensuite, Jean-Paul van Bendegem et Graham Priest ont explicitement proposé une extension du programme de Hilbert sur la base de la logique paraconsistante : [Bendegem 1993] et [Priest 1994]. Comme cette proposition revient à limiter l'arithmétique en éliminant les indécidables, nous l'évoquerons dans le cadre plus général des solutions visant à réaliser (partiellement) le programme de Hilbert au sens restreint en limitant la portée des mathématiques concernées (§ 4.5). 3) Enfin, récemment, dans un article éclairant, Stewart Shapiro a réalisé l'étude d'un véritable système formel constitué de PA étendu par la logique paraconsistante selon l'interprétation de Priest²¹⁷ : [Shapiro 2002]. Bien que l'auteur ne fasse pas mention du programme de Hilbert dans cette étude, il est intéressant d'en analyser le contenu dans ce cadre, dans la mesure où il s'agit de fait d'une solution qui conduit à contourner les deux théorèmes de Gödel de 1931.

Stewart Shapiro construit un système formel de l'arithmétique du premier ordre basé sur la logique paraconsistante au sens de Priest : PA^* , à partir de trois hypothèses :

- 1) il s'agit d'une extension (non conservative) de PA ; tous les théorèmes de PA sont des théorèmes de PA^* mais non l'inverse ;
- 2) le schéma d'induction de PA est étendu à toutes les formules ajoutées par PA^* ;
- 3) PA^* est simplement inconsistant, absolument consistant et sémantiquement clos.

Il applique alors à PA^* la démarche habituelle qui permet d'établir le premier théorème de Gödel de 1931 et montre que l'on a, dans le cas (particulier) présent, à la fois²¹⁸ : $\vdash_{PA^*} \text{Proof}_{PA^*}(g, \ulcorner G^* \urcorner)$ et $\vdash_{PA^*} \neg \text{Proof}_{PA^*}(g, \ulcorner G^* \urcorner)$ où G^* est l'énoncé de Gödel pour PA^* et g le nombre de Gödel d'une dérivation de G^* dans PA^* . Il s'agit bien entendu d'une contradiction, ce qui n'est pas problématique dans le cadre de PA^* qui est paraconsistant. Ce qui est beaucoup plus problématique est que l'analyse de l'énoncé $\text{Proof}_{PA^*}(g, \ulcorner G^* \urcorner)$ montre que celui-ci est non seulement purement arithmétique, mais qui plus est de type Δ_0 (sans quantificateur non borné) ; comment un tel énoncé pourrait-il être contradictoire ? Stewart Shapiro analyse trois réponses

²¹⁷ La voie retenue par Graham Priest est l'utilisation d'une logique à trois valeurs, la troisième caractérisant la situation (paraconsistante ou paradoxale) d'être à la fois vrai et faux. La caractéristique affichée de cette logique est d'être sémantiquement close : les prédicats de vérité la concernant peuvent être exprimés dans le système formel lui-même, donc sans nécessiter un métalangage pour cela comme en logique classique.

²¹⁸ Nous rappelons (§ 2.2) que $\text{Proof}(m, n)$ est une fonction caractéristique FRP, où n est le nombre de Gödel d'un énoncé φ et m le nombre de Gödel d'une dérivation de φ dans le système formel (quand elle existe), dont la généralisation existentielle conduit à la formule $\text{Prov}(n)$, que nous avons abondamment utilisée dans les démonstrations des deux théorèmes de Gödel de 1931.

possibles. Tout d'abord, si l'on suppose que PA est simplement (et absolument) consistant et que PA* est incorrect (certains de ses théorèmes sont non vrais dans le modèle standard) alors il suffit de retenir la solution : $\text{Proof}_{\text{PA}^*}(g, \ulcorner G^* \urcorner)$ est vrai et $\neg\text{Proof}_{\text{PA}^*}(g, \ulcorner G^* \urcorner)$ n'est pas vrai, pour résoudre le problème. Le prix à payer en est la perte d'intérêt de PA* pour lequel il faut abandonner au moins l'une des trois hypothèses qui ont permis sa construction. Une solution plus téméraire consiste à supposer que PA* est correct mais que PA est simplement inconsistant. L'auteur montre que l'on perd alors les liens naturels entre :

- 1) le fait d'être, respectivement de ne pas être, le code d'une dérivation dans PA* d'un énoncé φ ;
- 2) le fait que $\text{Proof}_{\text{PA}^*}(m, \ulcorner \varphi \urcorner)$, respectivement $\neg\text{Proof}_{\text{PA}^*}(m, \ulcorner \varphi \urcorner)$, soit vrai ;
- 3) la dérivabilité dans PA* de $\text{Proof}_{\text{PA}^*}(m, \ulcorner \varphi \urcorner)$, respectivement de $\neg\text{Proof}_{\text{PA}^*}(m, \ulcorner \varphi \urcorner)$ [Shapiro 2002, p. 826].

Cela se traduit en fait par la perte de l'isomorphisme entre les chaînes de caractères et les nombres naturels, pour des longueurs de celles-là et des valeurs de ceux-ci certes très grandes mais finies²¹⁹. La solution la plus extrême se traduit par la supposition que non seulement PA* est correct et PA simplement inconsistant, mais de plus, qu'une chaîne donnée de caractères peut à la fois être et ne pas être une dérivation légitime dans PA*. Il faut alors appliquer cette dernière position au résultat qu'il peut être montré de manière effective que g est le code d'une dérivation de G^* dans PA*. Nous sommes ici au cœur de l'attitude « dialéthiste »²²⁰. Signalons enfin l'ultime conséquence mentionnée par Stewart Shapiro de cette position : les deux énoncés $\text{Proof}_{\text{PA}^*}(g, \ulcorner G^* \urcorner)$ et $\neg\text{Proof}_{\text{PA}^*}(g, \ulcorner G^* \urcorner)$ sont également dérivables dans PA et même dans Q, qui doit donc être également considéré comme inconsistant. Il est à noter que l'analyse de Stewart Shapiro n'a pas été remise en question par Graham Priest [Berto 2007, p. 102, n. 28]. Face à ces réponses et à l'instar du choix de l'auteur, seule la première citée : la non correction de PA*, nous paraît acceptable²²¹. Il en résulte que le recours à la logique paraconsistante n'apparaît pas

²¹⁹ A la réserve près, non signalée dans [Shapiro 2002], que cette finitude découle elle-même de l'application de la démonstration du premier théorème de Gödel de 1931, qui fait implicitement appel à une longueur de caractère *infinie*, via la démarche de diagonalisation.

²²⁰ Selon Graham Priest lui-même [Priest 2000, p. 227-228], il convient de distinguer au moins deux conceptions de la logique paraconsistante (l'auteur en distingue en fait trois). La dernière est la « paraconsistance forte », nommée dialéthisme par l'auteur (terme forgé par lui mais qui n'est utilisé que dans le milieu des spécialistes de la logique paraconsistante). Elle se caractérise par la conviction qu'il existe de véritables contradictions (dans le monde réel). La première est la « paraconsistance faible », qui se caractérise par la conviction que les contradictions ne sont qu'apparentes et (ou) contingentes, liées aux imperfections de nos outils d'investigation et d'analyse.

²²¹ A noter un renforcement de ce choix par un résultat de Joachim Bromand : [Bromand 2002], publié exactement en même temps que [Shapiro 2002] et qui montre que l'affirmation de la clôture sémantique d'un système formel analogue à PA* est injustifiée : cette hypothèse conduit en fait à l'inconsistance absolue du

une solution crédible pour se soustraire à l'emprise des deux théorèmes de Gödel de 1931. Deux échappatoires, proposées par les tenants de la logique paraconsistante, sont cependant évoquées dans [Shapiro 2002] : 1) l'affirmation selon laquelle notre intuition concernant les nombres naturels ne serait plus applicable dès que ceux-ci sont très grands (ce qui est évidemment le cas de G), 2) l'idée que les énoncés indécidables comme celui de Gödel ne sont pas de véritables énoncés arithmétiques, ce qui autoriserait qu'ils puissent avoir des propriétés (en l'occurrence l'inconsistance simple) que n'ont pas les « véritables » énoncés arithmétiques. Ces voies en recouvrent en fait d'autres, qui ont été proposées par ailleurs et qui seront analysées dans le cadre plus large des solutions qui ont été envisagées pour réaliser au moins partiellement le programme de Hilbert au sens restreint, en limitant la portée des mathématiques concernées (§ 4.5).

Si nous faisons le bilan de l'analyse effectuée des différentes tentatives de dépassement du finitisme de Hilbert pour en sauver le programme au sens restreint, il ressort d'abord qu'aucune des solutions « traditionnelles » ([Gödel 1933b], [Gentzen 1936] et [Gödel 1958]) ne peut raisonnablement y prétendre, ce qui rejoint une remarque de Stephen Simpson [Simpson 1988, p. 353]. Par contre, la solution de Gentzen de 1936 *limitée à PRA* conduit à ce qui peut sans doute être considéré comme la limite supérieure d'une extension acceptable du finitisme. Cette solution permet alors d'obtenir une démonstration « finitiste » (en un sens étendu : au-delà de PRA) de la consistance de PRA, en cohérence avec l'interprétation du Programme que nous avons retenue. Cela n'autorise par contre nullement à aller au-delà, et en particulier cela ne permet pas de justifier sur la même base la consistance de PA (a fortiori de Z_2) et ne suffit donc pas pour sauver ce programme. L'introduction par Hilbert de la règle- ω , successivement en 1930 et 1931, est encore moins satisfaisante, en perdant tout l'intérêt épistémologique de simplicité et d'évidence de la base finitiste qui constituaient l'attrait du finitisme originel et en s'interdisant d'atteindre les objectifs du Programme. De fait, malgré les différentes tentatives de Bernays, l'école de Hilbert n'a pu fournir aucune justification philosophique sérieuse des différentes extensions du finitisme que nous avons analysées. Concernant les solutions plus récentes, l'inépuisabilité des mathématiques présentée par Torkel Franzen n'est pas utilisable comme solution du programme de Hilbert. Michael Detlefsen est quant à lui effectivement parvenu à proposer une interprétation cohérente du programme de Hilbert au sens restreint, qui permet de s'affranchir des deux théorèmes de Gödel de 1931. Cependant, dans le cadre de cette

système formel paraconsistant, ce qui est évidemment incompatible avec le fondement de la logique paraconsistante.

interprétation, il ne reste alors quasiment plus rien des traits fondamentaux du Programme originel et les objectifs de Hilbert ne peuvent donc être atteints. Enfin, la solution radicale de la logique paraconsistante qui consiste à considérer que PA et Q sont simplement inconsistants (mais absolument consistants), ne peut pas être sérieusement retenue. Il en ressort finalement qu'aucune des différentes extensions du finitisme de Hilbert qui ont été proposées ne peut prétendre sauver le Programme. On ne peut donc espérer au mieux qu'une réalisation partielle de celui-ci, en limitant alors le domaine des mathématiques concernées. Nous allons analyser maintenant les différentes voies qui ont été effectivement proposées en ce sens.

4 Peut-on réaliser partiellement le programme de Hilbert au sens restreint, en acceptant une réduction de la portée des mathématiques concernées ?

A défaut de pouvoir atteindre l'objectif du programme de Hilbert au sens restreint, même en acceptant d'en étendre le finitisme initial, se pose légitimement la question de savoir s'il est au moins possible d'en obtenir une réalisation *partielle*, pour un domaine limité des mathématiques, tout en prenant en compte les deux théorèmes de Gödel de 1931. Bien entendu, le fait d'accepter de ne pas pouvoir rendre compte de l'ensemble des mathématiques à partir d'une base finitiste, atténue grandement l'intérêt même du Programme tel que l'avait conçu Hilbert ; nous y avons suffisamment insisté au § 1 pour ne pas y revenir. Cette réserve importante étant émise, on peut dire qu'essentiellement deux voies ont été adoptées en vue de proposer une réponse positive à la question posée, avec de nombreuses variantes. La première est d'ordre mathématique et conduit à déterminer, pour tout domaine mathématique donné, les systèmes formels les plus forts, conservatifs par rapport à un système formel de base fixé qui est typiquement PRA ou PA : c'est la démarche des « reverse mathematics », apparue au milieu des années 1970. Le caractère partiel de réalisation du programme de Hilbert y sera alors défini, d'une part par l'étendue du domaine mathématique dont il est possible de rendre compte et d'autre part par l'écart de la puissance du système formel de départ à celle de celui qui peut être rattaché au finitisme originel, dont on devine qu'il va s'agir de deux caractéristiques antagonistes. La seconde voie est d'ordre philosophique et elle est plus radicale. Elle consiste à tenter de justifier que les énoncés indécidables issus des théorèmes de Gödel 1931 ne sont pas des énoncés véritablement arithmétiques et que pour ces derniers un système formel comme PA est en fait syntaxiquement complet : c'est la thèse de Daniel Isaacson dans [Isaacson 1987] et [Isaacson 1992]. Une telle position doit alors tenir compte non seulement des théorèmes de Gödel proprement dits, mais aussi de différents résultats également obtenus à partir du milieu des années 1970, montrant que certains énoncés *combinatoires finis* sont décidables dans Z_2 et indécidables dans PA. Dans ce cadre, il est à remarquer que la question cruciale de savoir si de tels énoncés sont ou ne sont pas véritablement « arithmétiques », a conduit les deux grandes figures que sont Solomon Feferman et Harvey Friedman à s'opposer sur ce point²²². Il convient

²²² Cette controverse intervient en fait dans un contexte quelque peu différent : sur la nécessité ou non « de nouveaux axiomes » (de nouveaux systèmes formels). Le débat porte ici sur la question de l'intérêt ou non d'étendre ZF (le système formel de Zermelo-Fraenkel) ou ZFC (lorsque l'axiome de choix lui est ajouté) et il

également d'évoquer le cas particulier des différentes propositions qui ont été faites en faveur d'un ultrafinitisme, c'est-à-dire d'un finitisme borné qui clôt la circonscription des différentes formes de finitisme et de constructivisme que nous avons proposée (Tableau III, § 3). Nous y incluons alors le cas extrême de la solution paraconsistante, qui peut naturellement lui être rattachée. Nous allons cependant d'abord présenter une synthèse de l'état de réalisation du Programme original de Hilbert à la fin de l'année 1930 (donc juste avant l'irruption des théorèmes de Gödel de 1931)²²³, celui-ci étant essentiellement soutenu par les théorèmes d'Ackermann de 1924, de von Neumann de 1927²²⁴ et de Herbrand de 1930.

4.1 Les résultats partiels obtenus avant 1931

La première démonstration antérieure à 1931 de la consistance simple d'une partie non triviale des mathématiques, revendiquée comme finitiste, est due à Ackermann. La démonstration initiale de celui-ci date de 1924²²⁵ ; elle a toutefois été reprise deux fois, d'abord en 1927²²⁶, à partir de la démonstration de von Neumann de la même année²²⁷, puis en 1940²²⁸, à partir de la démonstration de Gentzen de 1936. Le point important est que l'auteur, et à sa suite Bernays et Hilbert, s'est mépris en surestimant la portée réelle du domaine d'application de cette démonstration, ce qui explique les déclarations exagérément optimistes de Hilbert [Hilbert 1929 (1992), p. 179-180] et [Hilbert 1931 (1992), p. 193]. Une analyse détaillée de cette erreur et de ses corrections successives a été effectuée par Richard Zach dans [Zach 2003a]. La situation, aujourd'hui bien connue et non controversée, peut être résumée de la manière suivante : en 1924 Ackermann était persuadé que sa démonstration s'appliquait à un système formel ayant le langage de PRA mais du 2^{ème} ordre ; en 1927 il réalise que le domaine d'application ne dépasse pas le 1^{er} ordre, mais Hilbert est persuadé qu'il couvre quand même

dépasse clairement notre préoccupation. Nous nous limiterons donc dans un tel débat, à ce qui concerne les énoncés *arithmétiques* (exprimables dans PA).

²²³ Nous renvoyons au tableau I (§ 1.4) pour le bilan des systèmes formels du 1^{er} ordre plus faibles que PA et PRA, pour lesquels les théorèmes de Gödel de 1931 ne *s'appliquent pas*.

²²⁴ Cette synthèse est d'autant plus nécessaire dans le contexte de notre discussion, qu'il y a eu historiquement une erreur d'interprétation sur la portée de ces deux théorèmes (le fait est aujourd'hui bien connu), qui n'a été reconnue que tardivement par l'école de Hilbert [Hilbert, Bernays 1934 (2001) p. 459, 461].

²²⁵ « Begründung des « tertium non datur » mittels der Hilbertschen theorie der widerspruchsfreiheit », *Mathematische Annalen*, vol. 93, p. 1-36.

²²⁶ Texte non publié mais transmis à Bernays [Zach 2003a, p. 241-242].

²²⁷ « Zur Hilbertschen Beweistheorie », *Mathematische Zeitschrift*, vol. 26, p. 1-46.

²²⁸ « Zur Widerspruchsfreiheit der Zahlentheorie », *Mathematische Annalen*, vol. 117, p. 162-194.

l'ensemble : Z_1 (PA) de l'arithmétique du 1^{er} ordre ; en 1931 von Neumann met en évidence un contre-exemple à la démonstration (celle modifiée de 1927) ; enfin en 1932, Bernays prend conscience des limites de la preuve d'Ackermann [Parsons 2003b, p. 29-30] et en 1934 il montre que la démonstration ne porte en fait que sur un faible sous-système de Z_1 , où le schéma d'induction est réduit à des formules sans variables liées²²⁹. Cette explication permet de rendre compte de la contradiction apparente entre la démonstration d'Ackermann telle qu'elle est perçue en 1931 et le second théorème de Gödel de 1931, qui s'y oppose.

Au-delà du problème historique de l'évolution de la perception de la portée de la démonstration d'Ackermann, se pose la question de son caractère finitiste. Ce point a également été abordé par Richard Zach dans [Zach 2003a]. Il apparaît que la démonstration de 1924 (fondée sur la technique de « l'élimination des ε ») utilise l'induction transfinitie jusqu'à ω^{ω^ω} , ce qui pose évidemment le problème de sa justification finitiste. Comme le remarque Richard Zach, nous sommes conduits à une alternative :

Le fait que l'induction transfinitie dans la forme utilisée par Ackermann fut si facilement acceptée comme finitiste, non seulement par Ackermann mais aussi par Hilbert et Bernays, laisse ouvertes deux possibilités : ou ils étaient simplement dans l'erreur [...], ou la vision commune de ce que Hilbert pensait que devait être le point de vue finitiste doit être révisée. [Zach 2003a, p. 244]

Ce que montre Richard Zach est qu'Ackermann n'a pris conscience que bien après 1924 (en 1938) qu'il avait utilisé en fait une induction transfinitie [Ackermann 1938, cité dans Zach 2003a, p. 235]. Mais en opposition à Richard Zach, nous en déduisons un argument fort en faveur de la première branche de l'alternative : la non perception (jusqu'en 1930), par Hilbert et ses élèves, du caractère non véritablement finitiste de la démonstration de 1924, contre la seconde branche : l'interprétation (avant 1931), par Hilbert et ses élèves, du finitisme de manière élargie (ou libérale) par rapport à la conception classique. A l'appui de cet argument, on peut noter la circonscription non ambiguë du finitisme d'une démonstration, donnée par Weyl en 1929, dont on peut raisonnablement penser qu'elle traduit une conception partagée par l'école de Hilbert à cette époque :

Ce qu'il faut prouver, c'est qu'on ne se meut pas dans un cercle, et qu'après un nombre d'étapes correctives successives *qui peut être indiqué d'avance*, on parvient à une réduction qui n'échoue en aucun

²²⁹ La démonstration s'applique en particulier au système de Presburger (PA sans l'opérateur représentant la multiplication sur les entiers) [Hilbert, Bernays 1934 (2001), p. 455].

point de l'ensemble *fini* donné d'axiomes *s.* [...] J. von Neumann [1927] a démontré ce théorème de finitude. [Weyl 1929 (1994), p. 178-179 ; souligné par nous]

Une précision mentionnée dans [Tait 2010] confirme cette analyse. William Tait souligne que dans son article de 1924, Ackermann ne donne qu'une justification de l'induction jusqu'à ω^2 [Tait 2010, p. 92]. Cela, associé à la remarque précédente mentionnée par Ackermann en 1938, conduit légitimement à considérer qu'en 1924, Ackermann et Hilbert étaient persuadés que la démonstration ne mettait en jeu qu'une induction jusqu'à ω^2 , ce qui pouvait (à la rigueur) être considéré comme finitiste, sans pour autant supposer que le finitisme devait s'étendre jusqu'à ω^{ω^ω} (ω à la puissance ω à la puissance ω). Bien entendu, cette analyse ne s'oppose pas à l'évolution indéniable mais *ultérieure* du finitisme par Hilbert et son école : implicitement avec l'utilisation de la règle- ω et explicitement dans son introduction à [Hilbert, Bernays 1934], déjà citée au début du § 3.

La seconde démonstration antérieure à 1931 de la consistance simple d'une partie non triviale des mathématiques, revendiquée comme finitiste, est due à Herbrand : [Herbrand 1930] et [Herbrand 1931a]. Sa portée est très proche de celle d'Ackermann dont il vient d'être question : ici aussi le schéma d'induction est réduit à des formules sans variables liées²³⁰. Mais reposant sur une démarche complètement différente de celle d'Ackermann, son caractère finitiste y est moins ambigu (dans la limite de l'absence de circonscription précise du finitisme²³¹). En particulier, Herbrand récuse explicitement l'usage par Löwenheim et Skolem dans le théorème éponyme, des notions de vérité dans un modèle et de satisfiabilité, en exigeant au contraire la restriction à des démonstrations finitistes²³², à tel point que lorsqu'il utilise le terme : « vrai », il faut comprendre : « prouvable dans le système formel considéré » [Heijenoort 1967c, p. 526]. L'ambiguïté sur le contenu du finitisme évoquée à propos d'Ackermann n'est cependant pas totalement absente des écrits de Herbrand. En effet dans [Herbrand 1931b], en définissant la théorie prise en compte, l'auteur introduit le « groupe D » où apparaît explicitement la règle- ω de Hilbert. Même si ce « groupe D » est clairement distingué des autres groupes « A, B, C » qui correspondent à l'arithmétique classique (et au système considéré dans [Herbrand 1930]), à aucun moment Herbrand n'introduit de discussion sur l'évolution du

²³⁰ Mais les fonctions autorisées incluent la fonction d'Ackermann, donc les FRG.

²³¹ Ce qui se traduit par l'inclusion des FRG, considérées comme finitistes, ce que nous avons déjà discuté au § 1.2.

²³² Nous rappelons un point que nous avons déjà signalé : il est classiquement admis que jusqu'en 1933, l'école de Hilbert assimilait purement et simplement le finitisme de Hilbert à l'intuitionnisme de Brouwer (nonobstant la nuance apportée par [Tait 2006b] que nous avons mentionnée au § 3.1).

finitisme ainsi introduite. Par ailleurs, dans la mesure où l'article a été rédigé au moins en partie après que l'auteur ait pris connaissance des résultats de Gödel de 1931 (il y fait explicitement allusion), il n'est pas possible de savoir dans quelle mesure ces résultats ont influencé l'introduction du groupe D et donc l'évolution du contenu du finitisme²³³. Au passage, sur la question de savoir dans quelle mesure Herbrand était conscient des limites des possibilités d'extension de sa propre démonstration de 1930, on notera avec un sourire a posteriori, sa propre formulation de l'époque : « Il [le théorème général de 1930] permet d'abord de démontrer la non-contradiction de tous les axiomes de l'arithmétique, *avec cependant encore quelques petites restrictions* » [Herbrand 1931a (1968), p. 213 ; souligné par nous ; le commentaire entre crochets est nôtre].

4.2 Les résultats modernes obtenus dans le cadre des « reverse mathematics »

La locution : « reverse mathematics » désigne un programme de travail mathématique proposé par Harvey Friedman en 1974 : [Friedman 1975], qui s'est traduit par un grand nombre de résultats obtenus principalement durant les années 1970 et 1980. Son objectif consiste à déterminer, pour tout théorème mathématique donné, quel est le système formel *le plus faible* qui permette de le dériver (d'où le terme de « reverse mathematics »). En pratique, ce programme a conduit à proposer un nombre restreint de systèmes formels particuliers, qui sont tous des sous-systèmes de l'arithmétique du second ordre Z_2 et qui sont conservatifs par rapport à PRA ou PA au moins pour les énoncés Π^0_1 , tout en permettant de rendre compte d'une grande partie des mathématiques bien au-delà de l'arithmétique (voir le tableau IV ci-après). Il s'agit donc indiscutablement d'une démarche réductionniste en mathématique, qui n'a pas d'objectif fondationnel mais que l'on peut *aussi* associer à une démarche fondationnelle et à une réalisation (partielle) du programme de Hilbert au sens restreint, comme l'ont souligné Stephen Simpson [Simpson 1988, p. 356-357] et Wolfram Pohlers [Pohlers 1998, p. 210].

²³³ On se retrouve donc dans une situation d'incertitude historique analogue à celle évoquée au § 3.2 à propos de l'introduction de la règle- ω par Hilbert.

Tableau IV : Bilan des caractéristiques des principaux sous-systèmes formels de Z_2 obtenus dans le cadre des « reverse mathematics » et par Solomon Feferman

Nom	Auteur	Contenu	Ordinal limite	Conservativité	Capacité (exemples)
RCA_0^*	[Simpson, Smith 1986]	EA + Schème de compréhension (CA) ²³⁴ Δ_1^0	ω^3	EA / Π_2^0	?
$RCA_0 \equiv \Delta_1^0 - CA_0$	[Parsons 1970]	PRA + Schème de compréhension (CA) Δ_1^0 + Schème d'induction (IA) Σ_1^0	ω^ω	PRA / Π_2^0	Prouver l'existence de tous les ensembles récursifs
WKL_0^*	[Simpson, Smith 1986]	RCA_0^* + Lemme faible de König	ω^3 ou ω^ω ?	EA / Π_2^0 RCA_0^* / Π_1^1	?
WKL_0	[Friedman 1976]	RCA_0 + Lemme faible de König RCA_0 + Schème de séparation ²³⁵ Σ_1^0	ω^ω	PRA / Π_2^0 RCA_0 / Π_1^1	Rendre compte de 85 % des mathématiques [Simpson 1988] Dériver le théorème de complétude de Gödel de 1930
WKL_0^+	[Simpson 1988]	WKL_0 + Schème additionnel ²³⁶	ω^ω ou ω^{ω^ω} ?	PRA / Π_2^0 RCA_0 / Π_1^1	?
$ACA_0 \equiv \Delta_1^1 - CA_0$	[Friedman 1976]	RCA_0 + Schème de compréhension (CA) Δ_1^1	ε_0	PA / Δ_1^1 HA / Π_2^0	Effectuer l'analyse prédictive de Weyl
W	[Feferman 1988a]	Théorie formelle des types	ε_0 ou Γ_0 ?	PA / Δ_1^1	Rendre compte de 99 % des mathématiques [Feferman 2000a]
ATR_0	[Friedman 1976]	RCA_0 + Schème de séparation Σ_1^1	Γ_0 ²³⁷	IR (Feferman) / Π_1^1	Dériver la consistance de ACA_0
$\Pi_1^1 - CA_0$	[Friedman 1976]	RCA_0 + Schème de compréhension (CA) Π_1^1	$\theta\Omega\omega^0$ ₂₃₈	?	Dériver la forme finie du théorème de Kruskal ²³⁹

²³⁴ Le schème de compréhension Δ_1^0 est le suivant : $\vdash \forall x (\varphi(x) \leftrightarrow \psi(x)) \rightarrow \exists Y \forall x (x \in Y \leftrightarrow \varphi(x))$, où φ et ψ sont respectivement des énoncés Σ_1^0 et Π_1^0 .

²³⁵ La définition du schème de séparation est donnée par Stephen Simpson [Simpson 2009, p. 41].

²³⁶ La définition de ce schème additionnel est donnée par Stephen Simpson [Simpson 1988, p. 355].

²³⁷ ATR_0 est prédictif au sens de Feferman : l'ensemble infini des entiers naturels ω est pris comme une totalité, mais les univers successifs constitués à partir des sous-ensembles de ω sont construits selon un processus de définissabilité itérée.

²³⁸ Selon la notation de Kurt Schütte dans « Proof theory », Springer-Verlag, 1977 ; $\theta\Omega\omega^0$ (quoi qu'il recouvre exactement) étant situé bien au-delà de Γ_0 , $\Pi_1^1 - CA_0$ est clairement imprédictif au sens de Feferman (il se situe au-delà de notre tableau III du § 3).

²³⁹ Ce point est abordé au § 4.3.

Il est assez naturel de rattacher à ce programme les travaux menés par Solomon Feferman et ses collaborateurs, dont les premiers éléments sont antérieurs : [Feferman 1964] et [Feferman 1968]. Ceux-ci présentent en commun avec les « reverse mathematics » un caractère réductionniste et même instrumentaliste explicitement affiché [Feferman 1999, p. 109] et [Feferman 2000a, p. 410] et s'en distinguent essentiellement par le recours à un système formel de référence qui n'est prédicatif qu'en un sens faible. Comme nous l'avons déjà signalé, l'ensemble de ces réflexions s'inscrit par ailleurs dans le cadre plus général d'un débat en cours en philosophie des mathématiques, sur l'attitude à adopter à l'égard des systèmes formels ensemblistes forts²⁴⁰. Ce débat oppose typiquement deux approches radicalement différentes en philosophie des mathématiques :

1) se laisser toute latitude d'utiliser les systèmes formels ensemblistes les plus forts, y compris pour démontrer par exemple des énoncés simples d'arithmétique, ce qui revient implicitement à adopter une ontologie très riche avec évidemment un risque que ces systèmes formels s'avèrent contradictoires ;

2) chercher au contraire à contenir au « juste nécessaire » le système formel à utiliser en fonction du domaine mathématique considéré, en limitant le plus possible l'ontologie sous-jacente et le risque associé à l'utilisation d'un système formel qui pourrait s'avérer contradictoire.

Le représentant le plus éminent de la première voie est bien évidemment Gödel lui-même, sur la base de sa position réaliste et de sa propre interprétation des théorèmes de 1931²⁴¹, telle qu'elle est exprimée dans la note 48a (de l'original) :

Comme on le montrera dans la seconde partie de ce travail, la raison véritable de l'incomplétude inhérente à tous les systèmes mathématiques est que la formation des types plus élevés peut être poursuivie de manière transfinie (voir *Hilbert 1926*) tandis que dans tout système formel on ne peut en trouver au plus qu'une quantité dénombrable. On peut en effet montrer que les propositions indécidables construites ici deviennent décidables dès lors qu'on ajoute des types supérieurs appropriés (par exemple, le type ω au système P). On retrouve la même situation s'agissant du système d'axiomes de la théorie des ensembles. [Gödel 1931 (1989), p. 132, n. 3]

Cela a conduit Solomon Feferman à introduire l'expression : « programme de Gödel » (implicitement opposé au programme de Hilbert) pour désigner cette tendance [Feferman 1999, p. 99]. Comme nous l'avons déjà indiqué, Harvey Friedman s'est explicitement rattaché à ce

²⁴⁰ Ce débat est attaché à la question de la nécessité de « nouveaux axiomes » : [Feferman 2000a]. Comme nous l'avons déjà indiqué, en incluant des préoccupations purement ensemblistes, il dépasse le cadre arithmétique auquel nous nous limitons.

²⁴¹ Une telle interprétation est explicitement contestée par Harvey Friedman [Friedman 2000, p. 443].

courant [Friedman 1998, p. 804]. En sens inverse, Solomon Feferman ainsi que Stephen Simpson s'inscrivent dans le cadre de la seconde voie.

Dans le contexte du programme de Hilbert au sens restreint, le résultat majeur issu des « reverse mathematics » se rapporte au système formel WKL_0 , qui repose sur les axiomes de PRA complétés, du schème de compréhension pour les énoncés Δ^0_1 , du schème d'induction étendu aux énoncés Σ^0_1 , et du lemme faible de König (Tableau IV). WKL_0 est conservatif par rapport à PRA pour les énoncés Π^0_2 (donc a fortiori Π^0_1) et ce résultat est *prouvable dans PRA*. Cette dernière précision est évidemment cruciale pour notre discussion, comme le font remarquer Panu Raatikainen [Raatikainen 2003, p. 170] et John Burgess [Burgess 2010, p. 136], même si ce point ne semble pas avoir été souligné par les praticiens des « reverse mathematics »²⁴². Le caractère remarquable de WKL_0 est qu'une grande partie de l'algèbre et de l'analyse peut y être effectuée (au-delà du fait que le théorème de complétude de Gödel 1930 y est dérivable) : « J'estimerais qu'au moins 85 % des mathématiques existantes peuvent être formalisées dans WKL_0 ou WKL_0^+ ou dans des systèmes plus forts qui sont conservatifs par rapport à PRA pour les énoncés Π^0_2 » [Simpson 1988, p. 361]. C'est donc la réalisation partielle (à 85 % si l'on en croit Stephen Simpson) du Programme de Hilbert, sachant que si la consistance de PRA ne peut donc pas être établie dans le cadre de l'interprétation standard de la thèse de Tait, elle peut l'être dans le cadre de notre interprétation. D'une manière un peu différente, John Burgess fait remarquer que le résultat technique obtenu est en fait un peu plus fort, à savoir la conservativité de WKL_0^* par rapport à EA de notre tableau I (§ 1.4) [Simpson, Smith 1986]. Or EA est un fragment de PRA dont la limite supérieure de l'ordinal accessible est seulement ω^3 (au lieu de ω^ω pour PRA). Cela permet ainsi de limiter encore l'extension nécessaire du finitisme standard (associé à l'ordinal limite : ω) permettant d'assurer la consistance de EA. John Burgess est même plus audacieux en affirmant (dans le cadre de l'interprétation standard de la thèse de Tait) :

Ces résultats *sont* des réalisations partielles du programme parce que, tandis qu'un finitiste ne peut pas savoir que tout ce qui est prouvable dans PRA est finitistement prouvable, un finitiste *peut* savoir que tout ce qui est prouvable dans un fragment borné de PRA comme EFA [EA] est finitistement prouvable. Ce

²⁴² La première démonstration du caractère conservatif de WKL_0 par rapport à PRA : [Friedman 1976], faisait appel à des moyens de preuves bien supérieurs à PRA. La remarque attribuée à Harvey Friedman par Stephen Simpson, selon laquelle les moyens de démonstration utilisés ne dépassent pas ceux requis par le résultat de Gentzen de 1936 [Simpson 1985c, p. 144], ne répond évidemment pas à cette critique. Ce n'est que dans [Sieg 1985a] que l'on trouve une démonstration limitée aux seuls moyens de PRA.

fait positif est l'autre face du fait négatif que les fragments bornés n'épuisent pas la prouvabilité finitiste comme la prouvabilité de PRA le fait (selon la thèse de Tait).

[Burgess 2010, p. 139 ; le commentaire entre crochets est nôtre]

Il est également intéressant de signaler le second résultat remarquable obtenu par les « reverse mathematics ». Ce résultat se rapporte au système formel ACA_0 , plus fort que WKL_0 , sans le lemme faible de König mais avec un schème de compréhension étendu aux énoncés Π^1_0 . ACA_0 est conservatif par rapport à PA pour tous les énoncés arithmétiques et par rapport à HA pour les énoncés Π^0_2 . Bien entendu, encore plus de mathématiques peuvent y être effectuées²⁴³, ce qui a conduit à la conjecture de Feferman : « [...], j'ai été amené à conjecturer que pratiquement toutes les mathématiques scientifiquement applicables peuvent être formalisées dans des systèmes réductibles à PA, [...] » [Feferman 1999, p. 109]. Ce résultat ne pourrait évidemment être rattaché au programme de Hilbert au sens restreint qu'en acceptant une extension encore plus grande du finitisme de Hilbert : jusqu'à l'intuitionnisme de Brouwer (tableau III, § 3). C'est ce que suggère encore John Burgess²⁴⁴ :

Si nous augmentons le niveau du finitisme à celui du constructivisme, nous avons le résultat que le système nommé ACA_0 est une extension conservative (pour une large classe de formules) de l'arithmétique de Heyting HA. Ce résultat *permet* au constructiviste de conclure, étant donné la preuve d'un résultat (exprimable par une formule de cette classe) dans ACA_0 , que le résultat en question est constructivement prouvable. Cela provient du fait que le constructiviste peut savoir que tout ce qui est prouvable dans HA est constructivement prouvable. De nouveau, ce résultat positif est l'autre face d'un fait négatif, le fait que HA n'épuise pas la prouvabilité constructiviste. [Burgess 2010, p. 139]

Nous avons cependant montré qu'une telle position ne pouvait raisonnablement pas être rattachée au finitisme de Hilbert (§ 3.1). C'est évidemment encore plus vrai des systèmes plus forts, proposés dans le cadre des « reverse mathematics » et par Solomon Feferman, qui sont imprédicatifs (au sens de Feferman). Ces différents systèmes formels constituent cependant des réalisations tout à fait remarquables d'instrumentalisme en mathématique. Sur un plan philosophique, la clôture du débat entre Programme de Hilbert et « programme de Gödel » reste néanmoins suspendue à l'éventualité qu'un théorème d'arithmétique puisse être démontré dans un système formel fort (ensembliste) sans pouvoir l'être dans PRA (a fortiori dans PA). Cette

²⁴³ 99 % selon Solomon Feferman [Feferman 2000a, p. 411].

²⁴⁴ Dont le rattachement des systèmes formels des « reverse mathematics » aux niveaux de finitisme et de constructivisme est analogue au nôtre, et recouvre partiellement celui proposé par Stephen Simpson [Simpson 2009, Tableau 1, p. 43].

question a conduit à la conjecture de Friedman (1999), qui récusé une telle éventualité : « Tout théorème publié dans les Annals of Mathematics, dont le sujet concerne seulement les objets mathématiques finitaires (c'est-à-dire ce que les logiciens appellent les formules arithmétiques), peut être prouvé dans l'Arithmétique Élémentaire [EA] » [Avigad 2003, p. 258 ; le commentaire entre crochets est nôtre].

4.3 Les énoncés combinatoires finis, théorèmes de Z_2 et indécidables dans PA

A la fin des années 1970, plusieurs énoncés combinatoires finis exprimables dans PA, à commencer par celui de [Paris, Harrington 1977], ont été montrés indécidables dans PA ou dans des systèmes formels plus forts, tout en étant dérivables dans Z_2 ou dans des systèmes formels plus faibles. Ces résultats ont alors été présentés comme un témoignage en faveur du fait que l'indécidabilité arithmétique se rencontrait aussi sur de véritables énoncés mathématiques, et pas seulement sur des codages de propriétés métathéoriques à l'instar des théorèmes de Gödel de 1931 [Paris, Harrington 1977, p. 1133, note de l'éditeur : Jon Barwise]. Si cette conclusion est avérée, il en résulte l'impossibilité d'un instrumentalisme hilbertien « à la Feferman » concernant PA : [Feferman 2000a], soit une limitation extrêmement forte des possibilités de réalisation partielle du Programme de Hilbert, comme l'a souligné Jacques Dubucs [Blanché, Dubucs 1996, p. 378]. On peut cependant noter que des réserves ont également été émises très tôt contre une telle présentation, ainsi que cela apparaît dans un sarcasme de Craig Smorynski :

Cette indépendance [Paris, Harrington 1977] étant d'un caractère différent de celle de l'énoncé indécidable de Gödel ou de celles associées aux résultats d'indépendance en théorie des ensembles, le résultat a été beaucoup vendu par les logiciens recherchant l'acceptation sociale du reste du monde mathématique, comme quelque chose de très proche de la venue du millénaire.

[Smorynski 1982, p. 182 ; le commentaire entre crochets est nôtre]

L'importance de cette conséquence dans le cadre de notre débat justifie que nous analysions plus en détail ces énoncés et en particulier le ressort de leur indécidabilité. Ces énoncés sont essentiellement au nombre de cinq et le niveau précis de leur indécidabilité est indiqué dans le tableau V.

Tableau V : Principaux énoncés combinatoires finis, théorèmes de Z_2 indécidables dans PA

Énoncé ²⁴⁵	Forme finitaire du théorème de	Non dérivable dans	Dérivable dans
FR : [Paris, Harrington 1977]	Ramsey	PA	ATR ₀
FG : [Kirby, Paris 1982]	Goodstein	PA	ATR ₀
FGP : [Friedman, McAloon, Simpson 1982]	Galvin-Prickry	ATR ₀	Π^1_1 -CA ₀
FK : Friedman 1982 ²⁴⁶ [Simpson 1985b]	Kruskal	ATR ₀	Π^1_1 -CA ₀
FGK : Friedman 1982 [Simpson 1985b]	Kruskal généralisé	Π^1_1 -CA ₀	Z_2

Ils présentent tous la caractéristique commune d'être des énoncés combinatoires finis de type Π^0_2 , formes finies de théorèmes (infinis) déjà connus. Il existe également une famille de théorèmes concernant aussi des énoncés combinatoires finis, encore dus à Friedman, qui présentent la particularité commune que leur preuve requiert des axiomes d'existence ensemblistes de grands cardinaux (cardinaux de Malho) : [Friedman, 1998].

La première chose que l'on peut noter est que sous une apparence naturelle en mathématique, ces énoncés ont tous été « préparés exprès » en vue d'obtenir l'indécidabilité dans PA. Comme le remarque Solomon Feferman, aucun de ces énoncés ne correspond donc à un énoncé mathématique au sens « classique » [Feferman 1999, p. 108] et [Feferman 2000a, p. 406]. Surtout, il s'avère que tous ces énoncés, y compris les plus récents : [Friedman 1998], sont prouvablement équivalents dans PA à l'énoncé de consistance²⁴⁷ (plus exactement de 1-consistance) d'un certain système formel : PA lui-même pour FR et FG, ATR₀ pour FGP et FK, Π^1_1 -CA₀ pour FGK ainsi qu'un système beaucoup plus fort pour [Friedman 1998]²⁴⁸ [Feferman

²⁴⁵ Nous avons retenu le mode d'abréviation proposé par Solomon Feferman dans [Feferman 1999] (où FGP signifie : « Finite Galvin-Prickry »), qui nous a paru plus mnémotechnique que celui proposé par Craig Smorynski dans [Smorynski 1982] (où FFF y signifie : « Friedman Finite Forme »), outre l'ambiguïté de savoir en l'occurrence si FFF désigne la forme finie, du théorème de Kruskal ou bien du théorème de Kruskal généralisé. Nous l'avons étendu aux quatre autres théorèmes.

²⁴⁶ Résultats de Friedman obtenus en 1982, mais dont la première présentation succincte a seulement été publiée en 1985 dans [Simpson 1985b], sachant qu'une présentation encore plus succincte, très didactique, se trouve dans [Smorynski 1982].

²⁴⁷ La cause de l'indécidabilité est en l'occurrence à rechercher du côté des limitations des capacités déductives de PA à établir qu'une fonction très rapidement croissante est récursive. La fonction de choix qui permettrait de représenter (au sens formel) l'énoncé indécidable dans PA, croît plus vite que toute fonction prouvablement récursive, en relation avec l'ordinal supérieur de PA : ϵ_0 (Roman Murawski est particulièrement éclairant sur ce point [Murawski 1999, p. 303]). En d'autres termes, montrer dans PA que l'énoncé concerné est dérivable nécessiterait que l'on puisse représenter dans PA l'induction transfinie jusqu'à ϵ_0 , ce qui nous ramène à la problématique de la démonstration de [Gentzen 1936].

²⁴⁸ Il est à noter la situation un peu particulière de cet énoncé, qui a été relevée par Solomon Feferman [Feferman 2000a, p. 407]. Tous les autres énoncés sont dérivables dans Z_2 , ce qui conduit à une certaine assurance de leur validité (liée à l'assurance que l'on peut avoir de la 1-consistance de Z_2). En sens inverse, dans le cas de [Friedman 1998], il n'y a pas d'autre élément permettant de s'assurer de la vérité de l'énoncé que celui qui

1999, p. 108] et [Feferman 2000a, p. 406]. Il est bien sûr attendu que la dérivabilité de chacun de ces énoncés indécidables *implique* la consistance de PA ou du système plus fort dans lequel ils sont montrés indécidables, puisque tel est là, comme on s'en doute, le ressort de la démonstration utilisée pour établir leur indécidabilité (via le second théorème de Gödel de 1931). Mais le fait qu'il y ait une *équivalence* entre chaque énoncé et la 1-consistance de ce système formel, est évidemment un résultat beaucoup plus fort. Un tel résultat signifie que les énoncés en question sont en fait encore des énoncés métamathématiques, comme ceux directement issus des deux théorèmes de Gödel de 1931, même si ceux-là se présentent sous une forme mathématique plus naturelle.

4.4 La thèse d'Isaacson

Daniel Isaacson a proposé à partir de 1987 ([Isaacson 1987] et [Isaacson 1992]), la thèse forte suivante : 1) tous les théorèmes de PA sont des vérités (dans le modèle naturel de PA) arithmétiques_I²⁴⁹, 2) tous les énoncés vrais (dans le modèle naturel de PA), exprimables dans PA mais non dérivables dans lui (et possiblement dérivables dans des systèmes formels plus puissants) ne sont pas arithmétiques_I. Il en résulte évidemment que pour les énoncés vrais (dans le modèle naturel de PA) arithmétiques_I²⁵⁰, PA est (syntaxiquement) complet : [Isaacson 1987, p. 204], et ce nonobstant le premier théorème de Gödel de 1931. Dans cette thèse, est bien entendu fondamental le critère définissant ce qu'est un énoncé arithmétique_I. Pour cela, l'auteur se fonde sur une notion de concept arithmétique « d'ordre supérieur », dont il est d'abord nécessaire de cerner le contenu. Commençons par remarquer que la locution : « ordre supérieur », si elle implique bien le dépassement du premier ordre (au sens ordinaire en logique mathématique), ne se limite pas à cela. L'auteur est sans ambiguïté sur ce point :

Ce que je veux dire ici par « ordre supérieur » inclut son usage standard pour la quantification sur les ensembles d'individus, distinguée de la quantification du premier ordre sur les individus eux-mêmes.

consiste à supposer la 1-consistance du système formel, dont l'énoncé lui-même est une représentation : il y a manifestement là un cercle vicieux.

²⁴⁹ Dans tout ce paragraphe, nous utiliserons le terme arithmétique suivi de l'indice : « I », lorsqu'il s'agit de la notion « d'arithmétique » au sens particulier défendu par Daniel Isaacson dans [Isaacson 1987] et [Isaacson 1992].

²⁵⁰ Dans un contexte un peu différent, Hartry Field avait déjà proposé une distinction proche entre les énoncés « ésotériques » (à l'instar de G et du codage de la consistance de la théorie) et les énoncés « non-ésotériques » [Field 1984 (1996), p. 257]. Egalement, Don Fallis parle d'énoncés « pathologiques » pour désigner les codages d'énoncés métamathématiques qui s'opposent aux énoncés véritablement mathématiques [Fallis 1996, p. 266].

Mais je veux aussi inclure dans cette locution quelque chose de la notion d'infinitaire, au sens de présupposant une totalité infinie. [Isaacson 1987, p. 210]

Précisons ensuite que pour un énoncé arithmétique (exprimable dans le langage de l'arithmétique), la notion d'être ou de ne pas être arithmétique_I porte non sur le contenu de l'énoncé, mais sur les conditions d'établissement de sa vérité (au sens naïf ou dans le modèle naturel d'un système formel de l'arithmétique). Il s'agit donc d'une notion *épistémologique*, comme l'indique explicitement l'auteur [Isaacson 1992, p. 95]. Pour cela, le critère retenu est celui d'être en mesure de percevoir la vérité de l'énoncé « directement sur la base de l'expression claire (d'ordre supérieur) de notre saisie de la structure des nombres naturels » [Isaacson 1987, p. 217]. L'auteur ne donne qu'une seule précision sur ce qu'il entend par cette expression, mais qu'il n'exploite pas par la suite en renvoyant à [Gödel 1958] :

La vérité arithmétique implique de manière inhérente des concepts abstraits d'ordre supérieur dans le sens où Gödel en parle ici [Gödel 1958], nommément, ceux qui caractérisent les nombres naturels comme la plus petite structure fermée sous une application biunivoque et contenant un élément qui ne se trouve pas dans la portée de cette application. [Isaacson 1992, p. 99 ; le commentaire entre crochets est nôtre]

La subtilité est que même pour les énoncés arithmétiques_I, la reconnaissance de leur vérité implique l'utilisation de concepts d'ordre supérieur, parce que notre compréhension du concept de nombre naturel est elle-même d'ordre supérieur [Isaacson 1992, p. 108]. Le critère de démarcation entre énoncés arithmétiques_I et non-arithmétiques_I, est alors que la reconnaissance de la vérité des énoncés arithmétiques_I ne fait appel *qu'à* notre compréhension (d'ordre supérieur) de la structure des entiers naturels [Isaacson 1987, p. 217] et [Isaacson 1992, p. 96].

Ce n'est que dans un second temps que l'auteur affirme la coextensivité entre un énoncé arithmétique_I tel qu'il vient d'être défini et les théorèmes de PA [Isaacson 1992, p. 99], ce qui conduit alors à la thèse d'Isaacson proprement dite. Cette thèse nécessite une double justification : 1) il faut disposer d'un système formel au moins aussi fort que PA pour que ses théorèmes permettent de rendre compte de tous les énoncés arithmétiques_I et 2) il n'est pas nécessaire de faire appel à un système formel plus fort que PA pour cela. Concernant le premier point, l'auteur se contente de faire appel à l'évidence [Isaacson 1992, p. 100, 108]. Et ce n'est qu'implicitement qu'il donne, à la fin de [Isaacson 1992], des éléments en faveur de la nécessité de dépasser le système formel plus faible : PRA. Concernant le second point, l'auteur a l'honnêteté de reconnaître les limites de sa justification [Isaacson 1992, p. 100]. Si l'on accepte néanmoins ces éléments de justification, cette thèse présente l'avantage indéniable d'une

clarification de la ligne de démarcation entre énoncés arithmétiques₁ et non-arithmétiques₁, par rapport à la définition initiale. Il apparaît alors un lien direct avec les théorèmes de Gödel de 1931 dans la mesure où les énoncés non-arithmétiques₁ sont ainsi les indécidables de PA, au premier rang desquels se trouve l'énoncé G. L'auteur évoque explicitement ce lien, en précisant que les concepts d'ordre supérieur qui traduisent le caractère non-arithmétique₁ sont cachés dans le codage gödélien [Isaacson 1987, p. 210]. Dans le cas du premier théorème et de l'énoncé G, l'auteur précise que les concepts cachés sont d'une part la prouvabilité dans PA (qui donne le sens de G) et d'autre part la consistance de PA (dont l'hypothèse est nécessaire à l'établissement du théorème) : [Isaacson 1992, p. 96]. Une telle affirmation soulève deux objections : 1) d'un point de vue technique, la nécessité de disposer du sens de l'énoncé G n'est nullement requise pour établir le premier théorème (contrairement au second), 2) comme les énoncés arithmétiques₁ (vrais) sont les théorèmes de PA, cela implique de disposer déjà de la consistance de PA pour ceux-ci. La réponse de l'auteur à la première objection est qu'il fait intervenir la vérité de G et non seulement son indécidabilité [Isaacson 1987, p. 214]. Cela implique alors que l'on dépasse en fait le premier théorème pour faire intervenir, soit le second²⁵¹, soit le théorème de Tarski de 1933-1936. L'auteur n'évoque pas la seconde objection soulevée. Il affirme que la consistance de PA est un concept d'ordre supérieur, mais cela s'applique également aux énoncés arithmétiques₁. On conçoit tout à fait que le codage de la consistance de PA soit un concept d'ordre supérieur non arithmétique₁, mais si un tel codage intervient bien dans le second théorème il n'intervient pas dans le premier, ce qui ne répond donc pas non plus à l'objection. Pour la même raison, la thèse de l'auteur s'applique par contre très naturellement au second théorème de Gödel de 1931. Il en est de même pour les énoncés plus sophistiqués, qui ont l'apparence d'énoncés arithmétiques, tout en étant toujours le codage d'énoncés métamathématiques (en l'occurrence de consistance d'un certain système formel), contre [Smith 2007, p. 205] (avec les exemples du théorème de Matijasevic de 1970 en réponse au 10^{ème} problème de Hilbert²⁵² ainsi que les énoncés combinatoires finis que nous avons évoqués).

L'auteur discute également des liens entre sa circonscription des énoncés arithmétiques₁ et les notions classiques de finitisme et de constructivisme. Il affirme que le finitisme de Hilbert est arithmétique₁ mais que la réciproque est fautive [Isaacson 1992, p. 97], dans la mesure où

²⁵¹ Dont on rappelle qu'il nécessite des conditions (de dérivabilité) supplémentaires.

²⁵² Daniel Isaacson ne mentionne pas explicitement le théorème de Matijasevic. Mais Torkel Franzen, dans une discussion de la même problématique, renvoie au résultat de Matijasevic-Robinson-Davis-Putnam, selon lequel l'énoncé codant la consistance de PA est équivalent à la formulation qu'une certaine équation Diophantienne n'a pas de solution [Franzen 2004, p. 186].

l'aptitude qui permet de reconnaître qu'un énoncé est arithmétique_I ne peut pas être représentée dans PA, a fortiori dans PRA, et échappe donc au finitisme de Hilbert [Isaacson 1987, p. 207, n. 3]. Cette double affirmation est plausible en remarquant que la consistance de PRA, finitiste au sens de Hilbert, est bien décidable dans PA (même si elle ne l'est pas dans PRA en suivant la thèse de Tait selon l'interprétation standard). Pour l'auteur, les extensions du finitisme de Hilbert et en particulier le constructivisme associé au théorème de Gentzen de 1936, sont non-arithmétiques_I [Isaacson 1992, p. 99]. La justification qu'il en donne est que la consistance de PA est un concept d'ordre supérieur, qui n'est pas lui-même arithmétique_I et doit donc être supposée a priori, au même titre que l'aptitude à reconnaître la vérité d'un énoncé arithmétique_I. Là encore, cette affirmation paraît recevable. L'auteur suggère alors que l'arithmétique_I se situe au-delà du finitisme de Hilbert, trop restreint, mais en deçà d'un constructivisme à la Gentzen, trop étendu [Isaacson 1992, p. 99]. Une telle affirmation est par contre contestable dans la mesure où, comme nous l'avons indiqué, les conditions nécessaires pour établir la consistance de PA suivant la démarche de Gentzen dépassent certainement les moyens de PA par certains côtés, *tout en étant plus faibles par d'autres*. De fait, si l'on compare la force respective des présupposés conduisant aux énoncés arithmétiques_I d'une part, et des conditions nécessaires à la démonstration de Gentzen de l'autre, par exemple en s'intéressant à l'ordinal limite associé (tableau III du § 3), on aboutit à quelque chose d'analogue (à savoir l'ordinal limite : ε_0). Cette remarque rejoint celle de Michael Potter [Potter 2000, p. 254].

Enfin, dans l'article de 1992, l'auteur discute de manière détaillée de l'interprétation de la règle- ω ainsi que nous l'avons vu au § 3.2. De manière précise, il entend bloquer l'argument suivant, qui s'oppose à sa thèse :

- 1) la règle- ω est une expression naturelle de notre saisie de la structure des nombres naturels ;
- 2) lorsque l'on ajoute la règle- ω à PA, on reste donc dans le domaine arithmétique_I ;
- 3) or l'ajout de la règle- ω à PA rend PA correct, ce qui fait que tous les énoncés vrais de PA sont arithmétiques_I.

Pour cela, il conteste le point 1), en introduisant un distinguo subtil concernant PA²⁵³ entre les deux faits suivants :

- 1) pour *chaque* théorème, nous pouvons voir qu'il est vrai sur la seule base de la prise en compte de la structure de l'ensemble des nombres naturels (donc de manière arithmétique_I) ;

²⁵³ Distinguo déjà introduit par Gödel en 1951 [Boolos 1995, p. 292-293].

2) pour l'ensemble des théorèmes (ou pour l'affirmation que *tout* théorème est vrai), nous devons faire intervenir une réflexion de niveau supérieur (une réflexion de réflexion) qui dépasse la limitation précédente (et qui est donc non-arithmétique_I).

En attachant la règle- ω à cette seconde opération, il en déduit que celle-ci est elle-même non-arithmétique_I [Isaacson 1992, p. 108]. L'auteur introduit ensuite une version particulière de la règle- ω qu'il nomme (et estime) « finitiste » (au sens de Hilbert) dans le but d'étendre PA (voir infra le § 3.2). Il retrouve alors l'ordinal limite de PA : ε_0 et en déduit que cette règle- ω « finitiste » ne peut pas être elle-même arithmétiquement_I justifiée, tandis qu'elle permet de dériver certains énoncés Π_1 non dérivables dans PA [Isaacson 1992, p. 127-128]. Il introduit enfin une version restreinte de cette règle- ω « finitiste », afin d'étendre cette fois PRA (en fait le système formel : $I\Delta\Sigma_1$, de même force et de même ordinal limite). Il retrouve ici l'ordinal limite ω^ω , pour lequel il estime que cette extension peut être arithmétiquement_I justifiée²⁵⁴, mais montre que celle-ci ne conduit pas à dériver plus de théorèmes que ceux déjà dérivés dans PA [Isaacson 1992, p. 124-125].

A l'évidence, la thèse de Daniel Isaacson présente un intérêt indéniable dans le cadre des discussions sur les fondements des mathématiques, tout en possédant une forte cohérence interne. Son défaut majeur est la difficulté de justifier le critère retenu pour définir la vérité arithmétique_I qui en est le ressort. Si l'on s'en tient à la justification épistémologique proposée (se limiter à notre compréhension de la structure des nombres naturels), celle-ci est convaincante dans son esprit mais ne permet pas de tracer une ligne de démarcation précise entre énoncés arithmétiques_I et non-arithmétiques_I. Si l'on prend en compte la justification « formelle » (la vérité arithmétique_I est fixée par les théorèmes de PA), on obtient un critère de démarcation précis, mais en introduisant un caractère arbitraire dans le choix de PA (par rapport à des systèmes formels de l'arithmétique un peu plus faibles ou un peu plus forts). Dans les deux cas il en résulte une force d'évidence beaucoup plus réduite que celle du finitisme de Hilbert, par exemple. En ne retenant qu'une version affaiblie de la thèse d'Isaacson, il reste tout à fait légitime d'introduire une dichotomie parmi les énoncés arithmétiques (exprimables dans PA), qui soit philosophiquement signifiante et non *ad hoc*, entre : 1) ceux qui sont des énoncés métamathématiques, c'est-à-dire des codages de relations syntaxiques, au premier rang desquels l'énoncé G de Gödel ainsi qu'un énoncé codant la consistance simple *d'un système formel quelconque* (pas nécessairement PA lui-même), et 2) tous les autres énoncés, que l'on

²⁵⁴ L'ordinal limite ω^ω correspond à celui du finitisme élargi au sens de notre tableau III du § 3, au sein duquel la consistance de PRA peut être établie (ce que Daniel Isaacson ne mentionne pas explicitement).

nommera « purement arithmétiques »²⁵⁵. L'analyse que nous avons effectuée conduit alors au résultat qu'aucun énoncé « purement arithmétique » n'a été montré indécidable dans PA à ce jour. Il est donc possible de proposer une conjecture étendant à la fois la thèse d'Isaacson²⁵⁶ et la conjecture de Friedman : « Tout énoncé exprimable dans PA, qui n'est pas le codage d'un énoncé métamathématique concernant un système formel naturel, est décidable dans PA ».

4.5 La solution radicale de l'ultrafinitisme

Le principe général de la démarche ultrafinitiste consiste à prolonger l'idée présente à la base du constructivisme de Brouwer, en remplaçant ce qui y est autorisé *en principe* (sans tenir compte de limitations sur la taille des énoncés et sur le nombre d'étapes d'une dérivation), par ce qui est possible *en pratique* (en prenant en compte de telles limitations contingentes), comme le résume Michael Dummett (en utilisant le terme de « finitisme strict » à la place de celui « d'ultrafinitisme ») :

Le constructivisme traditionnel a autorisé que les constructions mathématiques par référence auxquelles les significations mathématiques sont données, puissent être celles que nous sommes capables d'effectuer seulement en principe. Il ne fait aucune différence sur le fait de savoir si elles sont ou non trop complexes ou simplement trop longues à effectuer en pratique, pour tout être humain, ou même pour la race humaine toute entière prise en collaboration. Le finitisme strict rejette cette concession aux vues traditionnelles et insiste plutôt sur le fait que les significations de nos termes doivent être données par référence aux constructions que nous pouvons réaliser en pratique, [...]. [Dummett 1975, p. 301]

Historiquement, il semble qu'une telle idée ait été suggérée pour la première fois par Bernays [Bernays 1935b, p. 91]. Celle-ci a ensuite été évoquée dans les années 1950, par Gödel [Gödel 1951 (1995), p. 313, 319], [Gödel 1953 (1995), p. 10-14] et [Gödel 1956 (2003), p. 375], par Borel [Borel 1952, p. 4, 13] et par L. van Dantzig [van Dantzig 1955, p. 273]. Elle a surtout été exploitée dans la période s'étendant des années 1960 au début des années 1980²⁵⁷, par Alexander Yessenin-Volpin : [Yessenin-Volpin 1961] et [Yessenin-Volpin 1970], par Rohit

²⁵⁵ Il peut bien entendu se faire que certains énoncés métamathématiques concernant un système formel naturel se présentent sous une forme qui les rende intéressants pour les mathématiciens « au travail ». Il en résulterait alors un hiatus entre une distinction épistémologique claire et la pratique « sur le terrain ».

²⁵⁶ Le fait qu'à ce jour aucun problème arithmétique (et non métamathématique) ouvert n'ait été montré indécidable, constitue évidemment un renforcement de la vraisemblance de la thèse d'Isaacson.

²⁵⁷ Les travaux plus récents sont essentiellement ceux qui ont été effectués sur la base de la notion de complexité de Chaitin-Kolmogorov, introduite par la Théorie Algorithmique de l'Information (voir le § 5.3), travaux dont la thèse de Cyrille Imbert présente une synthèse à la fois récente et détaillée : [Imbert 2008].

Parikh : [Parikh 1971] et par Robin Gandy : [Gandy 1982]. Sur un plan conceptuel il paraît difficile d'en nier l'intérêt en tant qu'option possible en philosophie des mathématiques (la première colonne de notre tableau III du § 4), ainsi que le suggère Robin Gandy de manière provocatrice :

Il n'est bien sûr admis en pratique que des inscriptions concrètes et une longueur limitée ; mais un auteur qui a fait cet aveu est porté à affirmer *qu'en principe* des inscriptions arbitrairement longues peuvent être écrites et que les opérations et tests de syntaxe élémentaire peuvent en principe être effectivement réalisés sur elles. Cela me semble aussi choquant que de dire *qu'en principe* les cochons peuvent voler, tout en admettant qu'en pratique il se trouve qu'ils n'ont pas d'ailes. [Gandy 1982, p. 129]

La contestation de ce point par Michael Dummett [Dummett 1975, p. 301, 323], basée sur l'assimilation d'une telle position au problème linguistique des termes vagues, nous paraît passer à côté du problème. Il est cependant clair qu'il n'y a pas de consensus sur le lien entre l'ultrafinitisme²⁵⁸ (qui présente lui-même différentes formes) et la typologie classique (incluant le finitisme et le constructivisme).

Concrètement, l'idée générale commune aux différentes formes d'ultrafinitisme, consiste à introduire une limite supérieure n_{sup} au nombre de caractères de toute dérivation prise en compte dans un système formel et, par extension, au domaine des entiers naturels sur lequel seront admis les opérateurs classiques : « + » et « × ». n_{sup} est bien entendu relié à des considérations *physiques*, en relation avec la quantité d'information accessible, à un être humain, à l'ensemble de l'humanité, ou au sein de l'Univers lui-même. Outre le choix de la valeur de n_{sup} ²⁵⁹, se posent alors plusieurs questions, dont la variété des réponses apportées se traduit par les différentes solutions d'ultrafinitisme qui ont été proposées, en particulier :

- 1) comment rendre compte des opérations sur les entiers lorsqu'elles font intervenir des valeurs supérieures à n_{sup} ?
- 2) que deviennent les notions métamathématiques classiques de consistance et de complétude syntaxique ?

²⁵⁸ Par exemple James Geiser ([Geiser 1974]) propose une filiation conduisant du réalisme à l'intuitionnisme (au sens de Brouwer), puis de l'intuitionnisme à l'ultra-intuitionnisme (terme utilisé par Alexander Yessenin-Volpin mais contesté par Georg Kreisel et Andrezej Ehrenfeucht [Kreisel, Ehrenfeucht 1967, p. 517], en proposant justement celui d'ultrafinitisme).

²⁵⁹ Au-delà de la question de son caractère accessible ou non, pour laquelle seul Emile Borel répond par la négative [Borel 1952, p. 4]. Robin Gandy considère quant à lui que c'est l'univers des fonctions d'entiers qui peuvent être établies, qui est une totalité ouverte [Gandy 1982, p. 133].

Commençons par donner une idée de la valeur de n_{sup} , valeur qui est en général liée à une certaine quantité d'information²⁶⁰. Robin Gandy en propose une quantification précise, en distinguant :

- 1) les inscriptions qui peuvent être envisagées « en pratique », pour lesquelles la plage proposée est de 10^{14} à 10^{20} bits ;
- 2) celles qui peuvent être envisagées « en principe » (en un sens évidemment différent de celui habituellement associé à cette locution), pour lesquelles la plage proposée est comprise entre 10^{50} et 10^{500} bits²⁶¹ [Gandy 1982, p. 131].

Le problème posé par un sens attribué aux opérateurs élémentaires sur les entiers : « + » et « × » (a fortiori l'exponentiation) nécessairement différent de celui classique, a déjà été soulevé par Gödel et L. van Dantzig [Gödel 1951, p. 319] et [van Dantzig 1955, p. 274]. La solution la plus simple qui a été proposée est de considérer que le successeur de n_{sup} est n_{sup} lui-même, comme le défend Jan Mycielski [Mycielski 1981, p. 627]²⁶². Mais quelle que soit la solution retenue, il est clair qu'elle va se traduire par des opérateurs « + » et « × » spécifiques. Qui plus est, comme cela est suggéré par L. van Dantzig [van Dantzig 1955, p. 274] et souligné par David Isles [Isles 1981, p. 133], les propriétés des opérations réalisées sur les entiers supérieurs à n_{sup} devront alors être postulées (puisqu'elles ne peuvent être prouvées), ce qui introduit une relativisation du choix effectué comme le précise L. van Dantzig [van Dantzig 1955, p. 275-276]. Ce problème est bien entendu amplifié avec l'opérateur exponentiation, ce qui nous ramène à la discussion que nous avons eue à ce sujet à propos du finitisme (§ 1.3). Pour illustrer le caractère particulièrement contre-intuitif de l'ultrafinitisme²⁶³, nous reprenons l'exemple donné par Robin Gandy avec deux énoncés successifs [Gandy 1982, p. 131-132] :

A) « $5^{5^5} + 7^{7^7} + 1$ » est un nombre premier ;

²⁶⁰ Ce qui pose déjà une question fondamentale. Soit B la capacité mémoire (en bit) prise en compte. On peut alors *dénoter* (abstraitement) tous les entiers jusqu'à 2^B (soit par exemple 256 si $B = 8$), la converse étant que la dénotation d'un entier n nécessite une quantité d'information de $\log_2(n)$ bits. Mais on ne peut *représenter* (concrètement, par exemple sous la forme d'une suite de bâtons), que les nombres jusqu'à B (donc 8 si $B = 8$). La différence est évidemment considérable. Si l'on prend l'exemple de la mémoire d'un être humain, estimée à 10^{11} bits, elle permet de dénoter tous les entiers jusqu'à $2^{10^{11}}$, mais elle ne permet de représenter que les entiers inférieurs à 100.000.000.000, ce qui est sans commune mesure.

²⁶¹ Les deux plages indiquées apparaissent cohérentes avec les ordres de grandeur donnés par Jean-Paul Delahaye [Delahaye 2006, p. 6-17 et 18-27] :

- 1) la mémoire naturelle d'un humain est d'environ $2 \cdot 10^{11}$ bits ;
- 2) la mémoire de l'ensemble de l'humanité est (en 2006) d'environ $2 \cdot 10^{23}$ bits ;
- 3) la quantité d'information qui peut être associée à l'ensemble de l'Univers visible est estimée à environ 10^{123} bits.

²⁶² Même si, en l'occurrence, pour cet auteur, n_{sup} n'est pas fixé a priori, ce qui déborde l'ultrafinitisme classique.

²⁶³ Et particulièrement déroutant, puisque cette position implique évidemment qu'il y ait un plus grand nombre premier.

B) « $5^{5^{(5^5)}} + 7^{7^{(7^7)}} + 1$ » est un nombre premier.

L'énoncé A) peut être vérifié « en pratique » (sous-entendu, par comptage), tandis que l'énoncé B) n'est pas vérifiable (toujours par comptage), même « en principe ». Se pose également le problème du codage gödelien dans un cadre ultrafinitiste, qui n'a jamais été abordé directement par les protagonistes. La solution la plus simple (et sans doute celle implicitement imaginée par eux) est de borner la gödelisation à n_{sup} . Elle présente deux inconvénients : 1) les résultats deviennent relatifs au choix (arbitraire) du codage gödelien retenu, 2) en pratique, tous les codages connus impliquent des entiers extraordinairement grands, qui interdisent de fait toute velléité de les écrire réellement sous forme d'une suite de bâtons²⁶⁴. Le second inconvénient cité est évidemment majeur, puisqu'il interdit de fait de pouvoir vérifier en pratique toute dérivation formelle ainsi codée. Une alternative simple, qui est celle utilisée par exemple par Robin Gandy dans [Gandy 1982], consiste à borner le nombre de *caractères* d'une représentation concrète (d'un énoncé ou d'une démonstration). Cela autorise évidemment des dérivations beaucoup plus longues (permettant de rendre compte de toutes celles envisageables en pratique par les mathématiciens), mais en perdant alors le caractère d'évidence évoqué par Hilbert dans la présentation de son finitisme, qui ne fait appel qu'à des comparaisons (possibles au moins en principe) sur des suites de bâtons.

Le questionnement métamathématique se pose en des termes spécifiques dans le cadre de l'ultrafinitisme. En effet, les questions, par exemple de la dérivabilité de tout énoncé donné au sein d'un système formel, ou bien de la non-contradiction de tout énoncé, n'ont à être vérifiées qu'en-deçà d'une certaine longueur de l'énoncé. En particulier si une contradiction ne survient que pour des énoncés dont la longueur est supérieure à n_{sup} , alors le système formel peut être considéré comme consistant (par rapport à cette valeur) au sens de l'ultrafinitisme (on ne peut plus appliquer sans discernement l'effet logique de *l'ex falso sequitur quodlibet*). De manière converse, pour établir la consistance d'un système formel au sens ultrafinitiste, il suffit de montrer que la consistance est assurée pour tout énoncé de longueur inférieure à n_{sup} . C'est la démarche qu'a utilisé Rohit Parikh, tout en considérant ouverte la question concernant PA [Parikh 1971, p. 506-507]. La difficulté est bien entendu que la démonstration doit elle-même être d'une longueur inférieure à n_{sup} . Un raisonnement simple illustre la difficulté de l'exercice. Supposons que l'on parvienne à une telle démonstration sans que cette démonstration ne fasse

²⁶⁴ Si l'on prend l'exemple donné par Ernest Nagel et James Newman [Nagel, Newmann 1958 (1989), p. 80, n. 1], le codage naturel de l'expression très simple : « $(\exists x)(x=s(y))$ » est supérieur à 2^{108} . Sa représentation concrète sous forme d'une suite de bâtons impliquerait donc une taille mémoire de 2^{108} , très supérieure à celle de l'humanité.

elle-même intervenir n_{sup} . En se replaçant dans un contexte finitiste et en faisant croître n_{sup} , on obtiendrait une démonstration finitiste de PRA ou PA, en contradiction avec le second théorème de Gödel de 1931. Gödel en faisant déjà la remarque [Gödel 1953 (1995), p. 14, n. 11 et p. 19, n. 38]. De fait, Pavel Pudlak précise, sur la base de [Parikh 1971], que la démonstration montrant qu'il n'y a pas de contradiction de longueur supérieure à n est elle-même d'une longueur d'environ n [Pudlak 1996, p. 68]. A titre d'exemple de résultat obtenu, concernant la théorie ultrafinitiste FIN mentionnée par Jan Mycielski [Mycielski 1981, p. 632-633], on peut établir la consistance de FIN dans PA (mais pas dans FIN elle-même) et FIN est essentiellement indécidable. Une situation analogue est présente pour la complétude sémantique. C'est ainsi que Robin Gandy fait remarquer qu'il peut exister des énoncés concrets de type Π^0_1 tout à fait simples, éventuellement classiquement faux mais indécidés (au sens de l'ultrafinitisme) : ils peuvent être pris comme hypothèse sans qu'aucune contradiction ne soit découverte « en pratique » [Gandy 1982, p. 135]. Que deviennent alors les deux théorèmes de Gödel de 1931 dans un tel contexte ? Bien évidemment, comme ces théorèmes (et tous ceux apparentés) font appel de manière essentielle à la non finitude des entiers (via l'argument de la diagonale), ils deviennent inopérants. Une autre conséquence de la position ultrafinitiste, est la perte des résultats métamathématiques « positifs » sur des systèmes formels faibles pour lesquels les théorèmes de Gödel de 1931 ne s'appliquent pas, à l'instar de Q ²⁶⁵. Ce point a été clairement souligné par Edward Nelson²⁶⁶ dans [Nelson E. 1992] et [Nelson E. 1993].

En évoquant l'ultrafinitisme il convient de mentionner deux cas particuliers. Le premier est Alexander Yessenin-Volpin, figure originale dans le monde des mathématiciens, dont l'objectif est en fait proche du programme de Hilbert puisqu'il vise à établir la consistance de ZF mais sur une base plus faible que celle du finitisme. En fait, comme le remarque Solomon Feferman [Feferman 2003, p. 58-59], ce n'est qu'une forme très faible de la consistance qui est en jeu : une contradiction peut seulement survenir dans une preuve d'une certaine longueur minimale, qui n'est plus alors vue comme une preuve concrète. Cette réserve, associée au fait que les deux articles de l'auteur : [Yessenin-Volpin 1961] et [Yessenin-Volpin 1970] ne contiennent qu'une

²⁶⁵ Nous rappelons les particularités métamathématiques du système formel Q (voir le tableau I § 1.4) :

1) Le premier théorème de Gödel de 1931 s'y applique, mais l'on peut établir l'incomplétude syntaxique de Q de manière triviale, sans faire appel à lui ;
 2) Le second théorème de Gödel de 1931 ne s'y applique pas, mais l'on peut tout de même établir l'impossibilité de dériver en son sein (le codage d') une formulation raisonnable de sa consistance : [Bezboruah, Shepherdson 1976]. Il est malgré tout possible d'établir de manière finitiste la consistance de Q par le biais du théorème de Hilbert-Ackermann de 1928. Il n'y a aucune contradiction entre ces deux dernières affirmations, simplement le finitisme (au sens classique) va au-delà du pouvoir d'expression de Q .

²⁶⁶ Celui-ci ne fait pas explicitement référence à l'ultrafinitisme, mais sa critique du « dogme finitiste » y conduit implicitement.

esquisse de preuve, s'est traduite par un accueil sceptique de la communauté des mathématiciens [Pudlak 1996, p. 66] et [Feferman 2003, p. 58-59, n. a¹]. Le second est le mathématicien singulier et solitaire Eduard Wette, qui n'affirme rien de moins qu'avoir prouvé l'inconsistance de la théorie élémentaire des nombres en se fondant sur un finitisme plus strict que celui de Hilbert : [Wette 1974]. Il a eu des échanges avec Bernays et a tenté sans succès d'échanger avec Gödel [Gödel 1975 (2003), p. 309] et [Bernays 1975 (2003), p. 313]. Il n'a cependant manifestement réussi à convaincre ni l'un ni l'autre, comme cela apparaît sans ambiguïté dans les deux lettres citées, et de manière plus nuancée dans une publication de Bernays [Bernays 1971/1976 (2003), p. 230].

Pour achever cette analyse de l'ultrafinitisme, il nous reste à évoquer la solution particulière proposée par Jean-Paul van Bendegem : [Bendegem 1993] et Graham Priest : [Priest 1994], qui consiste à associer l'ultrafinitisme à la logique paraconsistante pour proposer ce qui est présenté comme une extension du programme de Hilbert. Le recours à l'ultrafinitisme est non ambigu (même s'il est nommé « finitisme strict » par ces deux auteurs) :

Fixons désormais n [le « nombre » paraconsistant tel que $n = n + 1$ est vrai dans le modèle de PA de la logique paraconsistante selon l'interprétation de Graham Priest] comme quelque chose d'incroyablement grand, disons un nombre plus grand que le nombre de combinaisons de particules dans le Cosmos, plus grand que tout nombre qui peut être spécifié sur un mode sensible durant une vie, si grand qu'il n'a aucune signification physique ni réalité psychologique.

[Priest 1994, p. 338 ; le commentaire entre crochets est nôtre]

L'évolution par rapport au programme de Hilbert au sens restreint porte alors sur trois points :

- 1) le caractère « ni vrai ni faux » des énoncés idéaux de Hilbert, devient « à la fois vrai et faux » dans le cadre de la logique paraconsistante (selon l'interprétation de Priest) ;
- 2) la consistance simple du programme de Hilbert est remplacée par la consistance absolue, qui est affirmée démontrable [Priest 1994, p. 344] ;
- 3) la coupure entre les énoncés « réels » et les énoncés « idéaux » n'est plus renvoyée à l'infini comme chez Hilbert, mais située au-delà d'une certaine borne finie.

L'énoncé G de Gödel du premier théorème appliqué par exemple à PA est alors naturellement renvoyé au-delà de la limite « finitiste » introduite, où il est considéré comme « à la fois vrai et faux », en suivant l'interprétation dialéthéiste habituelle que nous avons déjà présentée. Nous constatons qu'une telle solution cumule alors les limites de l'ultrafinitisme classique que nous venons d'évoquer et les critiques fortes que nous avons opposées à l'approche paraconsistante (§ 3.5).

Quel bilan ressort finalement des tentatives très diverses de réalisation partielle du programme de Hilbert au sens restreint que nous avons analysées ? Nous rappelons pour mémoire, ce constat étant bien connu, la portée très réduite des différents théorèmes des années 1930 qui ont prétendu établir de manière finitiste la consistance de l'arithmétique. Dans la même direction, les résultats plus récents obtenus dans les années 1970 et 1980 dans le cadre des « reverse mathematics » vont beaucoup plus loin. Ceux-ci établissent en effet qu'une grande partie des mathématiques existantes peut être formalisée dans des systèmes formels qui sont montrés conservatifs par rapport à PRA (au moins) pour les énoncés Π^0_1 . Il s'agit donc bien *de fait* d'une réalisation partielle mais significative du programme de Hilbert au sens restreint selon notre interprétation. Le caractère partiel de la réalisation réside dans la limite de l'étendue du domaine des mathématiques concernées. Des solutions ont été proposées pour repousser cette limite, mais qui se rapportent alors à des systèmes formels d'autant plus forts que PRA que les extensions obtenues sont importantes. Or pour ces systèmes formels, il n'est plus possible d'en établir la consistance sur une base finitiste même raisonnablement étendue. On retrouve ici le constat attendu que les deux exigences, d'une part de s'en tenir à un finitisme fort, et d'autre part de rendre compte de l'ensemble des mathématiques, sont antagonistes. Puisque les théorèmes de Gödel de 1931 s'opposent à la réalisation du Programme de Hilbert, une solution radicale consiste à les exclure *a priori* des mathématiques (en l'occurrence de l'arithmétique) en arguant du fait qu'ils font intervenir en fait des énoncés *métamathématiques*. C'est la thèse d'Isaacson, ou tout du moins une version générale de celle-ci. Une objection naturelle est l'irruption dans les années 1970 de plusieurs théorèmes, montrant que certains énoncés combinatoires finis théorèmes de Z_2 sont des indécidables de PA. La réponse est que ces énoncés s'avèrent être encore des énoncés métamathématiques qui sont donc également exclus par la thèse d'Isaacson, même s'ils se présentent sous une forme mathématique plus naturelle que l'énoncé G de Gödel, ou le codage brut de l'affirmation de la consistance d'un système formel naturel de l'arithmétique. Une conjecture, étendant celle de Friedman, peut être alors naturellement proposée : « Tout énoncé exprimable dans PA, qui n'est pas le codage d'une expression métamathématique concernant un système formel naturel, est décidable dans PA ». Il va de soi que la thèse d'Isaacson et la conjecture qui en a été induite ne peuvent prétendre réaliser le programme de Hilbert au sens restreint, dans la mesure où la question de la démonstration finitiste en un sens raisonnable de la consistance de PA amène une réponse négative. La solution encore plus radicale de l'ultrafinitisme, envisagée principalement dans les années 1960 et 1970, n'y réussit pas d'avantage alors qu'elle est par ailleurs assaillie de difficultés spécifiques, ce qui explique sans doute son abandon (au moins par les

mathématiciens). Les différentes voies évoquées et en particulier la thèse d'Isaacson n'en constituent pas moins une approche intéressante concernant l'autre préoccupation de Hilbert, sur l'absence de problèmes mathématiques insolubles, que nous allons aborder maintenant d'une manière beaucoup plus générale.

5 Le programme de Hilbert au sens large : les résultats « négatifs » en logique formelle et en informatique théorique s'opposent-ils à la justification de l'absence de problèmes mathématiques insolubles ?

L'autre grande préoccupation de Hilbert en philosophie des mathématiques, au-delà de la résolution finitiste du problème des fondements (le programme au sens restreint), était la justification de l'absence de problèmes mathématiques insolubles (l'ensemble conduisant au programme au sens large). Les principaux résultats « négatifs » obtenus concernant PA : son incomplétude syntaxique (premier théorème de Gödel de 1931) et son indécidabilité (théorème de Church de 1936 et de Turing de 1936-1937), sont habituellement présentés comme une réfutation de cette dernière position de Hilbert, ce qui nécessite déjà de donner un contenu technique précis à la position de Hilbert, au moins pour l'arithmétique. Dans un contexte complètement différent, mais également à partir d'extensions du théorème de Turing de 1936-1937, les figures centrales de la Théorie Algorithmique de l'Information que sont Grégory Chaitin et Cristian Calude défendent quant à elles une position qui s'inscrit radicalement en faux contre celle de Hilbert. Ces auteurs n'affirment en effet rien de moins que l'existence de vérités arithmétiques « absolument » indécidables²⁶⁷, et partant de *l'existence* de problèmes mathématiques insolubles. Une telle position a à son tour été contestée, principalement par Panu Raatikainen. Nous pouvons souligner par ailleurs que pour l'ensemble des débats que nous venons d'évoquer, la prise en compte de la thèse de Church-Turing intervient de manière cruciale en permettant de donner un contenu technique précis à l'indécidabilité ; or un certain nombre de propositions ont été avancées récemment en faveur d'un dépassement de celle-ci. De ce dépassement, il en résulterait un concept élargi de décidabilité (aussi bien d'un problème que d'un système formel), dont on pourrait espérer qu'il permette de réduire, voire d'éliminer, les situations d'indécidabilité rencontrées en arithmétique et partant de reposer la question de la possibilité d'une justification de l'absence de problèmes mathématiques insolubles²⁶⁸. Au préalable, il est intéressant d'évoquer les résultats peu connus de Tarski de 1940-1967 et 1948, qui conduisent, en contrepoint du 1^{er} théorème de Gödel de 1931, à une réponse « positive » en

²⁶⁷ C'est-à-dire indécidables pour tout système formel répondant aux critères généraux d'effectivité et non seulement pour un système formel particulier (à l'instar de l'énoncé G du premier théorème de Gödel de 1931).

²⁶⁸ Dans cette discussion, nous nous limiterons au cadre général des situations « en principe », c'est-à-dire où les contraintes contingentes fixant une limite à la taille des énoncés et au nombre d'itérations autorisées ne sont pas prises en compte, quand bien même ce nombre dépasserait les possibilités offertes par l'Univers observable (se reporter à notre présentation de l'ultrafinitisme au § 4.5).

faveur de la complétude syntaxique de véritables théories mathématiques, à savoir, l’algèbre élémentaire et la géométrie élémentaire.

5.1 Le théorème de Tarski de 1940-1967 et 1948²⁶⁹ sur la complétude syntaxique et la décidabilité de l’algèbre élémentaire ainsi que de la géométrie élémentaire

L’importance épistémologique de ce résultat de Tarski provient du fait qu’il s’agit du seul résultat de *complétude* syntaxique obtenu pour une véritable théorie mathématique, c’est-à-dire une théorie mathématique utilisée par les mathématiciens *avant* l’établissement du résultat en question²⁷⁰. Malgré cette importance, ce théorème a été à peu près ignoré des commentateurs des théorèmes de Gödel de 1931 (à l’exception de Judson Webb [Webb 1968, p. 163, n. 12]), dont il constitue pourtant le contrepoint naturel, en en faisant de ce fait ressortir les particularités. Il a également été ignoré des commentateurs du programme de Hilbert au sens large, à l’exception notable d’Hourya Sinaceur [Sinaceur 1991 (1999), p. 301 – 354], alors qu’il constitue pourtant indéniablement un élément positif en sa faveur. Le théorème de Tarski de 1940-1967, 1948 se rapporte en fait à plusieurs théories mathématiques : la théorie élémentaire du corps ordonné des nombres réels, la théorie élémentaire du corps des nombres complexes, et la géométrie euclidienne du plan. Les théories en question sont élémentaires : par exemple pour la théorie élémentaire du corps ordonné des nombres réels, seules sont autorisées les variables d’individus constituées de nombre réels, en excluant à la fois les sous-ensembles de celles-ci ainsi que les relations entre celles-ci. La limite de ce théorème est de ne pouvoir représenter l’ensemble des entiers naturels (et donc l’arithmétique), que ce soit explicitement ou implicitement, pour ne pas tomber sous le coup du 1^{er} théorème de Gödel de 1931.

²⁶⁹ La publication de ce résultat de Tarski a une histoire compliquée, qui explique la variabilité des dates qui lui sont associées. Comme l’auteur s’en explique lui-même en introduction de [Tarski 1940-1967], il existe en fait trois versions différentes de cette publication. La version originale, rédigée en 1939, aurait dû être publiée en 1940, mais du fait de la guerre elle ne l’a été que beaucoup plus tard : en 1967, avec un très petit nombre de tirages. Entre temps, Tarski a remanié l’article (en changeant son titre), ce qui a conduit à une publication dans une première édition en 1948 et dans une seconde édition (avec quelques notes additives) en 1951. Nous nous référerons donc, *pour les mêmes résultats*, à la fois à [Tarski 1940-1967] et à [Tarski 1948]. Toujours d’un point de vue historique, il convient de signaler que de l’aveu même de Tarski, si l’article a été rédigé (dans la première version) en 1939, les résultats correspondants ont été obtenus vers 1930. Leur établissement est donc contemporain de celui des deux théorèmes de Gödel de 1931 [Sinaceur 1991 (1999), p. 326].

²⁷⁰ Il s’oppose donc en cela au résultat de complétude syntaxique le plus connu : celui de Presburger de 1929, qui porte sur un système très pauvre (et jugé artificiel par les mathématiciens) : l’arithmétique élémentaire *sans* l’opérateur de multiplication.

Explicitement, en s'interdisant d'adjoindre la théorie des nombres à la théorie considérée : « Tandis que chaque entier individuel peut être défini dans le système, le concept général de nombre entier n'y trouve pas place » [Tarski 1940-1967 (1974), p. 206]. Implicitement, en excluant la fonction exponentielle complexe ainsi que les fonctions trigonométriques [Tarski 1940-1967 (1974), p. 242]. Sans aucune ambiguïté il concerne, selon la terminologie moderne, la complétude syntaxique des théories considérées dont Tarski mentionne explicitement la définition [Tarski 1948 (1986), p. 337, n. 5]. La décidabilité s'en déduit donc directement, comme l'indique le titre de la seconde publication, même s'il y a eu un certain flottement sur ce point dans la première publication [Tarski 1940-1967 (1974), p. 208]. Cette évolution de formulation est à relier au fait que pour l'ensemble de la communauté des logiciens, le résultat que la complétude syntaxique entraîne d'une manière générale la décidabilité (pourvu que le système formel soit tel que l'ensemble de ses théorèmes soit récursivement énumérable), n'est devenu clair qu'après 1936 et les travaux sur la récursivité (nous aurons à y revenir au § 5.2). Sa démonstration présente un caractère constructif (au moins en un certain sens²⁷¹), comme cela est souligné par Tarski lui-même [Tarski 1940-1967 (1974), p. 206] et [Tarski 1948 (1986), p. 357, n. 15]. Ce caractère constructif ne doit toutefois pas induire en erreur : si la démonstration est sans aucun doute *effective* (elle peut être effectuée en principe), elle n'en est pas *effectuable* pour autant (elle ne peut être appliquée en pratique, compte tenu de sa complexité exponentielle²⁷²). Cette limitation peut donc rassurer a posteriori les mathématiciens qui craignaient avant 1931, qu'une démonstration de décidabilité n'enlève tout intérêt à leur activité (nous allons revenir sur ce point au § 5.2). En sens inverse, comme le fait remarquer Hourya Sinaceur, ce résultat de Tarski s'inscrit en faux contre une certaine association entre complétude syntaxique et simplicité (voir stérilité) de la théorie associée, que les mathématiciens avaient pu imaginer à partir des premiers résultats « positifs » obtenus (le théorème de Presburger de 1929 en particulier, mais également celui de Herbrand de 1930) [Sinaceur 1991 (1999), p. 330, 336].

²⁷¹ Selon Panu Raatikainen un résultat récent de Simpson et Friedman en renforce considérablement le caractère constructiviste : celui-ci établit que la consistance de la théorie élémentaire du corps ordonné des nombres réels peut être démontrée au sein de PRA (donc de manière finitiste au sens de Hilbert via la thèse de Tait) [Raatikainen 2003, p. 170].

²⁷² De manière plus précise, le résultat de Fisher et Rabin de 1974 montre que la théorie élémentaire des réels a une complexité au moins exponentielle, c'est-à-dire égale à 2^{cn} où n désigne le degré maximal des polynômes de la formule à décider et c est une constante positive [Sinaceur 1991 (1999), p. 328 – 329].

5.2 L'optimisme rationaliste de Hilbert confronté à l'incomplétude syntaxique et à l'indécidabilité de PA

« L'optimisme rationaliste »²⁷³ désigne a posteriori une conviction affichée très tôt par Hilbert, mais surtout exprimée par lui avec force lors de la conférence de 1900²⁷⁴ :

Je dois dire avant tout ceci : qu'il doit être possible d'établir le caractère correct de la solution [d'un problème mathématique] par le moyen d'un nombre fini d'étapes, basées sur un nombre fini d'hypothèses qui sont impliquées dans la formulation du problème et qui doivent être formulées exactement. Cette exigence de déduction logique au moyen d'un nombre de processus fini est simplement l'exigence de rigueur dans le raisonnement. [...]

Le problème survient alors : montrer l'impossibilité de la solution [d'un problème mathématique] sous les hypothèses données et dans le sens envisagé. [...] C'est probablement ce fait important ainsi que d'autres raisons philosophiques, qui conduit à la conviction que tout problème mathématique défini doit être nécessairement susceptible d'un traitement exact, soit sous la forme d'une réponse exacte à la question posée, soit par la preuve de l'impossibilité de cette solution et avec cela de l'échec de toute tentative. [...]

Nous entendons en nous l'appel perpétuel : le problème est là, cherchez la solution. Vous la trouverez par la raison pure, dans la mesure où en mathématiques il n'y a pas d'*ignorabimus*. »

[Hilbert 1900b, p. 441, 444 et 445 ; le commentaire entre crochets est nôtre]

Comme ce point est central dans le cadre le plus général de notre réflexion et comme il a donné lieu à des incompréhensions de la part d'un certain nombre de commentateurs, nous proposons d'explicitier la notion d'optimisme rationaliste défendue par Hilbert sous la forme de l'exigence du respect de quatre critères successifs, pour tout problème mathématique²⁷⁵.

1) Le problème mathématique est *bien posé*, c'est-à-dire précisément formulé. Ce point est non problématique, il est à la fois explicite dans la formulation de Hilbert et non contesté.

2) La réponse qui consiste à *démontrer que le problème n'a pas de solution, est considérée comme une résolution licite* du problème. Ce point est évidemment crucial quant au

²⁷³ Selon Hao Wang, cette locution a été introduite en 1972 par Gödel, qui partageait avec Hilbert la conviction associée [Wang 1993, p. 119].

²⁷⁴ Selon Constance Reid, le refus de « l'*ignorabimus* » en mathématique qui caractérise l'optimisme rationaliste, est une réaction de Hilbert contre la prise de position opposée du biologiste Emil Dubois-Raymond, en 1872 : « *ignoramus et ignorabimus* » à fort retentissement à l'époque (elle concernait, d'une manière générale, les problèmes philosophiques liés à la nature de la matière et de l'esprit et l'origine de la conscience) [Reid, 1970, p. 13, 72].

²⁷⁵ Auxquels il convient d'ajouter la remarque suivante : dans le cas particulier des problèmes métamathématiques et arithmétiques, le problème porte nécessairement sur une infinité d'instances et non sur une instance particulière (avec les exemples emblématiques des conjectures de Fermat et de Goldbach).

positionnement sur l'existence ou non de l'*ignorabimus* en mathématiques. Il est pourtant quelquefois omis²⁷⁶, alors qu'il est explicitement mentionné par Hilbert sans ambiguïté possible. Qui plus est, comme le signalent Hourya Sinaceur [Sinaceur 1990d, p. 553] et Judson Webb [Webb 2005, p. 494-495], c'est justement le constat historique de la démonstration de l'*impossibilité* d'un certain nombre de problèmes mathématiques qui a convaincu Hilbert du bienfondé du refus de l'*ignorabimus* (en mathématiques).

3) La résolution du problème mathématique est associée à des *contraintes particulières*, qui doivent être explicitées. Ce point est important, puisque toute réponse négative sous la forme d'une preuve d'impossibilité sera nécessairement associée à ces contraintes²⁷⁷. Il est également explicite dans la formulation de Hilbert et non contesté.

4) La résolution du problème est associée à une exigence épistémologique générale d'*intersubjectivité* : la solution doit être publiquement communicable en un temps fini et doit pouvoir être vérifiée, *en principe*, par n'importe quel être humain²⁷⁸ sans impliquer aucun engagement philosophique particulier. Ce point est important dans la discussion qui va suivre sur les versions fortes de l'optimisme rationaliste. L'exigence du caractère publiquement communicable de la solution est également à peu près explicite dans la formulation de Hilbert, sous une forme annonçant les critères modernes d'effectivité [Hilbert 1900b, p. 441]. Compte tenu de la position finitiste défendue par ailleurs par Hilbert, son adoption par celui-ci peut difficilement être mise en doute. Quant à la neutralité philosophique demandée, l'on rappelle qu'elle est un élément fondamental du finitisme de Hilbert (§ 1).

A cette présentation générale, il convient d'ajouter une précision historique. Si le refus de l'*ignorabimus* en mathématiques est une conviction exprimée de manière constante par Hilbert, au moins depuis la conférence de 1900 et jusqu'à l'allocution de Königsberg le 8 septembre 1930, le statut qu'il a accordé à cette conviction a évolué. D'abord considérée comme une évidence, Hilbert en est venu à penser qu'elle pouvait être justifiée de manière rigoureuse. Constance Reid situe le changement vers 1917 [Reid 1970, p. 174]. Toujours est-il que cette dernière attitude est évoquée sans ambiguïté en 1926, où elle est même présentée comme un objectif de la théorie de la démonstration [Hilbert 1926 (1972), p. 236-237].

²⁷⁶ En particulier, si Brouwer identifie explicitement le principe du tiers exclu et l'affirmation selon laquelle tout problème mathématique est résoluble [Brouwer 1928 (1992), p. 248], Jean Largeault, introduit une réserve à cette identification, sur ce point précis [Largeault 1992e, p. 241].

²⁷⁷ Ces contraintes sont évidentes dans la formulation des exemples historiques les plus anciens et les plus connus de problèmes mathématiques résolus négativement : exprimer tout nombre réel positif sous la forme du quotient de deux entiers naturels ou bien établir la trisection d'un angle avec la règle et le compas.

²⁷⁸ Une référence à l'intersubjectivité humaine associée à l'optimisme rationaliste est mentionnée par Hao Wang [Wang 1974, p. 90, souligné par nous] : « [...], *chacun* peut résoudre le problème pourvu qu'il soit capable de réaliser les opérations que requiert l'algorithme et de suivre exactement les règles telles qu'elles sont données. ».

Cette volonté de justifier rigoureusement le refus de l'*ignorabimus* en mathématiques nécessite alors d'introduire des critères d'exigence supplémentaires, qui conduisent à des formes renforcées de l'optimisme rationaliste tel que nous l'avons énoncé, *en nous limitant à l'arithmétique* puisque c'est sur ce domaine que porte notre discussion. Le premier renforcement : optimisme rationaliste₁, porte sur trois points. Tout d'abord le critère (1) est traduit par l'exigence (faible) de disposer d'un système formel au sein duquel la solution sera établie, sans imposer ce système formel *ni même qu'il soit unique pour l'ensemble des problèmes*, mais il doit être naturel²⁷⁹. On y ajoute l'exigence supplémentaire que les théorèmes du système formel soient récursivement énumérables, ce qui revient à adopter les critères d'effectivité modernes habituels d'un système formel. Enfin, et il s'agit clairement de la contrainte la plus forte, le critère (3) est exprimé sous la forme de l'exigence que la solution soit *unique* pour l'ensemble des instances du problème. On est alors conduit à une demande de *décidabilité de tout problème arithmétique*²⁸⁰, sachant qu'une démonstration d'indécidabilité constitue également la résolution (négative) du problème²⁸¹ (par le critère (2), inchangé). La limite de cette démarche est bien entendu que l'on voit mal comment elle est susceptible de trancher entre le refus et l'acceptation de l'*ignorabimus*. Positivement, on ne saurait demander de l'appliquer successivement à chaque problème possible (ensemble évidemment infini). Négativement, comment savoir si un problème actuellement irrésolu²⁸² est susceptible de le rester à jamais ? Pour cela, il faudrait disposer d'une méta-réponse : être en mesure de démontrer que l'on ne peut pas établir, ni que le problème est décidable, ni qu'il est indécidable. Sans

²⁷⁹ Il s'agit surtout d'interdire la solution *ad hoc* qui consisterait à introduire l'énoncé traduisant le problème que l'on cherche à résoudre (ou sa négation) comme *axiome* du système formel.

²⁸⁰ Nous rappelons qu'un *problème* arithmétique (binaire) est décidable, s'il existe une procédure effective unique permettant de répondre de manière certaine : « oui » ou « non » à la question posée par le problème, pour chacune des instances possibles de ce problème, à l'issue d'un nombre fini d'étapes. Il est sans doute utile de préciser ici un point élémentaire. Ce point est que tout résultat d'indécidabilité porte, par définition, sur l'absence d'une méthode *uniforme* de résolution des différentes instances du problème arithmétique posé. Il n'interdit donc nullement une résolution spécifique adaptée à chaque instance ou à chaque famille d'instance du problème. Pour prendre un exemple, le théorème de Ramsey fini est un indécidable de PA (théorème de Paris, Harrington de 1977 : voir infra au § 4.3) : il n'y a pas de méthode de résolution uniforme pour l'ensemble de ses instances, néanmoins chacune de ses instances prise individuellement est décidable dans PA.

²⁸¹ L'exemple emblématique étant la résolution négative du 10^{ème} problème de Hilbert par Matijasevic en 1970. La coexistence, lors de la conférence de 1900, de la formulation générale de l'optimisme rationaliste et de cette demande de décidabilité pour un problème arithmétique particulier, montre par ailleurs qu'Hilbert séparait clairement les deux démarches dès cette époque.

²⁸² Les exemples ne manquent pas, même en arithmétique. Pour ne citer que les plus faciles à formuler : 1) Existe-t-il des nombres parfaits impairs (un nombre parfait est égal à la somme de ses diviseurs différents de lui-même, comme $6 = 1 + 2 + 3$, 28, 496) ? ; On a pu montrer en 2012 que s'il en existe, ils sont supérieurs à 10^{1500} , mais sans pouvoir montrer qu'il n'en existait pas [Ochem, Rao 2012] ; 2) Un nombre pair est-il toujours décomposable en la somme de deux nombres premiers (conjecture de Goldbach) ? On a pu montrer en 2014 que s'il en existe, ils sont supérieurs à 4×10^{18} , mais sans pouvoir montrer qu'il n'en existait pas [Oliveira, Herzog, Pardi 2014].

parler de savoir si une telle preuve est même envisageable²⁸³, le spectre d'une régression à l'infini se profile alors. Un second renforcement est donc introduit : optimisme rationaliste₂, qui porte sur deux points supplémentaires. On impose d'abord cette fois la condition que le système formel soit *unique*, donc qu'il soit en mesure de rendre compte de toute l'arithmétique ; tout problème arithmétique (binaire) se ramène alors à un énoncé exprimable dans le système formel. On ajoute ensuite la demande majeure qu'il y ait une solution, non seulement unique pour l'ensemble des instances de chaque problème, mais également unique pour l'ensemble des problèmes eux-mêmes²⁸⁴. Cela revient à formuler l'exigence de la *décidabilité de ce système formel* que l'on renforcera par celle de sa complétude syntaxique²⁸⁵. Cette fois, on est bien en mesure de trancher entre le refus et l'acceptation de l'*ignorabimus* (absence versus présence d'énoncés indécidables dans le système formel). Le prix à payer est bien entendu l'extrême renforcement nécessaire de la formulation initiale de l'optimisme rationaliste. Par ailleurs, par l'intermédiaire de l'introduction de l'exigence d'un système formel unique à même de rendre compte de l'ensemble de l'arithmétique, on voit alors apparaître un lien entre ce dernier renforcement et le Programme lui-même. On devine que la réponse à la question de l'impact des résultats d'indécidabilité et d'incomplétude syntaxique de PA sur la justification de l'optimisme rationaliste de Hilbert, va dépendre de la prise en compte ou non, par ailleurs, du Programme.

Remarquons d'abord que la prise en compte du Programme dans sa plus grande généralité contraint en effet l'expression de l'optimisme rationaliste, même sous sa forme doublement renforcée de l'optimisme rationaliste₂. Comme nous venons de le dire, le renforcement de l'optimisme rationaliste impose l'existence d'un système formel unique, capable de rendre compte au moins de l'ensemble de l'arithmétique (premier point du second renforcement que

²⁸³ Selon Judson Webb, Kaufman, après avoir établi en 1930 une « preuve de complétude pour la théorie des nombres » à partir d'un raisonnement philosophique, affirmait avoir montré l'impossibilité de même prouver l'indécidabilité du problème de la décision [Webb 1980, p. 160].

²⁸⁴ Contrairement à ce qu'un certain nombre de mathématiciens ont cru ou ont craint (comme Hardy en 1929 selon Robin Gandy [Gandy 1988, p. 62]), les mathématiques n'en seraient pas rendues triviales pour autant, dans la mesure où la connaissance du caractère résoluble *en principe* du problème ne fournit pas en soi de solution de celui-ci qui soit mathématiquement acceptable *en pratique* (avec le contre-exemple du théorème de Tarski de 1940-1967, 1948 : voir infra au § 5.1).

²⁸⁵ Nous rappelons qu'un *système formel* est décidable, s'il existe un algorithme de décision effectif permettant de déterminer, pour tout énoncé du système formel, si celui-ci est ou non un théorème du système formel. Dans le cas le plus courant où le système formel est de plus syntaxiquement complet, en appliquant cet algorithme à tout énoncé φ on discrimine les deux solutions possibles (on suppose que le système formel est simplement consistant) : φ est dérivable ou φ est réfutable. Si le système formel est *indécidable*, il est nécessairement syntaxiquement *incomplet* et il existe des énoncés indécidables. Pour être exhaustif, il faut mentionner le cas particulier des systèmes formels quelque peu artificiels, *décidables* tout en étant syntaxiquement *incomplets*. Dans ce cas, en appliquant l'algorithme à la fois à φ et à $\neg\varphi$, on discrimine également la troisième solution possible où φ est indécidable.

nous avons évoqué). Mais la prise en compte du Programme impose également que la démonstration de l'optimisme rationaliste soit elle-même finitiste, puisque selon Hilbert, l'optimisme rationaliste est lui-même une proposition métamathématique, donc qu'elle soit un théorème de PRA selon l'interprétation standard du finitisme (en suivant la thèse de Tait). Par contre, la question de savoir ensuite si, au-delà de la conviction d'optimisme rationaliste, l'école de Hilbert était persuadée ou non de la décidabilité des théories du 1^{er} ordre : (ce qui a été appelé l'*Entscheidungsproblem*) et en particulier de PA, nécessite une discussion. Il existe en effet une tension concernant l'école de Hilbert entre les positions affichées par celle-ci concernant PA à l'égard, d'une part de sa décidabilité, d'autre part de sa complétude syntaxique²⁸⁶. Le problème de la complétude syntaxique de Z_2 ²⁸⁷ (donc de PA) est énoncé en 1929 et considéré comme ouvert [Hilbert 1929 (1992), p. 183 : problème IV]. Il est assez peu discuté (et non controversé) que l'école de Hilbert était persuadée d'une réponse positive (donc en faveur de la complétude syntaxique de Z_2). Une confirmation un peu ambiguë en est donnée par Hilbert [Hilbert 1931 (1992), p. 191], elle est plus précise de la part de Herbrand²⁸⁸ et elle est surtout formulée explicitement par Bernays [Bernays 1930/1976 (2003), p. 80]. Le problème de la décidabilité de Z_2 (et donc de PA) est énoncé et considéré comme ouvert dès 1917 [Sinaceur 1990d, p. 550]. Par contre, dès 1927, von Neumann s'est clairement exprimé en faveur d'une réponse *négative* à son sujet²⁸⁹ (donc en faveur de l'*indécidabilité* de Z_2) [von Neumann 1927, cité dans Gandy 1988, p. 62]. En 1933, Schütte écrit également qu'une solution générale à l'*Entscheidungsproblem* n'est probablement pas possible [Gandy 1988, p. 63]. La difficulté technique qui survient alors est que les deux positions sont contradictoires : la complétude syntaxique d'un système formel (répondant aux critères d'effectivité modernes habituels) entraîne sa décidabilité. Cette contradiction ressort directement d'un commentaire de Bernays :

²⁸⁶ Ou d'une forme de complétude qui lui est coextensive. En particulier, [Hilbert 1929] et [Bernays 1930] se réfèrent à la saturation (ou complétude au sens de Post), qui se définit très simplement par le fait que l'ajout au système formel de n'importe quel énoncé non dérivable le rend simplement inconsistant. On montre que d'une manière générale (sans se limiter au 1^{er} ordre), ces deux formes sont coextensives. On peut noter par ailleurs que pour ce qui concerne la logique des prédicats du 1^{er} ordre, son incomplétude syntaxique est établie dans [Hilbert, Ackermann 1928].

²⁸⁷ Rappelons que jusque dans les années 1920, pour l'école de Hilbert, il était entendu par « arithmétique », l'arithmétique du second ordre, représentée par le système formel Z_2 . Une distinction claire entre premier et second ordre pour l'arithmétique, n'apparaît qu'en 1934 [Hilbert, Bernays 1934 (2001), p. 218].

²⁸⁸ Qui considère qu'une réponse négative entraînerait une réfutation de l'optimisme rationaliste lui-même [Herbrand 1931a, p. 213-214].

²⁸⁹ Il est d'ailleurs à noter qu'il associe lui-aussi cette position à une crainte de la perte d'intérêt des mathématiques (en cas de décidabilité), en assimilant donc décidabilité « en principe » et « décidabilité » « en pratique » [von Neumann 1927, cité dans Gandy 1988, p. 62] (alors que la distinction est effectuée dès 1921 par Heinrich Behman au sein de l'école de Hilbert [Mancosu, Zach 2015, p. 167-168]).

Observons que cette exigence de clôture déductive [complétude syntaxique] ne va pas aussi loin que celle de *décidabilité* de toute question de la théorie [des nombres]. Cette dernière propriété signifie qu'il doit exister un procédé permettant de décider pour toute paire d'assertions contradictoires de la théorie laquelle des deux est démontrable.

[Bernays 1930/1976 (2003), p. 80, n. 1 ; le commentaire entre crochets est nôtre]

Deux explications sont alors possibles : soit les membres de l'école de Hilbert, et Bernays en particulier, avaient de bonnes raisons et la contradiction n'est qu'apparente, c'est ce que soutient Robin Gandy [Gandy 1988, p. 63], soit ils n'étaient pas conscients de la contradiction faute de disposer à cette époque d'un critère d'effectivité précis²⁹⁰, c'est ce que soutient Jacques Dubucs [Dubucs 1991, p. 35]. L'argument de Robin Gandy est que Bernays avait à l'esprit qu'en imposant un algorithme de décision uniforme, la décidabilité était plus contraignante que la complétude syntaxique qui n'impose pas cette uniformité. Mais cela revient justement à ne pas avoir vu que si la complétude syntaxique est acquise, un algorithme de décision en découle aussitôt (par l'énumération récursive de tous les énoncés et l'application successive du critère de complétude syntaxique à chacun d'entre eux), ce qui est très exactement l'argument mentionné par Jacques Dubucs [Dubucs 1991, p. 35]. L'ambiguïté de l'école de Hilbert quant à la différence entre les deux notions apparaît de manière flagrante dans l'un des derniers écrits de Herbrand, qui assimile « le cas le plus général de l'*Entscheidungsproblem* » à une expression coextensive à la complétude syntaxique, la saturation [Herbrand 1931a, p. 211], ou la complétude sémantique par rapport à un modèle [Herbrand 1931a, p. 213]. En nous appuyant sur le fait que Hilbert et Bernays soutiennent encore en 1934 : « En conséquence, nous sommes très éloignés d'une solution générale du problème de la décision » [Hilbert, Bernays 1934 (2001), p. 193], ce qui suggère la conviction d'une réponse positive, nous admettrons que ceux-ci au moins étaient bien convaincus de la décidabilité de Z_2 ²⁹¹ (donc de PA), ce qui recouvre le second point du second renforcement que nous avons évoqué. Finalement, si l'on prend en compte à la fois : la conviction générale du refus de l'*ignorabimus*, l'expression du Programme et la conviction de la décidabilité de Z_2 , on est bien conduit de la part de l'école de Hilbert à la

²⁹⁰ Il a en effet fallu attendre 1936 et la thèse de Church-Turing (qui assimile les fonctions intuitivement calculables aux fonctions récursives générales, nous y reviendrons en détail au § 5.4), pour que l'on traduise la décidabilité par le fait que les théorèmes constituent un ensemble récursif (et pas seulement récursivement énumérable) des énoncés possibles du système formel.

²⁹¹ Si une position opposée a bien été exprimée, concernant l'analyse et la théorie des ensembles, elle l'a été seulement en 1939, dans un paragraphe consacré à l'élargissement du cadre de la méthode de la théorie de la preuve pour tenir compte des différents résultats « négatifs » survenus depuis [Hilbert, Bernays 1939 (2001), p. 342-343])

fin des années 1920, à la forme la plus forte (finitiste) de l'optimisme rationaliste₂. Il est clair que si le Programme avait pu être mené à bien et que Z_2 avait pu être montré syntaxiquement complet au sein de PRA, ce résultat aurait indéniablement constitué une justification de la manière la plus forte (finitiste) de l'optimisme rationaliste₂ pour l'arithmétique.

Quel est alors l'impact des résultats « négatifs » obtenus sur cette justification, en particulier sur PA : son incomplétude syntaxique (premier théorème de Gödel de 1931) ? Bien entendu, cette forme particulièrement forte de l'optimisme rationaliste est d'abord tributaire des contraintes du Programme lui-même, qui inclut l'exigence de la démonstration finitiste d'existence d'un système formel unique, au moins pour l'arithmétique. L'analyse détaillée que nous avons faite de l'impact des théorèmes de Gödel de 1931 sur ce Programme (§ 2), nous indique déjà les limites de la réalisation du programme au sens large, qui ne permet pas d'aller au-delà de l'arithmétique du 1^{er} ordre. Qu'en est-il alors, au moins pour l'arithmétique du 1^{er} ordre, de l'impact du premier théorème de Gödel de 1931 sur la seconde exigence : la démonstration finitiste de la complétude syntaxique pour les problèmes mathématiques exprimables dans PA ? Pour y répondre, il convient d'abord de revenir à la formulation initiale de l'optimisme rationaliste, qui porte explicitement sur des problèmes *mathématiques*, qu'Hilbert distingue (ou distinguera à partir des années 1920) soigneusement des questions *métamathématiques*. Or nous avons montré que tous les énoncés indécidables arithmétiques découverts à ce jour, sont en fait des codages déguisés de propriétés métamathématiques (§ 4.3). Si l'on admet que de tels codages ne peuvent être de ce fait raisonnablement considérés comme d'authentiques problèmes mathématiques (même s'ils en ont éventuellement l'apparence), il en résulte que l'existence de ces énoncés indécidables ne peut pas être retenue comme une réfutation de l'optimisme rationaliste sous sa forme la plus forte. Qui plus est si l'on accepte la thèse d'Isaacson, qui conduit à l'affirmation de la *complétude* syntaxique de PA pour les énoncés authentiquement arithmétiques (§ 4.4), on est conduit au contraire à une *confirmation* de l'optimisme rationaliste sous sa forme la plus forte, au moins pour PA. On n'en est cependant pas parvenu pour autant à la démonstration de l'optimisme rationaliste qu'espérait Hilbert. En effet, la nécessité de recourir à la thèse d'Isaacson correspond alors *de facto* à un engagement philosophique qui s'oppose à l'exigence de neutralité philosophique de notre critère 4 (et du finitisme de Hilbert). De fait, ce n'est pas dans le sens d'une confirmation de l'optimisme rationaliste que la plupart des commentateurs, à commencer par Gödel lui-même (mais également Herbrand [Herbrand 1931a, p. 213 – 214]), ont interprété ces résultats négatifs.

Gödel s'est en effet intéressé à l'optimisme rationaliste de Hilbert dès son premier travail publié : sa thèse de 1929²⁹² (sur la complétude sémantique par rapport à l'ensemble des modèles²⁹³ de la logique des prédicats du premier ordre) : [Gödel 1929 (1986)]. Il y est ensuite revenu en 1938²⁹⁴ : [Gödel 193? (1995)] puis en 1961 [Gödel 1961/? (1995)], sans qu'il n'ait publié aucune de ces réflexions. En 1929 Gödel reprend la formulation classique de l'optimisme rationaliste, mentionnant explicitement la possibilité d'une résolution négative par la preuve d'une impossibilité [Gödel 1929 (1986), p. 62-63]. En 1938, il s'appuie sur l'affirmation par Hilbert qu'une preuve de cette position peut être donnée pour en proposer une version extrêmement forte, qui le conduit à une requête de complétude syntaxique ce qui correspond à la version informelle de notre optimisme rationaliste₂ :

Vous connaissez tous les mots fameux de Hilbert selon lesquels tout mathématicien est convaincu que pour toute question mathématique précisément formulée, une réponse unique peut être trouvée [...] il a même pensé qu'une preuve mathématique pouvait en être donnée, au moins dans le domaine de la théorie des nombres. Comment pouvons-nous imaginer qu'une telle preuve puisse même être obtenue ? [...] Pour tout homme sans préjugé, cela ne peut signifier que ceci : étant donné une proposition mathématique arbitraire A, il existe une preuve soit de A soit de non A, où par « preuve », il est entendu quelque chose qui débute avec des axiomes évidents et procède par inférences évidentes. [Gödel 193? (1986), p. 164]

Il interprète alors son propre résultat de 1931 comme l'infirmité d'une démonstration de la sorte, qu'il traduit quant à lui par le fait qu'une telle justification doit se situer en deçà de tout formalisme, tout en précisant que l'optimisme rationaliste lui-même n'est pas remis en cause pour autant :

[...], alors notre problème devient susceptible d'un traitement mathématique et la réponse s'avère négative même dans le domaine de la théorie des nombres. Mais il est clair que cette réponse négative peut avoir deux significations différentes : 1) elle peut signifier que le problème dans sa formulation originale a une réponse négative, ou 2) elle peut signifier que lors de la traduction de l'évidence au formalisme, quelque chose a été perdu. Il peut être montré facilement que c'est la seconde signification qui est la solution [...]. Ainsi, le résultat est plutôt qu'il n'est pas possible de formaliser l'évidence

²⁹² Une précision bibliographique : la discussion concernant l'optimisme rationaliste de Hilbert se trouve en introduction de sa thèse : [Gödel 1929 (1986)], mais cette partie a été supprimée lors de la publication de l'article correspondant : [Gödel 1930 (1972)].

²⁹³ Nous rappelons que la complétude sémantique *par rapport à l'ensemble des modèles* est définie de la manière suivante (avec les notations métamathématiques modernes habituelles, où « \models » renvoie à la conséquence logique et « \vdash » à la dérivabilité) : pour tout énoncé φ du système formel, si $\models \varphi$ alors $\vdash \varphi$.

²⁹⁴ Martin Davis [Davis 1995, p. 163] fait état de 1938 comme année vraisemblable de rédaction de [Gödel 193? (1990)].

mathématique même dans le domaine de la théorie des nombres, mais la conviction dont parle Hilbert reste entièrement non affectée. [Gödel 193? (1986), p. 164]

En 1961 il reste sur la position de 1938, tout en ajoutant l'idée qu'une formalisation sera toujours possible mais avec de « nouveaux axiomes », c'est-à-dire au moyen d'un système formel spécifique dont les caractéristiques doivent être trouvées, nécessairement d'une manière non mécanique²⁹⁵ (il évoque pour cela le recours à la phénoménologie de Husserl) [Gödel 1961/? (1995), p. 385]. Cela revient à réduire l'exigence d'optimisme rationaliste₂, pour se ramener à notre optimisme rationaliste₁.

Au-delà de la tentative de Hilbert de prouver l'optimisme rationaliste, il vient naturellement la question de savoir si celui-ci pourrait être justifié ou surtout réfuté par une autre voie. Si l'on ne se fixe aucune contrainte sur les méthodes de résolution acceptées pour résoudre un problème mathématique la question risque fort d'être vide de sens comme le remarquait déjà Gödel en 1929 [Gödel 1929 (1986), p. 64-65] et comme cela a également été rappelé ensuite [Fraenkel, Bar-Hillel 1958, p. 314], [Feferman 2006, p. 149-150] et [Koellner 2006, p. 154]. Ainsi que le souligne fort justement Andrzej Mostowski [Mostowski 1952, p. 3], pour espérer obtenir un tel résultat (fût-il dans le sens d'une réfutation) à partir des différents théorèmes obtenus sur les systèmes formels, il faudrait d'abord que l'on puisse construire (ou s'entendre sur) un système formel à même de rendre compte de l'ensemble des mathématiques. Le premier théorème de Gödel de 1931, formulé sous la forme de l'incomplétude sémantique par rapport au modèle²⁹⁶ naturel de PA, semble interdire cet espoir même en se limitant à l'arithmétique, si l'on omet de distinguer comme nous l'avons fait, énoncé mathématique et codage d'énoncé métamathématique. Gödel et à sa suite Jean Ladrière [Ladrière 1957, p. 410-411] ainsi que d'autres, en ont tiré la conclusion que le recours se situait en-deçà de la formalisation, fût-ce en faisant appel à une extension sans limite de systèmes formels de plus en plus puissants [Myhill 1960, p. 465] ou à l'utilisation de modèles non standards [Webb 1980, p. 116-117]. De telles solutions, qui s'apparentent clairement à une fuite en avant, impliquent cependant un certain engagement philosophique (anti-intuitionniste) et achoppent elles-aussi sur la condition d'absence d'engagement philosophique incompatible avec l'exigence d'intersubjectivité de

²⁹⁵ Il faut sans doute préciser qu'en 1961 comme en 1938, il entend restreindre les problèmes mathématiques auxquels s'applique l'optimisme rationaliste à ceux « susceptibles d'une réponse par oui ou par non ». On ne voit cependant pas très bien quel type de problème mathématique bien posé pourrait ne pas s'y ramener.

²⁹⁶ Nous rappelons que la complétude sémantique *par rapport à un modèle* est définie de la manière suivante : il existe au moins un modèle tel que tout énoncé vrai dans ce modèle est dérivable. On montre encore que d'une manière générale (sans se limiter au 1^{er} ordre), cette forme de complétude est coextensive à la complétude syntaxique.

notre critère 4. Elles ne sauraient donc être non plus interprétées comme des démonstrations de l'optimisme rationaliste.

5.3 L'affirmation, en Théorie Algorithmique de l'Information, d'une limite au domaine des vérités mathématiques accessibles

Des résultats précis ont été obtenus en Théorie Algorithmique de l'Information, qui semblent conduire à une limitation forte de la connaissance en mathématique. Ces résultats sont des extensions du théorème de Turing de 1936-1937, obtenues en exploitant la notion de complexité algorithmique²⁹⁷ (ou complexité de Kolmogorov ou encore complexité de Chaitin-Kolmogorov). Ils s'expriment principalement sous la forme de trois théorèmes :

- 1) pour établir la complexité algorithmique d'une suite binaire finie donnée : s , il faut un système formel dont la complexité algorithmique est au moins égale à celle de s ([Chaitin 1974]) ;
- 2) pour établir les n premiers bits du développement binaire d'un nombre Ω_U ²⁹⁸, il faut un système formel dont la complexité algorithmique est au moins égale à n ([Chaitin 1987a]) ;
- 3) la probabilité qu'un énoncé de longueur n soit dérivable dans un système formel tend vers zéro quand n tend vers l'infini, tandis que la probabilité que cet énoncé soit vrai (au sens intuitif) reste strictement positive ([Calude, Jürgensen 2005]).

²⁹⁷ La notion de complexité algorithmique a été proposée indépendamment par Andreï Kolmogorov en 1965 et par Grégory Chaitin en 1966. Nous rappelons que la complexité algorithmique d'une suite binaire finie quelconque : s , est, par définition, la taille (longueur) en bits du plus petit (court) programme qui engendre s sur une machine de Turing universelle, sans tenir compte du nombre d'étapes nécessaires. Elle est toujours exprimée à une constante additive près, liée au codage de la machine de Turing universelle utilisée. Elle est définie de deux manières possibles, soit pour des programmes limités se terminant par des « blancs » auquel cas elle est en général notée $K(s)$, soit pour des programmes auto-délimités (se terminant nécessairement par l'instruction « fin ») auquel cas elle est en général notée $H(s)$. Elle est caractérisée dans les deux cas par deux théorèmes importants : le théorème d'invariance et le théorème de majoration. Le théorème d'invariance indique que si u et v désignent deux langages de programmation donnés, on a : $\forall s, |K_u(s) - K_v(s)| < c_{uv}$ (respectivement : $|H_u(s) - H_v(s)| < c_{uv}$), où c_{uv} est une constante ne dépendant que des deux langages u et v . Le théorème de majoration est le suivant : $\forall s, \exists c', K(s) < \text{longueur}(s) + c'$ (respectivement : $H(s) < \text{longueur}(s) + 2\lceil \log_2[\text{longueur}(s)] \rceil$), où c' est une constante, fonction du langage de programmation utilisé, traduisant la longueur du codage de l'instruction : « imprimer s ».

²⁹⁸ Les nombres Ω_U ont été introduits par Grégory Chaitin en 1974. Ils sont la probabilité qu'une machine de Turing universelle U (à arrêt auto-délimité) s'arrête (à la suite d'un nombre fini d'étapes). Ils dépendent donc du codage de la machine universelle utilisée. Ce sont des nombres réels (compris entre 0 et 1), présentant la propriété particulière de « non-calculabilité » dans le sens où les nombres réels « calculables » sont la limite d'une suite : $x_n = p_n/q_n$, infinie croissante de nombres rationnels calculable par une machine de Turing universelle, avec la condition de convergence : $\forall n, |x_n - x| < 1/2^n$.

Du premier de ces résultats, Grégory Chaitin a induit un « principe heuristique » de portée plus générale : « Un ensemble d'axiomes de complexité algorithmique n ne peut pas produire un théorème de complexité algorithmique (substantiellement) supérieure à n » [Chaitin 1992, p. 68]²⁹⁹. Il l'a exprimé sous la forme d'un slogan souvent cité : « Je voudrais être en mesure de dire que si l'on a 10 livres [environ 5 kg] d'axiomes et un théorème de 20 livres [environ 10 kg], alors ce théorème ne peut être dérivé de ces axiomes » [Chaitin (1982), p. 300 – 301 ; les commentaires entre crochets sont nôtres]. Il a ensuite interprété ces résultats sous la forme d'une double affirmation de grande portée philosophique : 1) Le ressort des résultats d'incomplétude (premier théorème de Gödel de 1931) et d'indécidabilité (théorème de Turing de 1936-1937) se trouve dans les limites mises en évidence au moyen de la notion de complexité algorithmique et est lié à la présence de hasard authentique au sein des mathématiques, 2) Il en résulte une limite absolue au domaine des vérités mathématiques (au sens intuitif) accessibles [Chaitin 2005 (2009), p. 64] et [Chaitin 2005 (2009), p. 197]. La force de cette seconde affirmation apparaît dans les titres mêmes de ses ouvrages et articles récents : « Les limites des mathématiques » ([Chaitin 1998]), « L'inconnaissable » ([Chaitin 1999]), « Les limites de la raison » ([Chaitin 2006]). A cela s'ajoute l'usage réitéré du terme « dramatic » (avec son double sens possible en anglais de spectaculaire et de dramatique) pour qualifier les résultats obtenus, dont il est affirmé qu'ils y conduisent [Davis 1978, p. 265]. D'une manière assez surprenante, un tel enthousiasme immodéré a été repris par les différents commentateurs de ces résultats³⁰⁰ sans prise de distance, voire même avec une surenchère dans l'usage des superlatifs, par exemple : « L'ultime en indécidabilité » [Stewart 1988]. Il faut en fait attendre 1989 pour une première critique sérieuse : [Lambalgen 1989], puis la fin des années 1990 (soit plus de vingt ans après la publication des premiers de ces résultats), pour des critiques plus incisives : [Fallis 1996] et surtout [Raatikainen 1998] et [Raatikainen 2000]³⁰¹, qui sont principalement au nombre de trois.

La première critique, qui s'adresse au principe heuristique de Grégory Chaitin, est élémentaire : tout système formel finiment axiomatisé aura nécessairement une complexité

²⁹⁹ Il rappelle indiscutablement une remarque bien antérieure de Gödel, beaucoup plus circonscrite et non publiée : « On peut prouver que pour résoudre tous les problèmes de type Goldbach $[P_1]$ d'un certain degré de complexité k , on a besoin d'un système d'axiomes dont le degré de complexité sous réserve d'une correction mineure est $\geq k$ » [Gödel 1953 (1995), p. 32, n. 39].

³⁰⁰ Par exemple Martin Davis [Davis 1978, p. 286].

³⁰¹ On peut mentionner également les commentaires de Jaakko Hintikka et Torkel Franzel [Hintikka 2000, p. 29] et [Franzen 2005, p. 142-148]. Par ailleurs Jean-Paul Delahaye soutient en 1994 plusieurs remarques qui anticipent [Raatikainen 1998] et [Raatikainen 2000], mais qui n'ont pas été exploitées en ce sens par l'auteur [Delahaye 1994, p. 168, 172].

algorithmique (évaluée sur ses axiomes) finie³⁰², alors qu'un tel système formel permet évidemment de dériver des théorèmes dont la complexité est aussi grande que l'on veut (par exemple : « $n = n$ »), ce qui invalide le principe. Cette critique est essentiellement celle de [Fallis 1996], mais elle avait été anticipée par Grégory Chaitin lui-même dans la seule discussion critique de ses propres résultats qu'on lui connaisse : [Chaitin 1992, p. 68-70]. La réponse proposée par Grégory Chaitin est que le principe heuristique peut être exprimé sous une forme affaiblie : « Un ensemble d'axiomes de complexité algorithmique n ne peut pas produire un théorème *qui affirme qu'un objet spécifique est de complexité algorithmique (substantiellement) supérieure à n* ». Cette réponse renvoie alors directement à la deuxième critique. Cette deuxième critique, qui émane de Torkel Franzen [Franzen 2005, p. 143], porte sur la filiation qui conduit des deux résultats obtenus : [Chaitin 1974] et [Chaitin 1987], au principe heuristique (non affaibli) qui en est inféré. Ces résultats se rapportent en effet à la complexité algorithmique (directement pour le premier et indirectement, via la construction des nombre Ω_U , pour le second) et non à des théorèmes quelconques³⁰³. Ils conduisent donc seulement à une limitation de l'aptitude à déterminer la complexité algorithmique elle-même et non à une limitation de la capacité d'un système formel à dériver des théorèmes, ce qui n'a évidemment pas la même portée. La portée réelle du principe heuristique affaibli proposé par Grégory Chaitin dépend alors de la pertinence de l'utilisation de la complexité algorithmique pour caractériser la quantité de théorèmes que peut dériver un système formel, ce qui amène à la troisième critique. La pertinence de l'utilisation de la complexité algorithmique pour caractériser la force ou la capacité d'un système formel est toujours implicite dans les textes de Grégory Chaitin et de Cristian Calude³⁰⁴ (l'autre grande figure moderne de la Théorie Algorithmique de l'Information). C'est sa contestation qui constitue la troisième critique, soulevée dans [Lanbalgen 1989] et traitée de manière détaillée dans [Raatikainen 1998] et [Raatikainen 2000]. Ce que montre Panu Raatikainen est qu'il n'y a justement pas de lien direct entre, d'une part la force ou la capacité relative de deux systèmes formels traduite de manière précise par la quantité respective de théorèmes qui peuvent en être dérivés, et d'autre part la complexité algorithmique de ces systèmes formels évaluée sur la base de leur axiomes (et

³⁰² Dans le cas de systèmes formels contenant un schéma d'axiomes, comme PRA et PA, il y a une difficulté à résoudre pour déterminer la complexité algorithmique associée.

³⁰³ Qui plus est, comme le fait justement remarquer Don Fallis, le résultat de Chaitin de 1974 porte non pas sur les théorèmes d'un système formel, mais sur toute expression de ce système formel qui peut donc être également un énoncé réfutable ou indécidable (et même une expression mal formée) [Fallis 1996, p. 265-266].

³⁰⁴ Au-delà de la reconnaissance que si la théorie algorithmique de l'information indique qu'il y a beaucoup d'indécidables, elle n'en fournit explicitement aucun comme le signale Jean-Paul Delahaye [Delahaye 1994, p. 119]. Ce point est à relier au caractère intrinsèquement non effectif de la détermination de la complexité algorithmique, même si celle-ci peut être majorée (à une constante près) via le théorème de majoration.

éventuellement de leurs règles d'inférence). Techniquement, cela provient de la relativité de la constante qui intervient dans la définition de la complexité algorithmique appliquée à un système formel et qui dépend du double codage nécessaire utilisé, de la codification gödelienne (pour exprimer les axiomes du système formel sous la forme d'une suite binaire), et de la machine de Turing universelle. Selon le choix (arbitraire) retenu pour ce double codage, la complexité algorithmique relative de deux systèmes formels peut être inversée, ou conduire au résultat contre intuitif qu'avec n'importe quel codage raisonnable, celle de PRA sera par exemple supérieure à celle de Z_2 ³⁰⁵ (dont les axiomes et les règles d'inférence s'expriment plus simplement). Ce résultat s'étend³⁰⁶ (pour la même raison) à l'évaluation des nombres Ω_U (qui dépend de la machine de Turing universelle retenue) et partant aux résultats qui en sont issus. Une telle conclusion est évidemment dévastatrice pour l'interprétation proposée par Grégory Chaitin. La réponse de Cristian Calude, qui n'en conteste pas le fond [Calude 2002b, p. 508], est donnée dans [Calude, Jürgensen 2005]. Ce que ce dernier résultat montre³⁰⁷, traduit par le troisième théorème que nous avons cité, est que lorsque la longueur des énoncés tend vers l'infini l'effet de relativité de la constante s'estompe et l'on retrouve alors, à la limite, le principe heuristique non affaibli. C'est évidemment pour tomber de Charybde en Scylla : quel sens et quel intérêt mathématique peut avoir un théorème dont l'énoncé possède (ou même seulement tend vers) une longueur infinie ? Pour tout énoncé mathématiquement intéressant, dont la longueur est nécessairement finie et souvent faible, les résultats de la Théorie Algorithmique de l'Information ne savent donc rien nous dire³⁰⁸, quand bien même il s'agirait d'énoncés indécidables que l'on sait établir tels par ailleurs³⁰⁹.

Que reste-il finalement de la position philosophique forte soutenue par Grégory Chaitin (ainsi que Cristian Calude) ? Manifestement peu de choses. En se limitant au principe heuristique affaibli, non contestable, qui ne portent que sur les théorèmes qui concernent

³⁰⁵ Jean-Paul Delahaye aborde incidemment le problème en suggérant d'utiliser la profondeur logique de Bennett : [Bennett 1993 (1995)] plutôt que la complexité algorithmique, pour caractériser la force d'un système formel [Delahaye 1994, p. 118]. Nous n'avons cependant pas connaissance d'une tentative d'exploitation de cette idée.

³⁰⁶ Cette analyse est cohérente avec le cas particulier de Solovay en 2000, qui, selon Cristian Calude, produit un codage particulier ne permettant de connaître (avec certitude) *aucun* bit du nombre Ω_U associé à ce codage [Calude 2002a, p. 276]. Cependant Cristian Calude, qui fait mention de ce lien, n'en tire pas la même conclusion que Panu Raatikainen [Calude 2002a, p. 270].

³⁰⁷ En introduisant pour cela une nouvelle grandeur (afin d'éviter la première critique citée) : la *différence* entre la complexité algorithmique $H(s)$ et la longueur de la chaîne elle-même : s .

³⁰⁸ Jean-Paul Delahaye remarque que dans la mesure où les énoncés mathématiques intéressants soient confinés dans les décimales des nombres Ω_U , cela revient à ce que les énoncés mathématiquement intéressants soient confinés dans les premiers digits de Ω_U , ce qu'il qualifie de « réductionnisme involontaire extrême » [Delahaye 1994, p. 169].

³⁰⁹ A l'instar des énoncés du premier ou du second théorème de Gödel de 1931, contre Don Fallis [Fallis 1996, p. 264], comme le reconnaissent Cristian Calude et Helmut Jürgensen [Calude, Jürgensen 2005, p. 3] (que nous ne pouvons que suivre sur ce point).

l'aptitude à établir la complexité algorithmique, il subit frontalement l'objection de [Raatikainen 1998] selon laquelle la complexité algorithmique du système formel n'est pas représentative de l'aptitude de celui-ci à dériver des théorèmes. Le recours à la Théorie Algorithmique de l'Information ne permet donc pas de juger de la quantité de théorèmes qu'un système formel peut dériver, donc de la force de celui-ci. Le recours récent proposé par Cristian Calude et Helmut Jürgensen : [Calude, Jürgensen 2005] ne résout rien, puisqu'il ne s'applique qu'aux énoncés de longueur extrême qui n'ont rien à voir avec les énoncés et les problèmes utilisés par les mathématiciens. La double thèse défendue (bruyamment) par Grégory Chaitin à longueur d'articles selon laquelle, d'une part l'origine des phénomènes d'incomplétude et d'indécidabilité est à relier au hasard en mathématiques, d'autre part il en résulte une limite au domaine des vérités arithmétiques accessibles, est donc clairement réfutée³¹⁰. Elle ne saurait donc conduire à réfuter l'optimiste rationaliste de Hilbert sous une forme renforcée (optimisme rationaliste)³¹¹. Bien entendu, cette remarque ne remet pas en cause la validité *technique* des résultats obtenus en Théorie Algorithmique de l'Information, ni même l'intérêt philosophique, réel, mais modeste³¹², de ceux-ci.

5.4 La thèse de Church-Turing et les tentatives de dépassement de celle-ci

Etant donné que les FRG, totales et partielles, interviennent de manière cruciale dans la thèse de Church-Turing, quelques précisions techniques préalables à leur sujet sont utiles. Nous avons déjà introduit les FRG au § 1.2 (le qualificatif de total étant omis lorsqu'il n'y a pas de risque de confusion avec les FRG partielles). Les FRG partielles ont été introduites par Stephen Kleene en 1938. Elles constituent une classe de fonctions d'entiers encore plus large que celle des FRG

³¹⁰ Cette réfutation s'étend évidemment aux extrapolations qui en ont été faites sur l'inaccessibilité de la certitude et les limites de la raison, par exemple [Brisson, Meyerstein 1995, p. 179, 234-235].

³¹¹ Il convient également de signaler que les résultats obtenus sur la base de la Théorie Algorithmique de l'Information le sont toujours dans un cadre ensembliste à partir d'une démarche non effective (à l'instar de [Bienvenu et al. 2014] qui conduit à des théories dont tous les indécidables Π_1 sont éliminés, mais en faisant alors appel à un oracle « à la [Turing 1939] »). En ne répondant ainsi pas aux critères définissant l'optimisme rationaliste₁, ils sont (d'une manière bien plus générale) non recevables dans le cadre de la présente discussion.

³¹² Selon Panu Raatikainen, ces résultats ne sont qu'une conséquence du (ou une autre manière de traduire le), résultat classique de Turing sur l'indécidabilité du problème de l'arrêt [Raatikainen 2000, p. 218]. Une telle position est certainement réductrice, dans la mesure où ces résultats nous disent indiscutablement quelque chose sur les implications philosophiques de l'indécidabilité et de la vérité en mathématiques, même si c'est sur des bases qui vont bien au-delà des exigences de Hilbert. Il n'en reste pas moins que les réponses apportées restent tributaires des engagements philosophiques (ontologiques et épistémologiques) sous-jacents, qui en limitent d'autant la portée.

totales, puisqu'il est autorisé que la fonction ne fournisse pas de résultat pour certaines valeurs de l'argument³¹³. Si l'on s'intéresse maintenant aux *ensembles* respectifs des FRP, des FRG totales et des FRG partielles, on a alors les résultats importants rassemblés dans le tableau VI, que nous allons utiliser dans ce qui suit.

Tableau VI : Propriétés des *ensembles* des FRP, des FRG totales et des FRG partielles

Ensemble	Sujet à l'argument diagonal	Enumérable par programme
{FRP}	OUI ³¹⁴	OUI
{FRG totales}	NON	NON ³¹⁵
{FRG partielles}	NON ³¹⁶	OUI

Historiquement, c'est en 1936 que paraissent presque simultanément quatre propositions : [Church 1936a], [Kleene 1936]³¹⁷, Post [1936]³¹⁸ et [Turing 1936-1937], qui affirment toutes, selon la présentation standard, l'identification entre :

- 1) l'idée *intuitive* d'effectivité³¹⁹, d'algorithme aboutissant toujours, de calculabilité à aboutissement certain³²⁰ ;

³¹³ Nous rappelons qu'un ensemble d'entiers est *récurif* s'il peut lui être associé une fonction caractéristique FRG totale, ce qui est associé à la décidabilité, tandis qu'un ensemble d'entiers est *récurivement énumérable* s'il peut lui être associé une fonction caractéristique FRG partielle, ce qui est associé à la semi-décidabilité. Intuitivement, un ensemble d'entiers est récurivement énumérable s'il est possible de définir un algorithme qui produit une liste de ses éléments *non ordonnée* et possiblement avec répétition. Un ensemble récurivement énumérable est récurif si et seulement si son complément par rapport à \mathbb{N} est récurivement énumérable.

³¹⁴ Historiquement, c'est en utilisant une construction diagonale qu'Ackermann a obtenu la première fonction (qui porte son nom), récurive mais non FRP (§ 4.1).

³¹⁵ C'est ce résultat qui entraîne l'impossibilité d'utiliser l'argument diagonal à l'ensemble des FRG totales et partant, l'impossibilité d'étendre de nouveau de cette manière cet ensemble (en maintenant le critère que la fonction doit être calculable pour toute valeur de l'argument). C'est lui qui a conduit Judson Webb à évoquer, un « champ protecteur invisible » à propos de l'incomplétude syntaxique (en toute rigueur ce serait plutôt l'indécidabilité) d'un système formel répondant aux conditions d'applications des théorèmes de Gödel de 1931, et des « anges gardiens » à propos des énoncés indécidables correspondants, qui sont une conséquence directe de ce résultat [Webb 1980, p. 202-203, 208].

³¹⁶ Dans le cas de l'ensemble des FRG partielles, si l'argument diagonal ne peut s'appliquer, ce n'est évidemment pas par suite de l'impossibilité de l'énumérabilité par programme (puisque'il est au contraire énumérable par programme), mais pour des raisons techniques internes.

³¹⁷ Exploitant une remarque formulée par Gödel en 1934, sur la base d'une suggestion que celui-ci attribuait à Herbrand [Sieg 1994, p. 82]. Rappelons à ce sujet le fait historique bien connu qu'en 1934 Gödel n'a pas perçu cette thèse parce qu'il n'était pas convaincu que les fonctions récurives « de la manière la plus générale » coïncidaient avec les FRG totales, faute d'avoir réalisé l'impossibilité d'étendre l'ensemble des FRG totales par diagonalisation (voir le tableau VI) [Gandy 1988, p. 64] et [Dubucs 1991, p. 48-49].

³¹⁸ Sachant que les résultats correspondants ont vraisemblablement été obtenus dans les années 1920, mais n'ont été publiés qu'en 1936, après la formulation de la thèse par Church [Gandy 1988, p. 92].

³¹⁹ En fait depuis [Herbrand 1931a], l'expression : « effectivement calculable » est utilisée pour rendre : « intuitivement calculable ».

³²⁰ Une formulation un peu plus précise, mais toujours informelle, consiste à dire qu'une procédure effectivement (intuitivement) calculable est une procédure figée, permettant de répondre à la question posée de façon uniforme et mécanique, sans effort d'invention et sans rien attendre de l'intuition et de l'intelligence.

2) une notion mathématique précise de fonctions, respectivement, λ -définissables, récursives générales totales (FRG totales), finitaires au sens de Post et Turing-calculables qui se terminent, qui ont été ensuite montrées équivalentes entre elles ainsi qu'à d'autres fonctions encore proposées ultérieurement, et ce de manière extensionnelle.

Cette identification, qui a reçu le nom de thèse de Church-Turing (ou thèse de Church), permet de disposer d'un critère d'effectivité mathématiquement précis. Comme ce critère intervient dans la définition de la notion de décidabilité d'un système formel, si la définition intuitive d'effectivité est suffisante pour l'établissement de la décidabilité, une telle précision était nécessaire pour pouvoir en établir *l'indécidabilité*. C'est ainsi que l'indécidabilité du calcul des prédicats du 1^{er} ordre (et donc de PA) a pu être établie, en réponse à *l'Entscheidungsproblem*, peu après la thèse énoncée. En 1938 Kleene a proposé une thèse plus large, quelquefois nommée thèse de Kleene, en supprimant l'exigence d'un aboutissement certain dans la formulation de la partie intuitive de l'identification et en faisant intervenir les FRG *partielles* au lieu de totales et de Turing-calculables (qui ont leur équivalent dans les autres formulations) : [Kleene 1938]. S'il n'est évidemment pas possible de rendre compte de la notion de décidabilité d'un système formel de cette manière, son intérêt réside dans le fait que la notion de calcul la plus naturelle n'implique pas que celui-là aboutisse toujours, comme Gödel l'a souligné (selon Hao Wang) [Webb 1980, p. 223-224]. C'est ainsi qu'une machine de Turing prise sans restriction fournit nécessairement des FRG partielles (ce qui se traduit par le résultat sur l'indécidabilité de l'arrêt). Par ailleurs, sur un plan philosophique, il n'y a alors pas de différence de position entre les points de vue classique et intuitionniste en mathématiques concernant la thèse de Kleene, ce sans aucune restriction sur les FRG partielles prises en compte³²¹.

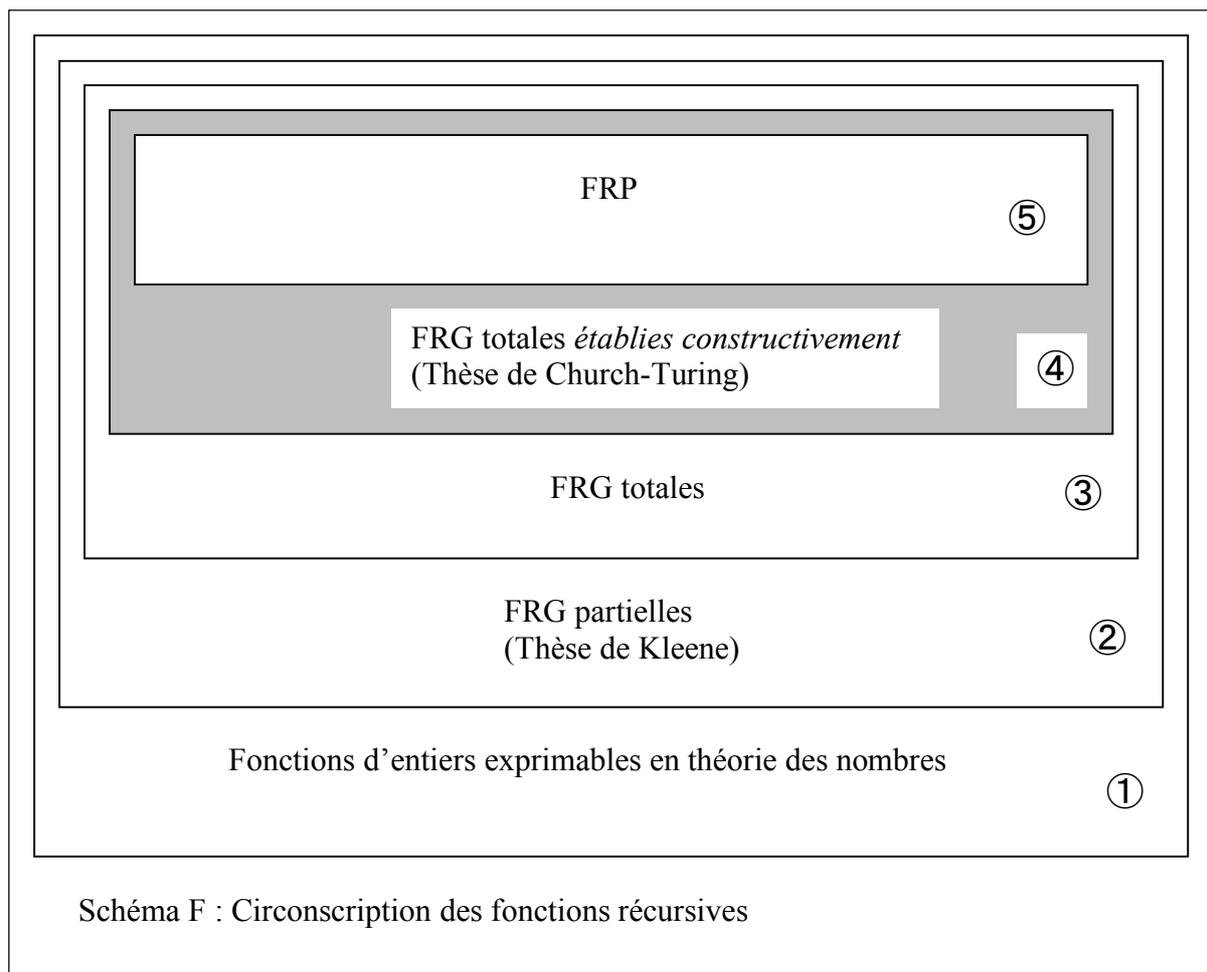
La situation est différente dans le cas de la thèse de Church-Turing : pour rendre compte de la manière la plus naturelle de la notion de fonction effectivement (intuitivement) calculable qui intervient dans la thèse de Church-Turing, il est apparu nécessaire de préciser l'identification précédente en introduisant une *contrainte* sur les FRG *totales* retenues (ou leur équivalent en fonctions Turing-calculables ou λ -définissables). Cette contrainte est déjà présente (mais discrète) dans l'article séminal de Church³²², et traduit l'exigence que les fonctions retenues soient définies *de manière constructive* :

³²¹ La généralisation par Kleene est la forme de la thèse de Church-Turing qui est acceptée à la fois par Gödel et par Georg Kreisel [Webb 1980, p. 223-224] (point sur lequel nous reviendrons au § 6.1).

³²² En contradiction avec ce qui est affirmé par Elliott Mendelson et Robert Black [Mendelson 1990, p. 226, n. 1] et [Black 2000, p. 247-249].

Le lecteur peut objecter que cet algorithme ne peut pas être utilisé pour fournir un calcul effectif de la valeur particulière exigée de F^i tant qu'une preuve constructive de la relation exigée $f_{ni}^i(k_1^i, k_2^i, \dots, k_{ni}^i) = k^i$, n'est pas ultimement trouvée. Mais s'il en est ainsi, cela signifie simplement qu'il devrait prendre le quantificateur existentiel qui apparaît dans notre définition d'un ensemble d'équations de récursions, *dans un sens constructif*. Ce que le critère de constructivité doit être, est laissé à la discrétion du lecteur. [Church 1936a, p. 351, n. 10 ; souligné par nous]

Cette contrainte est présentée de manière plus détaillée et associée à un critère qui semble plus précis (nous allons y revenir), dans l'ouvrage de référence de Kleene [Kleene 1952 (2009), p. 317-318, 319]. Une telle position revient à introduire une circonscription, parmi les FRG totales, entre celles dont la démonstration du caractère FRG est constructive et les autres. Cela conduit finalement à une circonscription des fonctions d'entiers en cinq niveaux (voir le schéma F ci-après), où la thèse de Kleene (par rapport aux FRG partielles) est associée au niveau (2), tandis que la thèse de Church-Turing (par rapport aux FRG totales établies constructivement) est associée au niveau (4).



Une conséquence emblématique de cette réduction des FRG totales retenues dans la formulation de la thèse de Church-Turing concerne la fonction d'entiers $f(i)$ définie de la manière suivante : $f(i) = 1$ si la conjecture de Goldbach est vraie et $f(i) = 0$ si la conjecture de Goldbach est fausse³²³. Une telle fonction est rejetée comme ne s'appliquant pas à la thèse de Church-Turing parce que l'on ne dispose pas d'une procédure effective pour la mettre en œuvre, bien qu'elle puisse être considérée (en un sens absolu) comme étant FRG totale, ainsi que le souligne Jeremy Seligman [Seligman 2002, p. 212]. Par ailleurs, outre le caractère naturel de la restriction introduite, celle-ci permet de respecter une neutralité ontologique à l'égard du réalisme en mathématiques, par exemple entre les positions classique et intuitionniste, neutralité que Kleene appelle explicitement [Kleene 1967 (1971), p. 235], contre Elliott Mendelson [Mendelson 1963, p. 202-203] et en contradiction avec ce qui est affirmé par Carol Cleland [Cleland 2006, p. 142-143]. Il n'en reste pas moins que même si celui-ci affirme avoir établi un lien entre l'intuitionnisme (au sens de Brouwer, formalisé par Heyting) et la thèse de Church-Turing [Kleene 1952 (2009), p. 509], il reste à préciser la nature du critère de constructivité retenu comme l'invite Church à la fin de son article séminal [Church 1936a, p. 351, n. 10]. Ce point mérite une discussion et l'on peut se demander si en introduisant la restriction en question Church et Kleene avaient conscience d'ouvrir une boîte de Pandore. En effet, la thèse de Church-Turing a été introduite pour donner un contenu technique précis et univoque à la notion de décidabilité d'un système formel, qui n'est définie que de manière intuitive. Mais si l'on est amené pour cela à introduire une ambiguïté au niveau de la circonscription des FRG prises en compte par le biais de l'exigence de constructivité ajoutée, l'on introduit au mieux une relativité (s'il apparaît plusieurs critères possibles) et au pire une source d'indétermination (si le critère reste intuitif). Commençons par préciser que même dans ce dernier cas il n'en résulterait pas que l'on soit revenu au point de départ ou que l'on n'ait fait que déplacer le problème. En effet, pour les FRG utilisées en pratique par les mathématiciens (à commencer par l'ensemble des FRP) il n'y a pas d'ambiguïté sur leur caractère constructif, et de fait les mathématiciens sont empiriquement d'accord à leur sujet³²⁴. Le problème est donc essentiellement conceptuel et philosophique (comme le montre l'exemple que nous avons donné de la fonction définie par rapport à la vérité de la conjecture de Goldbach). Il se ramène à demander si l'on est en mesure de faire état d'une notion primitive et consensuelle de démonstration *constructive*, ainsi que

³²³ La vérité et la fausseté pouvant être alors définies intuitivement ou par rapport au modèle naturel de PA.

³²⁴ En particulier, il n'a pas été mis en évidence de fonction proprement mathématique (contrairement à celle basée sur la vérité versus fausseté de la conjecture de Goldbach), qui traduirait cette ambiguïté.

l'énonce Göran Sundholm³²⁵ [Sundholm 2014, p. 14]. Il est important sur ce point de bien distinguer cette requête de constructivité du respect de l'ensemble des exigences intuitionnistes, qui incluent également en particulier la prise en compte du sujet créateur au sens de Brouwer (se reporter infra au § 3.1)³²⁶. Il est en effet clair que dans le cadre de la circonscription précise de la thèse de Church-Turing qui est notre objectif, cette dernière exigence, outre la complication qu'elle introduirait, non seulement n'est pas requise, mais doit être *rejetée* en tant qu'engagement philosophique particulier, eu égard à notre critère de neutralité philosophique. Cependant, même en se limitant à la seule requête de constructivité, l'on ne peut que constater que ni Church ni Kleene n'en ont vraiment défini le contenu. Church l'évacue tout simplement dans la note 10 de [Church 1936a] que nous avons citée. Quant à Kleene, dans les remarques plus précises que nous avons mentionnées [Kleene 1952 (2009), p. 317-319], il n'indique pas comment il faut interpréter « prédicat effectivement décidable » et « démonstration effective ». Ce dernier point avait déjà été relevé en 1958 par Heyting³²⁷ [Heyting 1958, p. 106]. En 1967, Kleene propose un nouvel argument :

Nous avons admis sans y regarder de près la réciproque de la thèse de Church : *si une fonction est Turing-calculable (ou récursive générale, ou λ -définissable), alors elle est intuitivement calculable (ou effectivement calculable)*. Si nous avons à défendre cette implication contre un intuitionniste ou bien un constructiviste de quelque autre nuance, qui considère qu'un algorithme n'existe que s'il est démontré, selon ses propres critères, qu'il donne toujours une réponse, il nous suffirait de lui demander ce qui suit : si l'hypothèse qu'une fonction est Turing-calculable est vérifiée selon ses critères de constructivité, la conclusion proposée l'est aussi. Sous cette forme, on ne voit pas bien comment elle pourrait être mise en doute. C'est seulement si on permet une interprétation non constructive pour l'hypothèse et si en même temps on exige une interprétation constructive pour la conclusion, que la réciproque de la thèse de Church peut être tenue pour douteuse. [...]. [Kleene 1967 (1971), p. 249, n. 171]

³²⁵ Le panorama brossé par Göran Sundholm sur cette question est particulièrement éclairant sur le fait que l'on ne soit pas parvenu à ce jour à une solution universellement acceptée par les intuitionnistes [Sundholm 2014, p. 14-28]. La situation est encore complexifiée dans le cas de Georg Kreisel, qui entend de plus rendre compte, à la suite de Gödel, de « rigueur informelle » et de « raisonnement non mécanique » (en introduisant pour cela la notion de définitions « h-effectives » pour « humainement effectives », au-delà des définitions « m-effectives » pour « mécaniquement effectives », ces dernières correspondant à la machine de Turing) : [Kreisel 1967a] et [Kreisel 1972].

³²⁶ Sachant que la prise en compte de ce sujet créateur peut conduire à un débordement partiel du champ intuitionniste au-delà du champ classique, comme cela est mentionné par Martin Löb, Georg Kreisel et Robert Black [Löb 1959, p. 160], [Kreisel 1972, p. 314-315, 321] et [Black 2000, p. 247].

³²⁷ Il semble que ce problème soit ensuite resté ignoré pendant cinquante ans, avant de réapparaître très récemment [Coquand 2014, p. 161] et [Sundholm 2014, p. 13]. On peut noter en particulier la remarque de Thierry Coquand : « La théorie des FRG est habituellement présentée, sans tenir compte de savoir si les résultats sont obtenus en utilisant la logique intuitionniste ou non » [Coquand 2014, p. 163].

Le problème est qu'au-delà du fait que le contenu du critère de constructivité n'y est toujours pas indiqué, l'on ne voit pas en quel sens il serait possible, sans circularité, de faire état de constructivité d'une fonction qui n'est définie que par le fait qu'elle est *intuitivement* calculable. En 1959 Rózsa Péter, qui défend également la nécessité de l'exigence de constructivité introduite par Church, affirmait justement qu'il est impossible d'éviter un cercle vicieux [Heyting 1959, p. 70], [Moschovakis 1968, p. 472] et [Mosconi 2014, p. 48]. Parmi les intuitionnistes, qui sont évidemment directement concernés par la question posée, il semble qu'il n'y ait eu que Bishop qui ait proposé une solution en 1967, en introduisant pour cela un terme primitif spécifique : « règle », mais défini seulement intuitivement ce qui nous ramène à la critique de Rózsa Péter. Finalement, l'indétermination liée à l'ambiguïté de la notion de démonstration constructive (que nous avons déjà souligné au § 3.8), apparaît comme étant sans doute le prix à payer de la restriction introduite sur les FRG (ou les fonctions Turing-calculables) qui interviennent dans la thèse de Church-Turing pour respecter l'exigence d'intersubjectivité.

Nous en venons maintenant à la justification de la thèse de Church-Turing. Les deux arguments présentés classiquement à l'appui de la thèse de Church-Turing sont : d'une part le constat empirique que depuis qu'elle a été proposée (c'est-à-dire depuis maintenant quatre-vingt ans) aucun contre-exemple sérieux n'a été proposé, d'autre part l'équivalence extensionnelle entre ses différentes formulations (FRG totales, fonctions Turing-calculables qui se terminent, λ -définissables, ...) pourtant intensionnellement très différentes. Il est cependant clair qu'aucun de ces deux arguments n'est dirimant, comme l'ont fait remarquer Georg Kreisel [Kreisel 1972, p. 321], Robin Gandy [Gandy 1988, p. 73] et Elliott Mendelson [Mendelson 1990, p. 228]. Pour en obtenir une justification plus pertinente, il est intéressant de revenir aux explications contenues dans, ou pouvant être déduites des, deux articles séminaux : [Church 1936a] et [Turing 1936-1937]. Dans son analyse « pas à pas », Church introduit, mais sans justification, l'affirmation cruciale selon laquelle les étapes prises en compte dans un calcul sont *récurives*. Il en déduit alors que les théorèmes de tout système formel contenant l'arithmétique sont FRG. Mais comme l'a fait remarquer Wilfried Sieg, faute d'une explication convaincante de la raison pour laquelle une telle hypothèse restrictive est introduite, la présentation effectuée est circulaire [Sieg 1994, p. 87] et [Sieg 2002, p. 394-395]. Au contraire, Turing explicite les limitations de l'agent humain calculant qu'il considère pour justifier les conditions restrictives qu'il impose à la « machine de Turing », ainsi que cela a été

particulièrement mis en évidence par les travaux récents de Wilfried Sieg³²⁸ [Sieg 2002, p. 395-397] et [Sieg 2006, p. 203]. Ces conditions restrictives³²⁹ sont essentiellement au nombre de deux :

- 1) la condition de bornage ; il y a une borne finie au nombre de configurations qu'un calculateur humain peut reconnaître immédiatement ;
- 2) la condition de localité ; un calculateur humain ne peut modifier que les sous-configurations immédiatement reconnaissables.

Comme l'a fait remarquer Wilfried Sieg, le point crucial est alors qu'une telle contrainte appliquée à une machine de Turing est équivalente aux contraintes d'effectivité habituelles appliquées à un système formel³³⁰ [Sieg 2006, p. 194-195]. Ces contraintes d'effectivité d'un système formel se traduisent en particulier par les caractères, récursif de l'ensemble des énoncés et récursivement énumérable de l'ensemble des théorèmes (ainsi que le fait que l'ensemble des FRG partielles est énumérable par programme : Tableau VI), ce qui introduit alors la justification manquante dans l'explication de Church³³¹. Il en résulte par ailleurs qu'il n'est pas envisageable d'étendre les possibilités d'une machine de Turing sans en modifier la finalité, en autorisant par exemple l'utilisation d'une quantité infinie d'information à chaque étape comme le suggère Robert Black [Black 2000, p. 256] ; nous allons y revenir. Par contre, l'utilisation d'une machine de Turing *universelle*³³², qui revient à introduire en entrée de la machine de Turing une liste d'instructions potentiellement infinie, ou plus exactement finie mais non bornée, est parfaitement acceptable sans remettre en cause les conditions restrictives introduites, contrairement à ce qu'indique Judson Webb [Webb 1990, p. 300].

Il est sans doute utile de préciser à cet endroit ce que recouvre exactement la notion de « machine de Turing ». Une « machine de Turing » n'est ni une machine numérique abstraite (au sens large) ni un être humain (au sens large) idéalisé, mais une machine numérique abstraite traduisant le comportement d'un être humain idéalisé lorsque celui-ci calcule de manière

³²⁸ Même si ceux-ci avaient été anticipés par Hao Wang [Wang 1974, p. 91-92].

³²⁹ Nous rappelons que ces restrictions sont par ailleurs associées à une *idéalisation* : la longueur du ruban est infinie.

³³⁰ Ce qui restreint *de facto* la classe des systèmes formels utilisables à ceux du 1^{er} ordre, dans la mesure où un système du deuxième ordre ne peut pas présenter les conditions d'effectivité requises. Cela rejoint une remarque de Saül Kripke [Kripke 2013, p. 82] et contraint alors la démarche générale proposée par Jaakko Hintikka [Hintikka 2000, p. 44].

³³¹ On peut également noter que bien que Post ait proposé une solution similaire à celle de Turing (dans [Post 1936]), celle-ci n'y est accompagnée d'aucune justification.

³³² Nous rappelons qu'une machine de Turing universelle est une machine de Turing capable de simuler le fonctionnement de toutes les machines de Turing possibles, pour un langage de programmation donné (Turing en propose une description particulièrement éclairante [Turing 1947 (2004), p. 383]).

routinière. Cette interprétation est indiscutablement présente dans l'article séminal de Turing, même si elle n'était peut-être pas suffisamment mise en évidence³³³ :

Un homme en train de calculer la valeur d'un nombre réel peut être comparé à une *machine* susceptible de se trouver dans un nombre fini d'états q_1, q_2, \dots, q_p , que nous appellerons ses *m-configurations*. ».
[Turing 1936-1937 (1995), p. 51]

Le comportement d'un homme en train de calculer est à tout instant déterminé par les symboles qu'il observe et par son « état mental » du moment. Nous supposons que ce calculateur [computer] observe un nombre maximal M de symboles (ou de cases) à la fois. [Turing 1936-1937 (1995), p. 78]

Cette interprétation avait cependant été quelque peu oubliée, vraisemblablement en relation avec l'avènement depuis, des ordinateurs, à quelques exceptions près comme Wittgenstein en 1942-1943 [Shanker 1987, p. 619]. Elle a été soulignée cinquante ans après par Robin Gandy [Gandy 1988, p. 80] et surtout par Wilfried Sieg [Sieg 1997, p. 171]. Ceux-ci ont même proposé d'utiliser un terme spécifique : « *computer* », pour désigner (aujourd'hui) un calculateur humain (en le différenciant ainsi d'un « *calculateur* » numérique) [Sieg 1994, p. 92, n. 62]. Il en résulte que la thèse de Church-Turing originale ne concerne pas les machines numériques (même idéalisées) en général³³⁴, contre Martin Davis et Andrew Hodges [Davis 2004, p. 200] et [Hodges 2006, p. 250]³³⁵, à rebours d'une présentation que l'on trouve couramment (y compris dans les écrits de Gödel [Gödel 193? (1995), p. 168] et [Gödel 1951 (1995), p. 304-305], sachant que cela intervient en particulier dans le contexte du débat « homme-machine », sur lequel nous reviendrons au § 6). Elle ne concerne pas d'avantage un être humain (même idéalisé) en général et en particulier pas un cerveau humain, ce que, selon Robin Gandy, Turing lui-même n'a jamais suggéré [Gandy 1988, p. 80]. La critique, par Gödel [Gödel 1972a (1990), p. 306] et par Georg Kreisel [Kreisel 1972, p. 319-320], du caractère trop simplifié du codage des états mentaux dans l'analyse de Turing est donc infondée (nous reviendrons à celle de Gödel au § 6.1). L'interprétation que nous venons de rappeler de la machine de Turing ainsi que la prise en

³³³ En remarquant qu'en 1936, le terme anglais « *computer* » utilisé par Turing, ne pouvait pas signifier autre chose que : *calculateur humain*.

³³⁴ La généralisation de la thèse de Church-Turing au cas d'une machine numérique (dont les limitations ne sont plus liées à des caractéristiques humaines mais à des considérations physiques), n'a été proposée que bien plus tard par Robin Gandy : [Gandy 1980] et a donné lieu depuis à une abondante littérature.

³³⁵ Il convient à ce sujet de rester prudent quant à l'interprétation des écrits de Turing lui-même durant sa participation au développement des machines numériques (donc bien après 1936). On peut effectivement y trouver mention d'un lien entre ces machines et les machines de Turing (au sens de [Turing 1936-1937]) : par exemple en 1947 [Turing 1947 (2004), p. 379, 383]. Mais ce sont alors les textes de conférences destinées à un public non spécialisé et l'on peut donc raisonnablement penser qu'il ne s'agit là, de la part de Turing, que d'une facilité à finalité pédagogique.

compte de l'exigence de constructivité dans la thèse de Church-Turing, s'opposent alors à l'affirmation selon laquelle la notion de calculabilité, qui intervient dans l'analyse de Turing et dans la thèse de Church-Turing, présente un caractère « absolu ». Une telle position a été défendue sous deux formes, que nous allons aborder successivement :

- 1) la version forte traduit directement l'idée que la récursivité (versus la calculabilité au sens de Turing) est « en soi » et non « pour les humains », avec l'exemple emblématique de la fonction d'entier définie par la vérité ou la fausseté de la conjecture de Goldbach ;
- 2) la version faible se limite à l'affirmation que le concept de calculabilité au sens de Turing est indépendant du choix d'un système formel particulier, contrairement par exemple à celui de dérivabilité.

La version forte, défendue par exemple par Elliott Mendelson et Bruno Marchal, provient manifestement, soit d'une confusion entre les FRG totales et les seules FRG totales *qui peuvent être établies constructivement* (respectivement, les niveaux 3 et 4 du schéma F), soit d'une interprétation de la thèse de Church-Turing faisant intervenir l'ensemble de toutes les FRG totales (et non l'ensemble des FRG totales *qui peuvent être établies constructivement*)³³⁶. Ce dernier point est clairement reconnu par Elliott Mendelson, celui-ci se réfugiant derrière la limitation de la récursivité à un cadre non constructiviste [Mendelson 1990, p. 226]. La réponse de Bruno Marchal est plus grossière puisqu'il se contente d'affirmer : « la règle pour la calculer [la fonction] existe, mais à l'heure actuelle, personne ne la connaît » [Marchal 1998, p. 71], mais elle revient évidemment à la même limitation à un cadre non constructiviste. Enfin, Gödel a défendu une position proche de cette version forte de l'objection lors de la conférence Gibbs de 1951. Il y oppose en effet les mathématiques « objectives » aux mathématiques « subjectives » : celles accessibles au moyen de démonstrations formelles. Il évoque d'ailleurs explicitement les conjectures (en 1951) de Fermat et de Goldbach [Gödel 1951 (1995), p. 313]. Il a cependant l'honnêteté de rattacher cette position à son engagement en faveur d'un réalisme au sens fort, particulièrement explicite dans le même texte [Gödel 1951 (1995), p. 313 et 323]. Cette position n'est donc pas d'avantage recevable dans le cadre de la neutralité ontologique (en mathématiques) associée à la thèse de Church-Turing.

Comme cela est souvent mentionné, la version faible est exposée dans [Gödel 1946 (1990)] :

³³⁶ Comme l'indiquait Kleene dès 1936, si les FRG totales peuvent être déterminées de manière purement arithmétique (donc en soi), ce n'est pas le cas des FRG totales établies de manière constructive [Kleene 1936, p. 738].

Il me semble que cette importance [du concept de récursivité générale ou de calculabilité au sens de Turing] est largement due au fait, qu'avec ce concept on a réussi pour la première fois à donner une définition *absolue* d'une notion épistémologique intéressante, c'est-à-dire, une notion qui ne dépende pas du formalisme choisi.

[Gödel, 1946 (1990), p. 150 ; souligné par nous ; le commentaire entre crochets est nôtre]

En fait, cette position avait déjà été explicitement énoncée par Gödel dès 1936 [Gödel 1936 (1986), p. 399], avec alors « absolu » entre guillemets. Comme l'a cependant mentionné Bernays en 1939 [Hilbert, Bernays 1939 (2001), p. 466-467] et comme cela a surtout été explicité par Wilfried Sieg ainsi que nous l'avons précisé, les systèmes formels concernés doivent répondre à des exigences fortes particulières qui sont techniquement équivalentes aux conditions d'effectivité habituelles³³⁷. Cette remarque relativise ainsi le caractère « absolu » mentionné par Gödel, comme Wilfried Sieg l'a souligné avec insistance [Sieg 1994, p. 88, n. 26], [Sieg 1997, p. 168], [Sieg 2002, p. 394] et [Sieg 2006, p. 194-195]. Dans le cadre de cette discussion, il convient enfin de mentionner le cas particulier d'un article récent de Michael Rescorla : [Rescorla 2007], qui défend quant à lui l'idée que la calculabilité au sens de Turing doit être relativisée à la sémantique utilisée, concernant en particulier les nombres entiers. On est bien entendu alors confronté à une démarche opposée à celle que nous avons discutée, puisqu'au lieu de chercher à « absolutiser » la récursivité il est ici question de la relativiser davantage ! Michael Rescorla dénonce le fait que l'analyse de Turing n'aille pas au-delà d'une analyse purement syntaxique (sur des chaînes de caractères) et n'introduise pas d'argument sémantique lui permettant de faire intervenir les nombres naturels et partant les FRG [Rescorla 2007, p. 258, n. 9]. Nous pouvons répondre qu'un tel lien sémantique existe (et existait en 1936) au travers du finitisme de Hilbert et de l'analyse détaillée que celui-ci a fournie pour en définir la base, ce que nous avons longuement discuté au § 1.1. Il est donc légitime de considérer qu'un tel lien est bien *implicitement* présent dans l'analyse de Turing (et vraisemblablement considéré comme évident par lui). Par ailleurs, les situations particulières pour lesquelles la prise en compte d'une sémantique conduit à des résultats différents de ceux naturellement attendus dans l'analyse que propose Michael Rescorla, vont au-delà de celles qui s'appliquent aux machines de Turing et à la récursivité.

Une autre question que l'on peut encore se poser sur la « thèse » de Church-Turing est le statut de celle-ci. Sur un plan historique, comme cela est souvent mentionné, Church a présenté

³³⁷ Comme le remarque Wilfried Sieg, l'introduction de ces critères d'effectivité contraint les systèmes formels et *restreint* donc la classe de ceux qui sont acceptés [Sieg 2006, p. 196]. L'universalité de l'adoption de ces critères pour les systèmes formels depuis 1936 a cependant pu conduire à oublier l'existence de cette restriction.

la thèse qui porte son nom comme une définition, ce qui a été vivement critiqué par Post qui considérait quant à lui qu'il s'agit au minimum d'une hypothèse de travail et éventuellement d'une loi :

L'objet du présent article est de proposer une *définition* de la calculabilité effective, qui est considérée correspondre de manière satisfaisante à la quelque peu vague notion intuitive, en des termes pour lesquels les problèmes de cette classe sont souvent établies, [...]. [Church 1936a, p. 346 ; souligné par nous]

Nous proposons cette conclusion actuellement comme une *hypothèse de travail*. Et à notre avis il en est de même de l'identification par Church de la calculabilité effective à la récursivité. [...] Le succès du programme ci-dessus changerait, pour nous, cette hypothèse non pas tant en une définition ou en un axiome, mais en une *loi naturelle*. [Post 1936, p. 105]

En fait, le travail déjà réalisé par Church et d'autres conduit cette identification considérablement au-delà du statut d'hypothèse de travail. Mais masquer cette identification sous une définition [comme le fait Church] cache le fait qu'une découverte fondamentale des limitations du pouvoir de mathématisation de l'Homo Sapiens a été faite et nous aveugle sur la nécessité de sa vérification continue.
[Post 1936, p. 105, n. 8 ; le commentaire entre crochets est nôtre]

[...] Nous estimons cependant qu'ultimement, « loi » [pour la thèse de Church-Turing] décrit mieux la situation [que « thèse »]. [Post 1944 (1965), p. 307, n. 4 ; les commentaires entre crochets sont nôtres]

Dans le cas de Post, cette position est à relier à sa conception plus large en philosophie des mathématiques, sur laquelle nous reviendrons au § 7.1. Quant à la position de Church, elle a été remplacée par celle qui est devenue consensuelle parmi les philosophes de la logique et des mathématiques, qui considèrent qu'il s'agit (justement) d'une « thèse ». Une position finalement peu différente a été récemment défendue par Maël Pégny, en reprenant la proposition qu'Elliott Mendelson associe à la vision standard de la thèse [Mendelson 1990, p. 229] suggérant le statut de « reconstruction rationnelle » :

La meilleure qualification possible du statut épistémique de la thèse nous paraît être celle de *reconstruction rationnelle*. Par ce terme, nous désignons une proposition théorique visant à substituer, dans un certain contexte d'usage scientifique, une expression rigoureusement définie à une expression intuitive, démunie de définition explicite, afin d'en préciser la signification et/ou l'extension.
[Pégny 2013, p. 32]

Une proposition plus audacieuse consiste à affirmer que la « thèse » de Church-Turing est en fait une démonstration mathématique informelle³³⁸, comme l'ont soutenu à la fois Robin Gandy et Elliott Mendelson³³⁹, le premier allant jusqu'à faire état à son sujet, de « théorème de Turing » [Gandy 1988, p. 76] et [Mendelson 1990, p. 230]. L'argumentation d'Elliott Mendelson est beaucoup plus développée que celle de Robin Gandy, mais les deux se ramènent à une démarche heuristique qui se borne à établir un parallèle entre la thèse de Church-Turing et, soit une définition (comme celle de la vérité par Tarski), soit une démonstration informelle (comme toutes celles qui prévalaient jusqu'à la fin du XIX^e siècle). Nous leur opposons deux objections. Dans le cas des définitions proposées il s'agit bien de « définitions », qui, même si elles peuvent présenter un caractère naturel, ne sont pas uniques, des définitions crédibles différentes pouvant être (et ayant été) proposées (par exemple pour la vérité, les théories de Kripke ou de Barwise et Etchemendy à la place de celle de Tarski). Quant à l'appel à la notion de « démonstration informelle », il s'apparente à un acte de foi, dans la mesure où la défense d'une telle idée ne s'est accompagnée à ce jour d'aucun argument convaincant en sa faveur emportant l'adhésion. En sens inverse et comme le souligne Janet Folina, un résultat mathématique peut tout à fait être jugé clair ou bien entraînant une évidence massive en sa faveur (comme la conjecture de Goldbach, qui a pu être empiriquement vérifiée jusqu'à 4×10^{18}), sans qu'il ne soit pour autant considéré comme prouvé [Folina 2006, p. 232-233].

Il reste à préciser la signification épistémologique de l'exploitation d'une machine de Turing. Comme nous l'avons vu, si la thèse de Church-Turing a été introduite dans le contexte de travaux conduisant à établir l'indécidabilité de certains systèmes formels particuliers, l'exploitation *sans restriction* d'une machine de Turing a une portée beaucoup plus large, puisqu'elle est équivalente à l'utilisation d'un système formel effectif (qu'il soit décidable ou indécidable), c'est-à-dire à la possibilité d'effectuer une dérivation au sens formel. Or d'une

³³⁸ Sachant que Nachum Dershowitz et Yuri Gurevich prétendent établir une démonstration *formelle* de la thèse de Church-Turing : [Dershowitz, Gurevich 2008], qui plus est avec une portée plus générale (incluant la thèse de Kleene et les machines de Gandy). Mais une telle allégation n'est pas justifiée. En effet, leur démonstration est établie sur la base de quatre postulats [Dershowitz, Gurevich 2008, p. 306], qui ne font finalement que reformuler, de manière plus abstraite, certains traits généraux de la thèse elle-même (ce qui recouvre l'appréciation par Wilfried Sieg du caractère « circulaire » de leur démonstration [Sieg 2013e, p. 119]).

³³⁹ Ainsi que très récemment, Saül Kripke, Stewart Shapiro et Robert Soare. Kripke en appelle à une « preuve intuitive rigoureuse » (dont il précise qu'elle n'est pas susceptible d'être formalisée dans ZF), tout en renvoyant pour cela à son ouvrage de 1996 non publié [Kripke 2013, p. 78, n. 5]. Stewart Shapiro se borne à renvoyer au constat historique que la plupart des démonstrations mathématiques publiées ne sont pas des dérivations formelles, sans relever qu'elles pourraient l'être [Shapiro 2013, p. 158]. Il établit également la comparaison avec la preuve analogue par Wilfried Sieg de l'argument de Turing, sans faire état de la différence entre un argument convaincant et une preuve [Shapiro 2015, p. 294]. Robert Soare s'oppose à la formulation de « thèse » au motif que ce terme renvoie à une réévaluation continue, ce qu'il refuse [Soare 2015, p. 300, 308] puis renvoie à [Kripke 2013].

manière générale, pour l'ensemble des mathématiques, une démonstration mathématique formelle est naturellement à rattacher au « contexte de la vérification » (qui s'oppose au « contexte de la découverte », en suivant la dichotomie néopositiviste classique). Pour rendre compte de la pratique mathématique, il est alors légitime d'associer à la démarche de vérification (en mathématiques) l'exigence épistémologique d'*intersubjectivité*, que nous avons introduite dans le cadre de la discussion sur l'optimisme rationaliste de Hilbert (critère (4) du § 5.2) et que nous exprimons sous la forme plus précise d'une demande d'intersubjectivité en principe à l'égard de tout être humain. Cette exigence entraîne en particulier, d'une part qu'une vérification en mathématique ne doit pas faire appel à l'innovation, l'intuition ou l'intelligence, quoi que recouvrent exactement ces notions, et d'autre part qu'une démonstration doit être publiquement communicable en un temps fini³⁴⁰. Il faut encore ajouter, cela sera important pour la suite de la discussion, que sur la base de la même exigence d'intersubjectivité, la vérification ne doit faire appel à aucun engagement invérifiable en principe³⁴¹, et en particulier à aucun engagement ontologique (typiquement en faveur d'une forme de réalisme) en mathématiques. Nous retrouvons alors la caractéristique majeure du programme de Hilbert (au sens large), de rechercher l'assentiment de tout mathématicien sur son ancrage finitiste, quelle que soit en particulier l'inclination philosophique de celui-ci. Ces caractéristiques traduisent ainsi les exigences épistémologiques générales du contexte de la vérification en mathématique, d'une manière beaucoup plus large que pour la seule situation de décidabilité, qui n'en représente qu'une situation particulière. Notre proposition sur la signification épistémologique de l'exploitation d'une machine de Turing consiste alors en deux points :

- 1) la machine Turing est une expression de l'exigence générale d'intersubjectivité d'une démonstration en mathématique dans le contexte de la vérification, lorsque celle-là est appliquée à un système formel³⁴² ;
- 2) l'exigence générale d'intersubjectivité d'une démonstration en mathématique, implique la nécessité que cette démonstration soit publiquement communicable (en un temps fini) et que sa vérification ne fasse appel à aucun engagement invérifiable en principe par n'importe quel être humain.

³⁴⁰ Ces conditions apparaissent également dans le contexte plus restreint qui caractérise la décidabilité, au travers des définitions informelles classiquement proposées des notions de calculabilité, d'algorithme, et d'effectivité.

³⁴¹ Il ne s'agit évidemment pas d'exclure de la pratique des mathématiques tout engagement invérifiable en principe, à commencer par celui du choix du système formel lui-même. L'exclusion concerne tout engagement invérifiable *supplémentaire*, non présent dans la théorie utilisée et ajouté (d'une manière *ad-hoc*) au cours de la vérification de toute preuve au sein de cette théorie.

³⁴² Il en résulte en particulier que l'exploitation d'une machine de Turing n'a la prétention de rendre compte que du « contexte de la vérification » et certainement pas du « contexte de la découverte », ainsi que le précise Solomon Feferman [Feferman 2006, p. 149].

On peut remarquer que bien que de très nombreux travaux aient été consacrés à l'analyse de la machine de Turing, en particulier récemment, ceux émanant de Robin Gandy, Jack Copeland, Oron Shagrir et surtout Wilfried Sieg, une telle proposition n'a pas été mentionnée explicitement par ces auteurs, mais certaines de leurs remarques en sont très proches [Gandy 1988, p. 76], [Sieg 1994, p. 101], [Copeland 2000, p. 23] et [Shagrir 2006, p. 410]. On peut également se demander pourquoi une telle proposition n'a pas été formulée par les protagonistes de l'époque³⁴³, puisqu'il faut attendre cinquante ans pour que Kleene fasse référence au caractère « publiquement communicable »³⁴⁴, comme caractéristique essentielle de la thèse de Church-Turing [Kleene 1987, p. 495]. Il nous faut encore nous assurer que cette proposition est bien cohérente avec les conditions d'établissement d'une machine de Turing et donc de manière équivalente, avec l'imposition des conditions d'effectivité habituelles à un système formel. Le point central est l'hypothèse que l'exigence épistémologique d'intersubjectivité est vérifiée par un être humain qui effectue un *calcul* sur des entiers naturels. Turing n'a sans doute pas fait état d'une telle hypothèse dans son article séminal, mais on peut penser qu'une telle idée était implicite (et évidente) pour lui dans le cas considéré de l'usage exclusif de papier et de crayon, ainsi que le souligne Maël Pégny [Pégny 2013, p. 151]. Rózsa Péter formulait dès 1957 une remarque très proche de notre affirmation en faisant état d'une procédure « qui doit être mécanique » [Péter 1957 (1967), p. 225]. Wilfried Sieg est quant à lui explicite sur le même point :

Je défends que le remplacement des théories axiomatiques en théories formelles fût motivé par des considérations épistémologiques. Un point central fut la requête que la vérification des preuves doive être effectuée d'une manière radicalement intersubjective. Il devait donc inclure seulement des opérations similaires à celles utilisées par un calculateur humain lorsqu'il conduit un calcul arithmétique.

[Sieg 1994, p. 101].

Nous pouvons en effet considérer, avec Rózsa Péter, que l'introduction d'une procédure finie et mécanique (également très proche dans l'esprit de la démarche qui a conduit Hilbert à

³⁴³ A deux exceptions près. Une allusion à ces deux éléments d'intersubjectivité est effectuée par Church dans le cadre de la description générale d'un système formel qu'il propose [Church 1956, p. 53-54 et 55]. On trouve une mention plus précise sur laquelle nous allons revenir dans [Péter 1957].

³⁴⁴ En renvoyant d'ailleurs pour cela au célèbre ouvrage de vulgarisation de Douglas Hofstadter sur le premier théorème de Gödel de 1931 : [Hofstadter 1979]. On peut cependant noter que dès 1967, Marwin Minsky faisait référence, à propos de la thèse de Church-Turing, à des processus « qui peuvent être capturés au moyen d'un nombre fini de mots » [cité dans Webb 1980, p. 220].

l'exigence de finitisme³⁴⁵) est le seul moyen qui ait été trouvé par les humains à ce jour, pour respecter la contrainte d'intersubjectivité³⁴⁶. Contre cette affirmation, Torkel Franzen a émis l'objection selon laquelle l'absence d'un accord entre les mathématiciens (par exemple intuitionnistes et classiques) sur ce qui constitue une preuve, ne permet pas de prétendre disposer d'une démarche universelle de preuve [Franzen 2005, p. 82]. La réponse est que le désaccord ne porte que sur les présupposés retenus mais non sur la démarche de vérification elle-même pour des présupposés donnés, qui pourrait être conduite par tout mathématicien, quelle que soit son inclination philosophique et quels que soient ces présupposés. Enfin, au-delà de la machine de Turing, notre proposition s'étend naturellement à la notion plus large de *calcul* (au sens de tout processus qui fournit un résultat numérique à partir de données d'entrée numériques) qui en dérive directement. En ce sens, pour mériter ce nom, nous considérons qu'un calcul³⁴⁷ doit raisonnablement être également nécessairement *publiquement communicable (en un temps fini), sans nécessité d'engagement invérifiable en principe*. Une telle exigence est fondamentale dans le cadre de la discussion qui va suivre, sur les propositions de dépassement de la thèse de Church-Turing³⁴⁸.

³⁴⁵ L'exigence d'intersubjectivité est ainsi le lien entre le programme de Hilbert au sens restreint et le programme de Hilbert au sens large, lien que n'a pas perçu Oron Shagrir qui oppose les deux approches [Shagrir 2006, p. 412].

³⁴⁶ Ce qui rejoint une remarque de Post [Post 1944 (1965), p. 310] et ce qui constitue une réponse à la question posée par Oron Shagrir [Shagrir 2006, p. 414].

³⁴⁷ Nous soulignons que sur le plan étymologique, le mot « raison » provient lui-même du latin « *ratio* », qui était utilisé pour traduire les opérations liées au *calcul*, à la comptabilité. Il recouvrait les actions de compter, de mettre en forme, d'organiser. On songe également à Leibniz et son *Calculus Ratiocinator*.

³⁴⁸ Sur un plan historique, il y a eu principalement trois objections, de nature différente, émises à la fin des années 1950 à l'encontre de la thèse de Church-Turing. Celles-ci émanent de Jean Porte, Rózsa Péter et Laszlo Kalmar et ont été critiquées par Elliott Mendelson dans [Mendelson 1963]. Ces objections, ainsi que leur critique, étant classiques (elles sont résumées dans [Moschovakis 1968]), nous ne les évoquons que brièvement. La critique de Jean Porte se ramène à une exigence d'effectivité humaine « en pratique », elle ne correspond donc pas à la conception standard de l'effectivité et de la décidabilité. La réserve (justifiée) de Rózsa Péter, concerne le point que nous avons déjà évoqué, de la nécessité d'une restriction de la thèse de Church-Turing aux FRG établies constructivement (Elliott Mendelson se contentant, dans sa réponse, du renvoi à une attitude classique en mathématiques). Le contre-exemple proposé par Laszlo Kalmar correspond quant à lui à une solution qui ne peut raisonnablement pas être qualifiée d'effective. On peut également mentionner, un peu plus tardivement, la critique de [Lee Bowie 1973], mais dont les deux arguments se ramènent en fait à ceux de Rózsa Péter et de Laszlo Kalmar (même si l'auteur s'en défend), dans le dernier cas en introduisant ce qui se ramène à un oracle [Lee Bowie 1973, p. 72]. On peut enfin mentionner la situation tout à fait particulière d'une contestation *de facto* de la thèse de Church-Turing, émise en 1982 par Tarski (donc tout à la fin de sa vie) dans le contexte technique spécifique des « théories équationnelles » relatée par Benjamin Wells dans [Wells 2002] (Tarski n'ayant rien publié sur le sujet). Selon Benjamin Wells, Tarski y affirme en effet que certaines familles de telles théories, dont on peut montrer qu'elles ne sont pas récursives, sont cependant « décidables » [Wells 2002, p. 303]. Mais la justification fournie à l'appui de cette affirmation reste très ténue [Wells 2002, p. 308], tout en ressemblant fortement à l'argument de [Kalmar 1959], ce qui appelle alors l'objection qui lui a été opposée par [Mendelson 1963].

Est-il alors possible d'envisager un dépassement de la thèse de Church-Turing³⁴⁹ ? Pour clarifier la discussion, rappelons que nous nous limitons au contexte dans lequel ont été introduites les machines de Turing, qui ne concerne que les situations « en principe », alors que la plus grande partie de la littérature sur le « dépassement de la thèse de Church-Turing » concerne en fait la prise en compte des contraintes « en pratique » (vitesse d'exécution et taille de la mémoire finies), qui concerne ce que Maël Pégny appelle « la thèse de Church-Turing étendue » [Pégny 2013, p. 18]. Cela répond en particulier à une remarque d'Oron Shagrir [Shagrir 2006, p. 414-415, n. 50]. Comme nous l'avons indiqué, l'exploitation d'une machine de Turing ne concerne pas le « raisonnement mathématique » en général, mais seulement le contexte de la vérification, à l'exclusion du contexte de la découverte. Cela lève déjà une ambiguïté qui apparaît en particulier dans les propos de Wittgenstein [Shanker 1987, p. 636], sur lesquels nous reviendront au § 7.1. Le critère certainement le plus éliminatoire est sans aucun doute la nécessité que la vérification d'une preuve, ou plus généralement l'exécution d'un calcul, soit *publiquement communicable* (en un temps fini). Une machine de Turing est potentiellement concernée par ce critère à trois niveaux, chacun d'entre eux ayant conduit à des affirmations de « dépassement » de la thèse de Church Turing. 1) Tout d'abord, les *entrées* de la machine de Turing ne peuvent être ni infinies, ni non définies a priori ni aléatoires. Cela répond directement à l'objection de Robert Soare [Soare 1996, p. 286] et à la question d'Oron

³⁴⁹ Il convient d'introduire ici une précision à l'égard d'un risque de confusion terminologique, bien mis en avant dans les différents écrits de Jack Copeland. Parmi les nombreux auteurs qui se sont exprimés sur les possibilités de dépassement de la « thèse de Church-Turing », la plupart sont des philosophes de l'informatique. Or pour ceux-ci, la préoccupation majeure et l'utilisation de la thèse de Church-Turing, sont radicalement différentes de celles des philosophes de la logique et des mathématiques : la consultation de [Goldin, Wegner 2008] par exemple, est particulièrement éclairante sur ce point. Ils cherchent en effet à rendre compte des processus informatiques susceptibles d'être appliqués par des ordinateurs (idéalisés) et non à assurer la vérification des démonstrations mathématiques. Ils s'intéressent donc en fait à une *extension* de la thèse de Church-Turing très différente de la thèse séminale : affirmant que tout ce que peut réaliser un ordinateur idéalisé peut être réalisé par une machine de Turing. Le risque de confusion est amplifié par le double sens du terme anglais : « computability », utilisé aussi bien pour le calcul que (plus généralement) pour les processus appliqués par des ordinateurs, pour lesquels Jack Copeland conteste l'utilisation du terme « calcul » [Copeland 2006, p. 151-153], contre [Hodges 2006]. [Bringsjord, Arkoudas 2006], dans lequel les auteurs entendent réfuter la thèse de Church-Turing sur la base d'un raisonnement cognitif portant sur « l'ensemble des histoires intéressantes », est emblématique d'une telle confusion. Mais le cas le plus flagrant est certainement la position défendue par Carol Cleland, en particulier dans [Cleland 1993], [Cleland 1995], [Cleland 2004] et [Cleland 2006]. Sur la base d'une notion élargie de « calcul », « d'effectivité » et de « procédure » ainsi que d'une notion restreinte de « mécanique », cet auteur entend en effet contester la conception standard (des logiciens et mathématiciens) de la thèse de Church-Turing, au motif que « L'aboutissement est qu'il n'y a pas de machine de Turing qui soit effective pour réussir une sauce hollandaise » [Cleland 1993, p. 301] (une position aussi extrême a bien entendu été critiquée, en particulier par [Horsten, Roelants 1995] et [Folina 2006]). Une position dans le même sens, mais beaucoup plus nuancée, a été défendue récemment par Solomon Feferman en faveur d'une extension de la notion d'algorithme au-delà de la logique (et de la métamathématique) : [Feferman 2015]. Il va de soi que les éventuelles solutions présentées en faveur d'un dépassement d'une telle version « informatique » de la thèse de Church-Turing, ne constituent en aucune manière un argument probant en faveur du dépassement de la version standard de la thèse de Church-Turing.

Shagrir [Shagrir 2006, p. 409-411]. Sont également éliminées sur la base de ce critère, les solutions à base de réseaux de neurones d'Hava Siegelmann analysées par Martin Davis, qui s'avèrent être non calculables parce que leurs paramètres d'entrée le sont [Davis 2004, p. 202-203]. Il en est de même des machines de Turing couplées à un environnement ouvert : avec Maël Pégny [Pégny 2013, p. 126], contre Jack Copeland et Richard Sylvan [Copeland, Sylvan 1999, p. 51]. 2) Ensuite le *fonctionnement* d'une machine de Turing ne peut faire intervenir ni l'infini en acte ni l'aléatoire. Cette restriction s'applique en premier lieu à la « machine de Turing » modifiée, étudiée par Turing lui-même sous la forme d'une « machine à oracle » (ou « machine-O ») dans [Turing 1939] (et reprise par Post dans [Post 1944]). En effet, même si la démarche procède étape par étape, l'explication fournie par Turing ne contient aucune information sur la procédure qui permettrait de calculer l'oracle³⁵⁰, l'objectif de Turing étant alors seulement exploratoire, en relation avec l'introduction de la notion de calculabilité *relative* (à l'oracle)³⁵¹ : avec Maël Pégny [Pégny 2013, p. 143], contre Robert Soare et Jack Copeland [Soare 1996, p. 286] et [Copeland 2000, p. 21]. Sont également éliminés les calculs interactifs à la Wegner et Golding discutés par Maël Pégny, qui font intervenir un algorithme potentiellement différent à chaque instance [Pégny 2013, p. 120]. Il en est a fortiori de même de l'appel à une éventuelle « calculabilité non-numérique » suggérée par Robert Soare et Jack Copeland [Soare 1996, p. 294] et [Copeland 2000, p. 32]. 3) Enfin, les *sorties* de la machine de Turing ne peuvent être infinies. Cela s'applique à l'objection du « calculateur prodige » introduite par Michael Rescorla [Rescorla 2007, p. 262] et par Maël Pégny [Pégny 2013, p. 67, n. 18]. Mais tout résultat (ou succession de résultats) proposé comme calcul ne sera reconnu comme tel par la communauté des mathématiciens que s'il est possible de s'y ramener (au moins en principe) au moyen d'une démarche « à la Turing 1936-1937 »³⁵², faute de quoi ce résultat ne sera considéré que comme une prophétie (ou un oracle). Il apparaît finalement ainsi qu'aucune des tentatives qui ont été proposées comme dépassement de la thèse de Church-Turing et qui sont en fait des tentatives de dépassement de la machine de Turing elle-même, ne parvient à répondre aux exigences d'intersubjectivité qui peuvent être associées à celle-ci.

Si nous faisons maintenant le bilan général de l'analyse effectuée sur le programme de Hilbert au sens large, nous aboutissons à un résultat beaucoup plus nuancé que celui qui est généralement proposé en réponse à la question : que nous disent les résultats « négatifs »

³⁵⁰ Nous avons déjà évoqué au § 3.3 le caractère profondément non effectif d'une telle démarche.

³⁵¹ Turing n'a d'ailleurs pas prétendu pouvoir dépasser ainsi les machines de Turing introduites en 1936.

³⁵² Ce qui rejoint la conclusion de Kleene (concernant en l'occurrence la dérivation dans un système formel), pour la situation similaire qu'il évoque, du « théoricien des nombres omniscient » [Kleene 1952 (2009), p. 303] et [Kleene 1967 (1971), p. 259, n. 179].

obtenus en logique et formelle et en informatique théorique, sur l'existence de problèmes mathématiques insolubles ? Pour pouvoir répondre à cette question à la suite d'une *démonstration*, comme le souhaitait Hilbert, il est en effet d'abord nécessaire de donner à celle-ci un contenu technique précis, incluant en particulier un critère *d'intersubjectivité* de la résolution des problèmes, ce qui conduit alors à deux interrogations de force différente : à minima sur la décidabilité de tout problème mathématique (optimisme rationaliste₁) et à maxima sur la complétude syntaxique d'un système formel, supposé unique, en mesure de rendre compte de l'ensemble des mathématiques, ou au moins de l'arithmétique du 1^{er} ordre (optimisme rationaliste₂). La réponse classique à l'optimisme rationaliste₂ est que l'incomplétude syntaxique de PA le réfute. Cette réponse ne prend cependant pas en compte le point technique que les indécidables arithmétiques qui sont cités à son appui sont en fait toujours des codages d'énoncés *métamathématiques* (et non des énoncés authentiquement arithmétiques), qui de ce fait ne peuvent prétendre être opposés à l'optimisme rationaliste₂. Pour aller plus loin dans le sens d'une démonstration de l'optimisme rationaliste₂ on peut d'abord envisager de faire appel à la thèse d'Isaacson, qui conduit à l'affirmation de la complétude syntaxique de PA pour les énoncés authentiquement arithmétiques. Cette stratégie échoue cependant, dans la mesure où l'appel à une thèse philosophique n'est pas compatible avec le critère de neutralité philosophique qui constitue l'une des conditions du critère d'intersubjectivité, qui est lui-même l'une des exigences du programme de Hilbert. Une autre solution possible, consiste au recours à des solutions particulièrement non effectives (en particulier de la part de Gödel), mais qui de ce fait impliquent également un engagement philosophique, et par là échouent à répondre aux exigences de Hilbert exactement pour la même raison. L'espoir de Hilbert d'une *démonstration* de l'optimisme rationaliste s'avère donc vain³⁵³. Quant à la réponse la plus connue (et tranchée) *contre* l'optimisme rationaliste₁ : celle de Grégory Chaitin, elle est clairement invalidée par l'analyse dévastatrice de Panu Raatikainen. Cette invalidation s'étend évidemment à la thèse forte sur les limites de la connaissance et de la raison que cet auteur (ainsi que ses thuriféraires) a cru pouvoir en déduire. Il ne s'avère donc possible ni de pouvoir démontrer ni de pouvoir réfuter l'optimisme rationaliste de Hilbert, qui reste donc à l'état d'une thèse métamathématique. Une solution radicale en vue du dépassement des résultats précédents consiste à envisager un élargissement de la thèse de Church-Turing. A la suite d'une analyse détaillée des conditions d'application de celle-ci, nous avons d'abord été amené à faire état de deux caractéristiques qui ont été soit oubliées soit ignorées jusqu'à une

³⁵³ Au-delà du constat que le théorème que le Tarski de 1940-1967, 1948, constitue bien une confirmation de l'optimisme rationaliste résultant d'une démonstration, certes pour un domaine très limité des mathématiques.

date récente. Tout d'abord la thèse fait intervenir les seules FRG (ou les seules fonctions Turing-calculables) qui peuvent être établies *constructivement*. Ensuite, comme cela a été souligné par Wilfried Sieg, une machine de Turing caractérise en fait *un être humain qui calcule*. Il en résulte deux conséquences philosophiques. 1) Il apparaît une certaine relativité, d'une part de la récursivité qui y intervient : « pour les humains » et non « en soi », d'autre part du concept de calculabilité, qui est bien indépendant de tout système formel particulier, mais avec la restriction forte que celui-ci respecte les conditions d'effectivité modernes habituelles. 2) Nous avons été amenés à proposer une signification épistémologique, à la fois précise et de portée générale de la machine de Turing (en retrouvant une proposition similaire de Sieg en 1994) : celle-ci exprime en fait l'exigence *d'intersubjectivité d'une démonstration formelle en mathématiques*, qui doit être publiquement communicable et dont la vérification ne doit faire appel à aucun engagement invérifiable en principe par n'importe quel être humain. Il en résulte une réfutation de tous les arguments qui ont été proposés en faveur d'un dépassement de la machine de Turing et donc a fortiori de la thèse de Church-Turing, sur la base du critère d'intersubjectivité introduit³⁵⁴. Cette analyse de la thèse de Church-Turing et de la machine de Turing nous amène assez naturellement au débat survenu à partir du début des années soixante et que nous allons aborder maintenant : sur la supériorité de l'être humain par rapport à la « machine », qui serait mise en évidence par les théorèmes de Gödel de 1931.

³⁵⁴ A posteriori, c'est le résultat opposé qui aurait été très surprenant. Nous avons en effet rappelé qu'il a un lien étroit entre la machine imaginée par Turing et les critères d'effectivité habituels appliqués à un système formel. Une éventuelle extension de la machine de Turing impliquerait donc un élargissement de ces critères d'effectivité (en cohérence avec l'idée de réduire les situations d'indécidabilité), sachant qu'aucune solution sérieuse en ce sens n'a été proposée depuis quatre-vingt ans. A la lumière de notre analyse, ce résultat n'est pas fortuit, mais provient fondamentalement du fait que les êtres humains n'ont pas trouvé d'autre solution que celle que nous venons d'évoquer au point 2, pour assurer l'exigence d'intersubjectivité d'une démonstration mathématique.

6 Les théorèmes de Gödel de 1931 permettent-ils de mettre en évidence une supériorité de l'esprit humain sur la « machine » ?

La question générale de savoir si la pensée à l'œuvre dans l'exercice des mathématiques doit être considérée comme « mécanique » (en un sens qui reste évidemment à préciser) a déjà été soulevée par Hilbert, qui y a répondu d'une manière positive mais avec une justification vague [Hilbert 1923 (1992), p. 140], [Hilbert 1928 (1992), p. 158] et [Hilbert 1931 (1992), p. 195]. Comme nous l'avons vu (§ 5.4), c'est seulement en 1936 que Turing a donné à cette question une réponse précise via l'introduction des machines de Turing, mais en se limitant alors aux opérations de calcul (de vérification). En 1961, la question est venue sur la scène philosophique au travers d'un article remarqué de John Lucas affirmant la supériorité de l'esprit humain sur la « machine », sur la base du premier théorème de Gödel de 1931 : [Lucas 1961]³⁵⁵. Ce point de vue a ensuite été défendu à son tour par Roger Penrose à partir de la fin des années 1980, puis, de manière beaucoup moins médiatisée, par Crispin Wright en 1995³⁵⁶. Il s'avère alors que pratiquement tous les grands noms de la philosophie de la logique et des mathématiques se sont exprimés sur ce sujet³⁵⁷, pour s'opposer aux deux premiers auteurs cités au cours d'un débat qui a lieu quasiment sans discontinuité depuis cinquante ans³⁵⁸. Enfin, comme nous l'avons déjà indiqué au § 3.5, dès la fin des années 1980, les tenants de la logique paraconsistante ont également cherché à justifier cette dernière avec une argumentation similaire à celle de John Lucas et de Roger Penrose. C'est cependant Gödel lui-même qui s'est engagé le premier dans un tel débat.

³⁵⁵ Sur le plan historique, une brève allusion résumant ce qui sera l'argument de John Lucas, a été proposée dès 1941 par Emil Post, mais n'a été publiée qu'en 1965 [Post 1941-1965, p. 417]. En 1950, Turing lui-même introduit de manière informelle un argument analogue [Turing 1950 (1995), p. 151-153]. Par ailleurs, en 1958 a été publié un argument également très proche de celui de Lucas, mais présenté de manière succincte [Nagel et Newman 1958 (1989), p. 93-94, 95]. Enfin, en 1960 un autre argument encore a été publié, aboutissant à la même conclusion mais plus grossier, basé sur le *second* théorème de Gödel de 1931 associé à « l'évidence » de la correction de l'arithmétique élémentaire (par rapport à la vérité naïve) [Myhill 1960, p. 466].

³⁵⁶ Pour être complet, il faut également mentionner Michael Dummett : [Dummett 1994] ainsi que Storrs McCall : [McCall 1999] et [McCall 2001], mais qui n'apportent pas d'élément nouveau par rapport à Lucas et Penrose. Il faut enfin signaler le cas particulier de [Detlefsen 2002], où Michael Detlefsen entend s'opposer au mécanisme sur la base, non du premier théorème de Gödel de 1931, mais du théorème de Löb de 1955. L'argument repose en fait sur une analyse fine mais particulièrement compliquée des contraintes qui peuvent être associées à la notion de croyance. Il est du coup une contestation très faible du mécanisme (traduit ici par le caractère récursivement énumérable de l'ensemble des croyances prises en compte).

³⁵⁷ Une liste de ces auteurs est donnée par Putnam [Putnam 2011, p. 332], liste à laquelle il faut encore ajouter Quine qui a également proposé un argument [Smart 1961, p. 109-110].

³⁵⁸ Avec l'exemple emblématique d'Hilary Putnam, qui s'est exprimé sur le même sujet, d'abord en 1960 (donc évidemment pas sur la base de [Lucas 1961] mais en fait sur celle de [Nagel et Newman 1958]), puis en 2011, c'est-à-dire cinquante et un ans plus tard !

6.1 La position de Gödel

Au-delà de son engagement en faveur de l'optimisme rationaliste que nous avons déjà évoqué (§ 5.2), Gödel est intervenu régulièrement dans les débats homme - machine et corps - esprit, de 1938 jusqu'à la fin de sa vie, avec comme point d'orgue la conférence « Gibbs » : [Gödel 1951]. Si sur une aussi longue période sa position a indéniablement évolué, elle est restée cohérente et s'est seulement radicalisée dans les derniers écrits : [Gödel 1964a] et [Gödel 1972a]. L'originalité et surtout la subtilité de cette position est toutefois rarement mentionnée dans la littérature, à l'exception notable de [Webb 1980, § 4.4] et surtout de [Webb 1990] ainsi également que [Sieg 2006]. Nous allons l'analyser de manière détaillée sous la forme de trois éclairages successifs : 1) l'interprétation par Gödel de la thèse de Church-Turing, 2) l'exploitation de ses propres théorèmes de 1931 dans le débat homme-machine, 3) son engagement dans le débat corps-esprit. Nous aborderons ensuite l'interprétation originale de Richard Tieszen : [Tieszen 2006], qui tente de rendre compte de la position de Gödel sur ces sujets par la prise en compte de la phénoménologie de Husserl.

Nous avons déjà évoqué l'implication de Gödel dans l'émergence de la thèse de Church-Turing, d'une part en en proposant lui-même une solution technique (l'utilisation des FRG), d'autre part en faisant valoir l'intérêt tout particulier de la machine de Turing parmi les différentes solutions techniques démontrées équivalentes (en particulier par rapport au λ -calcul et aux FRG). Mais si l'on analyse avec attention l'ensemble de ses écrits, l'on constate, comme l'a souligné Judson Webb [Webb 1990, p. 296], qu'à aucun moment il ne fait intervenir la notion d'*effectivité* (au sens naïf) ou d'*intuitivité* qui intervient pourtant de manière fondamentale dans l'expression standard de la thèse de Church-Turing (§ 5.4). Ce à quoi Gödel associe l'une des solutions techniques démontrées équivalentes (machines de Turing, λ -calcul, FRG, ...) est en effet, successivement et *exhaustivement* : « le raisonnement mathématique mécanisé » et « la calculabilité mécanique » [Gödel 193? (1995) p. 164 et 168], « dériver des formules du calcul des prédicats » [Gödel 1939 cité par Sieg 2006, p. 197], « le concept de procédure finie » et « le concept de fonction calculable d'entiers » [Gödel 1951 (1995), p. 304-305], « le formalisme aveugle » [Gödel 1962 (2003), p. 176], « le concept de système formel », « le concept de procédure mécanique, alias algorithme ou procédure calculatoire ou procédure combinatoire finie » [Gödel 1964a (1986), p. 369]. En d'autres termes, Gödel ne retient qu'une

forme très faible de la thèse de Church-Turing, où il associe la machine de Turing (et ses équivalents) à seulement : « calculabilité mécanique » (en suggérant qu'il pourrait y avoir une calculabilité non mécanique) et « procédure calculatoire et raisonnement, mécanisés » (en suggérant qu'il pourrait y avoir des procédures calculatoires et des raisonnements non mécanisés). Il ne fait en particulier jamais référence à l'effectivité (au sens naïf) ou l'intuitivité, ni plus largement à toute aptitude humaine générale, *même limitée à la vérification formelle d'une preuve*. Sur ce dernier point il existe par ailleurs manifestement une ambiguïté dans les écrits de Gödel lorsque celui-ci parle de « raisonnement », faute d'avoir introduit la distinction méthodologique entre (ce qui est classiquement reconnu comme) le contexte de la découverte et le contexte de la vérification. Si en 1964 on peut penser que seul le second est en jeu [Gödel 1964a (1986), p. 369-370], la remarque de 1967 avec la mention « d'appels renouvelés nécessaires à l'intuition mathématique » tend à inclure également le premier [Gödel 1967 (2003), p. 162]. La même absence de distinction est flagrante en 1972 [Gödel 1972b (1990), p. 306, remarque 3], où elle se traduit par un procès d'intention manifestement injustifié à l'égard de [Turing 1936-1937] qui l'a conduit à faire état de « l'erreur de Turing » dans le titre de cette remarque. Il oppose en effet des concepts qui concernent indiscutablement le contexte de la découverte, alors que l'analyse de Turing se limitait tout aussi indiscutablement à celui de la vérification.

Dans le débat homme - machine, dès 1938 Gödel défend une position tranchée mais mesurée :

Une autre manière d'exprimer ce résultat est celle-ci : il n'est pas possible de mécaniser le raisonnement mathématique, c'est-à-dire, il ne sera jamais possible de remplacer le mathématicien par une machine, même si vous vous confinez vous-mêmes aux problèmes de la théorie des nombres.

[Gödel 193? (1995), p. 164-165]

C'est cependant surtout en 1951, lors de la conférence « Gibbs » : [Gödel 1951], que Gödel s'est exprimé sur ce sujet même s'il y est revenu ultérieurement mais dans des documents non publiés : [Gödel 1961/?] ou privés : lettre à Léon Rappaport du 2 août 1962 [Gödel 1962 (2003), p. 176-177], lettre à David Plummer du 31 juillet 1967 [Gödel 1967(2003), p. 161-162] et commentaires rapportés dans [Wang 1974]. En 1951, en se fondant une nouvelle fois sur ses théorèmes de 1931 (en l'occurrence le second), il est conduit à l'alternative³⁵⁹ suivante : soit

³⁵⁹ En toute rigueur il ne s'agit pas d'une alternative, dans la mesure où Gödel envisage également la troisième possibilité que les deux options évoquées soient simultanément vérifiées.

l'esprit humain surpasse la puissance de toute machine de Turing, soit il existe des problèmes insolubles en théories des nombres. Bien entendu, sur la base de sa conviction en faveur de l'optimisme rationaliste, il est conduit à rejeter la seconde branche de l'alternative³⁶⁰ et comme il considère par ailleurs que le cerveau peut, quant à lui, être ramené à une machine de Turing³⁶¹ [Gödel 1951 (1995), p. 309, n. 13], il est amené à soutenir l'idée que l'esprit doit être distingué de son support biologique à rebours de « l'esprit du temps » [Gödel 1951 (1995), p. 311] et [Wang 1974, p. 326]. D'une manière globale, la lettre du 2 août 1962 à Léon Rappaport contient le résumé de sa position dans ce débat :

Mes théorèmes [de 1931] montrent seulement que la *mécanisation* des mathématiques, c'est-à-dire l'élimination de *l'esprit* et des entités *abstraites*, est impossible, si l'on veut avoir une fondation satisfaisante et un système des mathématiques. [...] Je n'ai pas prouvé qu'il y a des questions indécidables pour l'esprit humain, mais seulement qu'il n'y a pas de *machine* (ou de *formalisme aveugle*) qui peut décider toutes les questions de la théorie des nombres (même d'une sorte très spéciale).
[Gödel 1962 (2003), p. 176]

Il apparaît ainsi qu'en 1962 comme en 1938, Gödel considère que les théorèmes de 1931 (en fait le second) montrent l'impossibilité d'une mécanisation complète de la démarche mathématique (en particulier via l'impossibilité de dériver la consistance de PA dans PA). Pour autant, en aucun cas il ne soutient que ces théorèmes prouvent la supériorité de l'esprit sur la machine.

Comme nous l'avons indiqué, l'engagement de Gödel dans le débat corps – esprit apparaît dans ses écrits tardifs : [Gödel 1964a] et [Gödel 1972b]³⁶². Cet engagement est particulièrement

³⁶⁰ Georges Boolos (qui introduit le texte de 1951) remarque que l'existence de problèmes absolument indécidables n'est plus aujourd'hui considérée comme surprenante par la majorité des mathématiciens [Boolos 1995, p. 293-294]. Stewart Shapiro, dans un registre similaire, oppose la relativité de Quine au réalisme de Gödel dans ce débat [Shapiro 1998, p. 294]. Enfin Peter Koellner conclut à l'indécidabilité humaine de la dichotomie proposée par Gödel, à la suite d'une analyse technique détaillée basée sur une interprétation particulière des termes de cette dichotomie [Koellner 2016, p. 185].

³⁶¹ Au-delà du fait que cette position implique une interprétation de la notion de machine de Turing qui va bien au-delà de la notion classique (§ 5.4), on peut noter à ce sujet que Gödel donne l'argument d'un nombre nécessairement fini de neurones et de synapses [Gödel 1951 (1995), p. 311]. Si cela peut effectivement être associé au nombre fini des états d'une machine de Turing, il convient alors d'associer la longueur infinie du ruban de celle-ci à une utilisation du cerveau pendant une durée potentiellement infinie. Ce serait ainsi ce dernier point qui traduirait l'aspect idéalisé de l'être humain considéré. Cette interprétation est confirmée par le témoignage suivant rapporté par Hao Wang : « Par exemple, signifiait-il [Gödel] par "esprit", celui d'une seule personne ou bien les esprits de l'espèce humaine ? En réponse il dit qu'il signifiait un esprit individuel d'une durée de vie illimitée, [...] » [Wang 1993, p. 121].

³⁶² Sachant que les documents écrits par Gödel sur le même sujet, non publiés dans les *Collected Works* mais analysés par Pierre Cassou-Noguès, en introduisant des notions nouvelles non explicitées comme celle de « concept de concept », ouvrent sur plus de questions qu'ils n'apportent de réponses [Cassou-Noguès 2006, p. 74].

fort puisqu'il le conduit à suggérer l'existence de « procédures finies *non-mécaniques* » [Gödel 1964a (1986), p. 370] et « d'une procédure bien définie qui pourrait effectivement être conduite (et qui produirait une fonction non récursive en théorie des nombres) » [Gödel 1972b (1990), p. 306]. Une telle position est clairement en opposition avec la formulation standard de la thèse de Church-Turing³⁶³ que nous avons évoquée (§ 5.4). A fortiori, elle s'oppose frontalement à l'interprétation épistémologique que nous avons proposée de cette thèse à la suite de Wilfried Sieg : même au sein du seul contexte de la vérification elle entend autoriser des procédures de vérification qui iraient au-delà des possibilités d'une machine de Turing³⁶⁴. Toutefois, le recours à des extensions du finitisme comme celle de [Gödel 1958], au-delà du fait que celles-ci ne peuvent raisonnablement pas être considérées comme finitistes (§ 3.1) en étant associées à un engagement philosophique, achoppent sur le critère d'intersubjectivité lié à une telle interprétation. Le commentaire de Gödel lui-même sur les conditions d'acceptabilité d'une telle position « nécessitant une avancée substantielle dans notre compréhension des concepts de base des mathématiques » [Gödel 1972b (1990), p. 306, remarque 3], traduit de manière éloquente³⁶⁵ les limitations de la portée de cette position³⁶⁶. Il convient alors de remarquer qu'en parallèle à ce débat l'auteur avait été amené quelques années plus tôt à défendre une autre position forte, de nature épistémologique : l'existence de mathématiques « objectives » (s'opposant aux mathématiques « subjectives » accessibles par des démonstrations) : [Gödel 1951]. En cohérence avec la position ultérieure que nous venons d'évoquer, cela le conduit à supposer qu'il existe des moyens d'accéder à la vérité d'une proposition sans passer par une preuve ; comme élément de justification il évoqua l'utilisation d'une démarche inductive, qu'il propose d'appliquer aux énoncés de Goldbach et de Fermat [Gödel 1951 (1995), p. 313]. Il a alors l'honnêteté de rattacher cette position à son engagement par ailleurs en faveur d'un réalisme au sens fort [Gödel 1951 (1995), p. 313], engagement bien connu explicite dans le même texte³⁶⁷

³⁶³ Solomon Feferman et Judson Webb considèrent qu'il s'agit d'un revirement de la part de Gödel afin d'éviter une conclusion à la Post en faveur d'une limitation de la raison humaine [Feferman 1990, p. 281] et [Webb 1990, p. 293]. Judson Webb évoque également la suggestion de Solomon Feferman selon laquelle Gödel n'accepterait pas la première partie (en fait, l'explication véritable : se reporter au § 5.4) de l'explication de Turing (qui fait référence aux calculateurs humains) [Webb 1990, p. 297]. Il suggère que Gödel a peut-être oublié (en 1972) qu'en 1936 « calculateur » ne signifiait pas autre chose que « personne effectuant des calculs » [Webb 1990, p. 302].

³⁶⁴ En évoquant un finitisme non-mécanique Gödel a sans doute à l'esprit la démarche de [Gentzen 1936] ou la sienne de [1958].

³⁶⁵ C'est à leur propos que Kleene évoque des « promesses en l'air » auxquelles il ne croit évidemment pas [Kleene 1987, p. 494].

³⁶⁶ On peut également mentionner l'incompréhension mentionnée par Wilfried Sieg des arguments que Gödel avance pour justifier le caractère dynamique de l'esprit (en l'opposant à une machine de Turing) [Sieg 2009, p. 618].

³⁶⁷ Nonobstant le fait que Pierre Cassou-Noguès ait proposé d'interpréter le réalisme de Gödel, en particulier dans le texte de 1951, en un sens plus faible dans [Cassou-Noguès 2005].

[Gödel 1951 (1995), p. 320, 323]. Il est évident que le même engagement réaliste de Gödel peut être également associé à sa position dans le débat corps - esprit, qui se présente effectivement comme l'une de ses conséquences naturelles. Il en résulte alors un affaiblissement considérable de cette position ainsi subordonnée à un tel engagement philosophique, sachant que Gödel ne fournit par ailleurs aucun autre argument probant en sa faveur.

Comme nous l'avons indiqué, Richard Tieszen défend quant à lui l'idée qu'au-delà de l'engagement réaliste de Gödel en mathématiques, c'est la prise en compte de la *signification* des termes abstraits utilisés au travers de l'intentionnalité (au sens Husserlien), qui permet « d'accéder » aux suites infinies et de « voir » par exemple que l'énoncé G est vrai [Tieszen 2006, p. 233]. Comme on le sait, Gödel a effectivement évoqué l'intérêt de prendre en compte la phénoménologie de Husserl [Gödel 1961/? (1995), p. 385], mais il n'a laissé aucun écrit connu en ce sens. Au-delà de la difficulté de justifier une telle exégèse en l'absence d'écrits de Gödel, on peut émettre plusieurs objections à l'encontre de Richard Tieszen. Celui-ci affirme que pour qu'une croyance en faveur d'un énoncé métamathématique soit justifiée, il faut une évidence en sa faveur [Tieszen 2006, p. 239-240] ; mais d'où peut provenir une telle évidence et comment être assuré qu'il ne s'agit pas d'une illusion ? Il affirme ensuite que les machines (contrairement aux humains) n'ont pas d'intentionnalité ni d'intuition [Tieszen 2006, p. 246]. Le problème est que dans le cadre du débat en cours, la pertinence de cette spécificité humaine est justement le point principal contesté par les tenants du mécanisme. Pour espérer convaincre ceux-ci il faudrait donc disposer d'un argument en sa faveur *externe à la phénoménologie*, qui permette de distinguer l'être humain d'une machine (une sorte de « test de Husserl » par analogie avec le « test de Turing »), ce que l'auteur ne fournit évidemment pas. « L'estocade » que propose alors Richard Tieszen est que l'affirmation que l'esprit humain est une machine présuppose elle-même *in fine* l'intentionnalité [Tieszen 2006, p. 249-250]. Un tel argument serait effectivement probant si l'on disposait là encore d'une justification en sa faveur externe à la phénoménologie (par exemple s'il était montré qu'une telle affirmation est la condition de possibilité de toute communication entre êtres humains). Mais comme l'auteur se contente de présupposer l'intentionnalité (chez les humains) l'argument est évidemment encore circulaire. Il apparaît ainsi qu'en remplaçant comme mode explicatif de la position philosophique de Gödel un engagement réaliste (en mathématiques) par un engagement en faveur de la phénoménologie de Husserl, on ne fait de nouveau que tomber de Charybde en Scylla par rapport à une exigence d'intersubjectivité associée à l'absence de tout engagement philosophique.

6.2 L'argument de Lucas

L'argument de John Lucas a été originellement présenté dans [Lucas 1961] puis explicité dans [Lucas 1968a] et [Lucas 1999] (soit presque quarante ans après la publication initiale !). Comme nous l'avons indiqué il a suscité un très grand nombre de critiques, pourtant l'auteur affirme que ses différents contradicteurs n'ont pas réellement répondu à son argument³⁶⁸ et considère que celui-ci reste valable sur le fond [Lucas 1999, p. 106]. Cette apparente contradiction provient sans doute du fait que l'argument repose en fait, non pas sur les théorèmes de Gödel de 1931, mais sur une *démarche informelle inspirée de ceux-ci* associée à un argument « dialectique » appliqué à l'encontre de l'interlocuteur mécaniste fictif de son article séminal auquel il entend s'opposer [Lucas 1968a, p. 145] et [Lucas 1999, p. 104, 113]. L'argument de Lucas tel que nous pouvons le reconstituer est le suivant : pour tout système formel auquel s'applique le premier théorème de Gödel 1931, il existe au moins un énoncé non dérivable dont nous pouvons nous rendre compte qu'il est vrai (au sens naïf) sur la base de son sens. A l'inverse, tout modèle mécanique proposé en vue de simuler le fonctionnement de l'esprit humain ne peut prétendre à une telle aptitude. Donc l'esprit humain possède au moins une aptitude que tout modèle mécanique qui prétend le simuler ne possède pas. En évoquant son argument, l'auteur fait explicitement référence à l'attitude similaire de Gödel dans ce débat mais sans signaler la nuance forte que Gödel ne prétend nullement *déduire* cette position de l'utilisation de ses théorèmes de 1931 [Lucas 1999, p. 111]. La justification apportée par l'auteur en faveur de cette spécificité de l'esprit humain est vague : c'est parce que nous sommes « en dehors » du système formel³⁶⁹ que nous pouvons « voir » que l'énoncé est vrai [Lucas 1961 (1983), p. 81-82] ; cette aptitude serait à relier à la capacité d'avoir une conscience de soi, considérée comme une caractéristique de l'esprit humain [Lucas 1961 (1983), p. 96] ; elle serait également une caractéristique même de la démarche rationnelle [Lucas 1968a, p. 149]. Deux critiques à caractère général ont été émises contre l'argument de Lucas : 1) Si l'argument est celui du « recul » (que l'esprit humain est censé avoir par rapport à une affirmation comme celle de la vérité de l'énoncé de Gödel), nous n'avons pas besoin de faire état des théorèmes de Gödel de 1931 pour le justifier [Benacerraf 1967, p. 20] et [Webb 1968, p. 170], 2) Dans ce

³⁶⁸ Ce qui n'est pas tout à fait exact : Charles Chihara a essayé de répondre directement à l'argument original sous sa présentation informelle dans [Chihara 1972], sans proposer pour autant une réponse très convaincante, de notre point de vue.

³⁶⁹ Torkel Franzen a dénoncé le caractère ambigu d'une telle formule, souvent reprise dans ce débat [Franzen 2004, p. 180-181].

type de débat l'anti-mécaniste est placé devant l'alternative dirimante, soit de surestimer les possibilités d'un être humain, soit au contraire de sous-estimer celles d'une machine de Turing [Benacerraf 1967, p. 21-22], [Webb 1980, p. 235] et [Dubucs 1992, p. 74]. Des critiques plus précises et détaillées ont également été émises, que nous allons analyser successivement³⁷⁰.

L'objection la plus simple qui a été proposée contre l'argument de John Lucas, repose sur l'évidence de la *finitude* de tout esprit humain et de toute machine concrète, ce qui soulève le problème posé par l'idéalisation retenue à la fois pour l'esprit humain et pour la machine. Il convient alors de ne pas oublier deux éléments importants : 1) Si l'on borne le nombre des itérations possibles (traduit par la longueur du ruban sur la machine de Turing), ni les théorèmes de Gödel de 1931 ni celui de Turing de 1936-1937 ne s'appliquent (à cause de l'impossibilité d'appliquer la diagonalisation), 2) En sens inverse, en utilisant une machine de Turing *universelle* il est possible de rendre compte à la fois de l'ensemble des machines de Turing possibles. John Lucas défend une position classique (analogue à celle de Gödel), selon laquelle la confrontation porte sur un esprit humain *idéalisé* au travers de la prise en compte d'une durée de vie infinie, opposé à une machine idéalisée sous la forme d'une machine de Turing (dont le ruban est de longueur infinie) [Lucas 1961 (1983), p. 84]. Il évoque par ailleurs explicitement le fait qu'une telle idéalisation est nécessaire à l'application du premier théorème de Gödel de 1931 [Lucas 1999, p. 103] (même s'il semble l'avoir oublié quelques pages plus loin [Lucas 1999, p. 105]). Les éléments de justification qu'il présente en faveur de cette idéalisation sont par contre techniquement incorrects ou indigents : en confondant finitisme et finitude [Lucas 1999, p. 105], en faisant valoir que le premier théorème de Gödel de 1931 s'applique non seulement à PA mais également à Q (ce qui n'est pas contestable mais ne répond pas à la question de la finitude) [Lucas 1999, p. 107], ou en suggérant la possibilité d'un regroupement de plusieurs esprits humains (plusieurs mathématiciens) en contradiction avec l'idéalisation affichée (concernant *un* esprit humain) [Lucas 1999, p. 107]. Cette discussion l'amène à introduire une nouvelle aptitude qu'il considère encore inhérente à l'esprit humain (ou à la rationalité) : non seulement l'appréhension de la potentialité infinie [Lucas 1999, p. 109] mais également celle de l'infini en acte [Lucas 1999, p. 110], au travers de ce qui n'est rien d'autre qu'une expression informelle de la règle- ω (au sens classique) [Lucas 1999, p. 114].

³⁷⁰ Signalons au passage un argument technique global original proposé par Torkel Franzen contre l'argument de Lucas, faisant appel au théorème de Löb de 1955 [Franzen 2005, p. 179-180]. L'argument est le suivant : l'énoncé H_S de Henkin, qui code dans le système formel S l'énoncé dont l'interprétation dans la métalangue est « cet énoncé est dérivable dans S », est à la fois « vrai » (au sens intuitif, de par son sens) et néanmoins dérivable dans S (par le théorème de Löb de 1955). John Lucas a donc la charge d'expliquer en quoi la spécificité de G par rapport à H fait que son argument s'applique avec le premier mais non avec le second.

Indépendamment de la faiblesse de la justification apportée par John Lucas, une telle idéalisation de l'esprit humain n'a pas été remise en cause par les critiques (par exemple Paul Benacerraf et Stewart Shapiro [Benacerraf 1967, p. 18-19] et [Shapiro 1998, p. 275-276]), à l'exception notable de Judson Webbs dans [Webb 1980] qui fait appel, quant à lui, à une machine de Turing *universelle*. Judson Webb fait valoir qu'une comparaison homme-machine doit raisonnablement être effectuée avec une machine de Turing universelle (et non avec une machine de Turing), en renvoyant à l'article séminal de Turing [Webb 1980, p. 228]. L'argument qu'il fournit à l'appui de ce choix est cependant bref et superficiel : « Tout comme un humain peut réussir cela ou ce calcul ou tâche lorsque lui sont données les instructions ou l'éducation convenable, la machine de Turing universelle M_u le peut également » [Webb 1980, p. 228]. Il néglige en effet l'élément important que l'introduction d'une machine de Turing universelle ajoute un élément d'idéalisation *supplémentaire* à la machine de Turing (dont on sait déjà que le ruban est illimité) : l'infinité potentielle des calculs qui peuvent être effectués par l'ensemble (infini) des machines de Turing envisageables est virtuellement contenu dans la machine de Turing universelle. Le problème est alors que la comparaison avec un esprit humain idéalisé (sous la forme d'une durée de vie supposée infinie) n'est plus licite : il faudrait de plus prendre en compte l'ensemble des êtres humains possibles (l'ensemble des formations et des éducations qu'ils pourraient recevoir). Il est donc clair qu'avec ce choix Judson Webb *modifie les règles du jeu* introduites par John Lucas, ce qui réduit évidemment considérablement la portée de son argumentation qui repose essentiellement sur un tel choix (nous y reviendrons). Il est d'ailleurs à noter que dès 1968 (en réplique à [Benacerraf 1967]), donc avant la parution de l'ouvrage de Judson Webb, John Lucas ne contestait pas que son argument ne s'applique plus dans cette autre situation [Lucas 1968a, p. 146]. Une objection plus fondamentale à l'argument de Lucas porte sur l'assimilation par celui-ci (ainsi que par nombre de ses critiques) d'une « machine » à une machine de Turing. En effet, indépendamment du fait qu'une machine de Turing soit une machine idéalisée (le point précédent), ce choix pose deux problèmes : 1) Il néglige le fait qu'une « machine de Turing » n'est pas une machine, mais un être humain (idéalisé) qui calcule (§ 5.4), 2) Dans le débat mécanisme - anti-mécanisme, une « machine » est classiquement associée à une exigence de capacité de prédiction en principe, voire à un déterminisme, qui fait manifestement défaut à une machine de Turing (via l'indécidabilité du problème de l'arrêt). La seconde objection a été relevée par Paul Benacerraf [Benacerraf 1967, p. 13] mais surtout par Judson Webb [Webb 1980, p. 232, 245]. Il est manifeste que John Lucas est passé complètement à côté de ce point en 1961 [Lucas 1961, p. 96] et il n'y est pas revenu en 1968 ni en 1999. Ce que n'a pas vu John Lucas et qui a été souligné avec force par Judson

Webb, est que pour un dispositif *non fini* comme une machine de Turing, il ne suffit pas que le système opère selon des lois déterministes définies pour qu'il soit complètement prédictible. Il en résulte l'affaiblissement de l'argument de Lucas qui porte en fait sur une version non classique du mécanisme. Enfin, celui-ci, mais également ses contradicteurs, est muet sur le premier problème évoqué (une machine de Turing n'est pas une « machine » mais l'idéalisation d'un être humain calculant). Ce problème entraîne une autre limitation de son argument, puisqu'il faut ajouter l'hypothèse supplémentaire que les « machines » considérées présentent les mêmes limitations que les machines de Turing³⁷¹.

L'objection la plus répandue contre l'argument de John Lucas est cependant d'une autre nature. Elle porte sur l'absence d'une justification suffisante par celui-ci de la consistance (ou de la correction) du système formel utilisé, à la fois par « l'esprit humain » et par la machine de Turing introduite, qui est évidemment indispensable à l'utilisation du théorème de Gödel de 1931 (c'est l'objection majeure de [Putnam 1960]³⁷²). La question de la consistance a bien été abordée par John Lucas, qui est tout à fait conscient du fait qu'elle est cruciale, dès son article séminal [Lucas 1961 (1983), p. 90]. Pour l'esprit humain, il considère que la consistance du système formel utilisé (a minima celle de Q) est évidente [Lucas 1961 (1983), p. 93], étant caractéristique de l'être humain considéré à la première personne [Lucas 1968a, p. 156]. Il tente ultérieurement de proposer des résultats techniques en sa faveur, mais d'une manière partiellement incorrecte et absolument non pertinente (eu égard à toute la discussion que nous avons déjà eue sur ce sujet), qu'il s'agisse des démonstrations finitistes de la consistance, du calcul des propositions, du calcul des prédicats du 1^{er} ordre, ou même de PA via le théorème de Gentzen de 1936 [Lucas 1999, p. 118]. Il introduit même une confusion en précisant que le terme « consistant » ne doit pas être pris dans le même sens lorsqu'on l'applique à l'esprit humain et à un système formel [Lucas 1999, p. 120]³⁷³. Bien entendu de nombreuses critiques se sont élevées contre l'indigence d'une telle justification. Tout d'abord en faisant valoir qu'elle ne rendait pas justice au fait qu'une inconsistance peut fort bien ne pas être triviale et donc ne pas être d'emblée perçue par le concepteur d'un système formel (a fortiori par « l'esprit humain ») [Chihara 1972, p. 509-511] et [Boolos 1990, p. 656]), au-delà de l'ambiguïté de l'application d'une telle notion à l'esprit humain [Webb 1980, p. 193]. Ensuite en faisant état

³⁷¹ Ou encore admettre la thèse de Gandy qui introduit un lien entre les deux (§ 5.4).

³⁷² On la retrouve ultérieurement régulièrement [Chihara 1972, p. 509-511], [Webb 1980, p. 193], [Lee Bowie 1982, p. 282], [Krajewski 1991, p. 44-45], [Shapiro 1998, p. 282-283] et [Murawski 1999, p. 330].

³⁷³ Ce qui a été relevé par Judson Webb [Webb 1980, p. 193] et Torkel Franzen [Franzen 2004, p. 181], mais qui semble hors sujet dans le cadre de la présente discussion où c'est bien la consistance (simple) au sens classique qui est en jeu, et non une vague notion de « consistance » qui pourrait être appliquée au discours ordinaire.

d'une alternative dirimante³⁷⁴ : 1) S'il est entendu se limiter à des justifications précises de cette consistance, celles-ci seront selon toute vraisemblance récursives auquel cas l'esprit humain est tout aussi bloqué que les machines de Turing elles-mêmes (l'ensemble des machines de Turing consistantes n'étant pas récursif), 2) s'il est au contraire envisagé de produire des justifications non récursives (ce qui revient peu ou prou à s'opposer à la thèse de Church-Turing), cela correspond à un engagement philosophique (anti-mécaniste) dans un sens que l'argument global est justement censé justifier, ce qui revient à vicier celui-ci. Pour ce qui concerne la consistance du modèle mécaniste (de la machine de Turing) destiné à rendre compte de l'esprit humain auquel il entend s'opposer, John Lucas présente son argument comme « dialectique » (selon ses propres termes) : cette consistance est considérée comme une exigence indispensable, et comme elle ne peut être établie par le modèle lui-même il estime que c'est le défenseur (humain) du modèle qui doit s'engager à son propos (selon la tactique argumentative de renversement de la charge de la preuve) [Lucas 1999, p. 117-118]. Cette position a évidemment été contestée en faisant d'abord valoir que le mécaniste est fondé à fournir des arguments convaincants en faveur de cette consistance, sans pour autant s'engager à prouver celle-ci [Shapiro 1998, p. 283-284]. En soulignant ensuite qu'une simple affirmation par le mécaniste de la consistance du modèle qu'il entend proposer, ou une justification empirique ou métaphysique de celle-ci, ne saurait intervenir dans une démonstration mathématique (même informelle). Il est par ailleurs à signaler que John Lucas évoque comme seule voie de sortie possible pour le défenseur d'une position « mécaniste », la solution que l'esprit humain, ainsi que le modèle mécaniste qui prétend en rendre compte, s'avèrent tous les deux « ultimement » inconsistants (en faisant explicitement référence à Hegel) [Lucas 1968a, p. 157] et [Lucas 1999, p. 122]. Une telle solution conduit de fait à l'argument de la logique paraconsistante, que nous aborderons au § 6.4.

Le cœur de l'argumentaire de John Lucas repose sur l'affirmation de la « vérité » de l'énoncé G du premier théorème de 1931³⁷⁵. L'auteur est tout à fait conscient à la fois de l'importance et de la difficulté de la justification correspondante³⁷⁶, en particulier de celle de sa transmission à d'autres personnes : « Je pense que je sais ce qu'est la vérité, mais je sais que je ne peux pas dire à quelqu'un d'autre exactement ce que c'est » [Lucas 1968a, p. 148]. Il affirme néanmoins

³⁷⁴ Au-delà de l'argument technique plus faible (voir le § 2.3) introduit par Judson Webb, selon lequel certaines formules codant la consistance sont dérivables dans un système formel [Webb 1980, p. 193].

³⁷⁵ Bien entendu, sur un plan technique, les deux affirmations (celle de la consistance et de celle de la vérité) sont liées, la seconde, définie formellement par rapport au modèle standard associé au système formel, ne faisant que traduire la correction de celui-ci et donc son ω -consistance (voir la figure C du § 1.5).

³⁷⁶ Il fait explicitement référence au théorème de Tarski de 1933-1936.

que tout être humain y a accès [Lucas 1968, p. 148], qu'il s'agit d'une disposition du caractère rationnel [Lucas 1968a, p. 149] qui lui permet de « voir » que G est vraie, ce qui échappe à une machine [Lucas 1968a, p. 155]. Cette référence à la vérité « naïve » et à son universalité, que l'on trouve déjà chez John Myhill en 1960 [Myhill 1960, p. 463], a naturellement été contestée [Webb 1968, p. 162], [Dubucs 1992, p. 777], [Wright 1994, § 3 - 4 -5], [Shapiro 1998, p. 284] et [Franzen 2004, p. 236]. Ces auteurs montrent que l'on retombe sur l'alternative qui a été mise en évidence dans la discussion sur la consistance (ce ne sont que deux facettes d'une problématique similaire) : 1) si par « voir comme étant vrai » il est entendu « prouver », alors l'esprit humain ne peut certainement pas faire une telle chose, 2) si cela signifie quelque chose de plus faible comme « avoir d'excellentes raisons de penser », alors cela ne répond pas à l'exigence de certitude mathématique que l'on est en droit d'attendre dans le cadre d'une exploitation d'un théorème mathématique comme celui de Gödel de 1931. Cela montre que pour espérer vaincre le mécanisme, il faut nécessairement faire appel à une étape *créative* et non se limiter à l'application d'un algorithme, ce qui ouvre alors sur deux nouvelles questions : 1) quelle peut être (et comment justifier) cette étape créative ? 2) En quoi les théorèmes de Gödel de 1931 montreraient que l'esprit humain la possède ? La réponse de John Lucas est un appel à une capacité supposée de l'ordre de l'intuition. Mais comme le souligne Stewart Shapiro [Shapiro 1998, p. 291-293], cette position ne fait que refléter un engagement épistémique³⁷⁷. Il est enfin reproché à John Lucas de confondre *ce qui est vrai* (sous-entendu, dans le modèle standard) et *ce qui peut être connu* (donc défendable) *comme étant vrai* [Webb 1980, p. 220], [Franzen 2005, p. 116-119] et surtout [Burgess 2010, p. 133], particulièrement éclairant. Il en résulte deux conséquences : 1) L'être humain, pas plus que la machine, ne peut décider (et donc savoir) quels énoncés sont vrais (dans le modèle standard) [Webb 1980, p. 220], 2) Il est possible que quelque machine M puisse de manière factuelle coïncider avec l'humainement prouvable, sans qu'il soit prouvable qu'il en est ainsi³⁷⁸ [Burgess 2010, p. 133]. Dans l'article plus tardif de 1999, John Lucas évoque enfin le point technique selon lequel son interprétation naïve de la vérité correspond en fait au choix du modèle standard de l'arithmétique [Lucas 1999, p. 107]. Il n'en tire toutefois pas les conséquences qui s'imposent sur la non intersubjectivité

³⁷⁷ Stewart Shapiro analyse de manière détaillée la question de savoir si les résultats techniques (non effectifs) de complétude de [Turing 1939] et de [Feferman 1962] pourraient être utilisés à l'appui de l'argument anti-mécaniste, sans cet engagement épistémique [Shapiro 1998, p. 285-291]. Il conclut sans ambiguïté par une réponse négative (nous y reviendrons au § 6.3).

³⁷⁸ En retrouvant alors la conclusion bien connue de [Benacerraf 1967] selon laquelle, si je suis une machine de Turing, je ne peux pas savoir laquelle. C'est également la conclusion à laquelle parvient Peter Koellner à la suite d'une analyse détaillée récente, sur la base de l'utilisation d'un opérateur arithmétique au sein de la logique modale S4, pour exprimer la prouvabilité absolue [Koellner 2016, p. 164].

que ce choix introduit (nous allons y revenir), en se contentant de faire référence au caractère commun des affirmations contingentes des mathématiciens sur ce plan³⁷⁹ [Lucas 1999, p. 107].

Au-delà de l'ensemble des remarques qui viennent d'être formulées, un certain nombre des commentateurs de John Lucas ont été conduit à proposer au préalable une interprétation *technique* de son argument informel³⁸⁰, avant d'en discuter le bien-fondé. L'ambiguïté de l'argument original, que nous avons signalée, a alors naturellement conduit à des interprétations différentes³⁸¹. Nous avons fait une allusion aux plus anciennes : [Putnam 1960] et [Benacerraf, 1967], nous n'y reviendrons pas. Nous allons par contre aborder celles qui sont certainement les plus détaillées : [Webb 1980], [Dubucs 1992] et [Shapiro 1998].

Nous avons déjà évoqué l'analyse de Judson Webb : [Webb 1980], que nous allons maintenant compléter. Nous rappelons que sa caractéristique essentielle est de faire intervenir une machine de Turing *universelle*. L'auteur établit alors qu'une machine de Turing universelle peut faire aussi bien qu'un être humain³⁸², pourvu que celui-ci présente des informations « suffisamment précises » en faveur de son argument [Webb 1980, p. 231-232]. Il pousse même la provocation en suggérant qu'une machine de Turing universelle présente des points communs avec ce que l'on associe habituellement à l'intelligence humaine [Webb 1980, p. 235]. D'une manière plus nuancée il avance l'idée d'un *double* comportement humain, d'une part qui peut être décrit de manière précise et détaillée et dont il peut être rendu compte « mécaniquement », d'autre part dont il ne peut pas l'être³⁸³ [Webb 1980, p. 235]. Nous avons cependant déjà montré la limitation forte de l'argument de Judson Webb, qui porte sur deux points : 1) L'appel à une machine de Turing universelle conduit à sortir de la « règle du jeu » proposée par John Lucas, en introduisant une dissymétrie dans les idéalizations respectives de l'être humain et de la machine, 2) L'argument de Lucas revient justement à tenter de justifier (avec plus ou moins de succès), l'existence d'un comportement humain à la fois précis et non mécanisable (en anticipant ce qui sera la suggestion analogue de Gödel en 1964 et 1972 : § 6.1).

³⁷⁹ Sur ce point, on peut signaler l'argument faible de Torkel Franzen, de la relativité de l'attitude des mathématiciens quant à ce qui est considéré comme prouvé (en prenant pour exemple la théorie des ensembles), ce qui revient implicitement à contester l'idée d'une vérité mathématique universelle [Franzen 2005, p. 82].

³⁸⁰ Judson Webb évoque à son sujet : « la dramatisation pittoresque de problèmes particuliers dans les fondations des mathématiques » [Webb 1968, p. 170].

³⁸¹ Ce constat a permis à John Lucas une défense facile en faisant remarquer, d'une part que les objections de ses contradicteurs étaient diverses, et d'autre part qu'elles ne portaient pas sur l'argument original.

³⁸² On peut signaler qu'une telle affirmation (pour une machine de Turing *universelle*) a été soulignée par Turing en 1950 [Turing 1950(1995), p. 153].

³⁸³ Sans toutefois proposer une ligne de démarcation associée dans l'activité du mathématicien, et en particulier sans introduire la distinction classique entre les contextes de la découverte et de la vérification.

Jacques Dubucs a proposé une critique particulièrement éclairante de [Lucas 1961], en introduisant un argument original³⁸⁴ et convaincant³⁸⁵ : [Dubucs 1992]. Il propose en fait successivement deux traductions techniques possibles de l'argument de John Lucas. La première consiste à faire intervenir la règle- ω au sens sémantique (voir le § 3.2) appliquée à l'énoncé G : dans un système formel répondant aux conditions requises par le premier théorème de Gödel de 1931, on peut dériver chaque instance de l'énoncé (qui est un énoncé Π_1) mais non l'énoncé G lui-même, et cela en relation avec l'existence de modèles non standards du système formel³⁸⁶. La raison pour laquelle l'être humain peut alors « savoir » que G est vrai, est qu'il fait le choix de définir la vérité dans le modèle standard sur la base de la *signification* qu'il accorde à l'énoncé G, tandis que le système formel (ou la machine de Turing qui le représente) est *intentionnellement inerte* [Dubucs 1992, p. 77-78]. Une telle explication est extrêmement proche de celle que propose John Lucas lui-même [Lucas 1968a, p. 155] et [Lucas 1999, p. 114], mais le mérite de Jacques Dubucs est de faire ressortir un élément crucial dans le débat en cours, à savoir que la reconnaissance humaine de la vérité de G n'autorise aucune manifestation *externe* (donc publique) de sa propre validité³⁸⁷ [Dubucs 1992, p. 82-83]. L'argument est le suivant : à la suite d'une durée finie, tout ce que l'être humain pourra justifier est la vérité successive de n instances de G. Dans le cas de l'être humain idéalisé dont il est question il pourra donc tout au plus (à la limite) justifier la vérité successive de l'infini dénombrable de l'ensemble des valeurs de G, mais pour justifier la vérité de l'énoncé G lui-même il faudrait pouvoir aller au-delà (en faisant intervenir le nombre transfini : ω), donc au-delà du dénombrable et partant au-delà du communicable. La seconde traduction technique possible de l'argument de John Lucas par Jacques Dubucs est la transcription en théorie des modèles de la condition de correction appliquée à la vérité de l'énoncé G, à savoir sa correction dans le modèle standard. La conviction de la correction du système formel par rapport au modèle standard retenu implique alors de s'assurer que, d'une part les axiomes du système formel sont satisfaits dans le modèle, et d'autre part que les règles d'inférence préservent la vérité. Mais comme le fait là encore ressortir Jacques Dubucs [Dubucs 1992, p. 84-85], la première partie de cette

³⁸⁴ L'argument est partiellement anticipé par Michael Dummett sans se référer explicitement à [Lucas 1961] [Dummett 1963 (1978), p. 186].

³⁸⁵ Cette analyse ayant été publiée uniquement en français, elle a malheureusement été totalement ignorée des nombreux échanges qui ont eu lieu sur ce sujet, presque tous en anglais.

³⁸⁶ Comme nous l'avons déjà signalé, il se trouve que cette interprétation a été explicitement introduite par John Lucas [Lucas 1999, p. 114], après la publication de [Dubucs 1992] et très vraisemblablement sans avoir pris connaissance de celle-ci.

³⁸⁷ Il est intéressant de noter que John Lucas a lui-même suggéré une telle remarque mais sans en tirer les conclusions qui s'imposent [Lucas, 1999, p. 107, 114].

conviction n'est pas susceptible d'une communication publique quelle que soit « l'intuition » qui lui est associée³⁸⁸. Jacques Dubucs suggère alors deux possibilités de dépassement d'une telle limitation de l'argument de John Lucas. Tout d'abord l'objection pourrait être contournée en faisant appel à une logique d'ordre supérieur ; le problème est que le système formel qui en résulterait ne serait alors plus effectif, ce qui enlève tout bénéfice à une telle opération. Ensuite, même si nous ne pouvons fournir la garantie publique de nos choix privés concernant le modèle retenu, nous avons de bonnes raisons (empiriques) de supposer que nous pouvons nous entendre sur eux. S'il existe une telle convergence, c'est qu'il y a bien une attitude humaine qui échappe à la possibilité de mécanisation et dont le mécaniste devra rendre compte ; il s'agit donc ici d'une remarque *en faveur* de l'argument de John Lucas, mais qui en est une forme très faible.

Stewart Shapiro a présenté une synthèse globale du débat intégrant les différents arguments qui ont été présentés (à l'exception de celui de Jacques Dubucs) avec le mérite d'une grande qualité pédagogique ([Shapiro 1998]), via une circonscription que nous avons traduite sur la figure G. Il distingue trois ensembles d'énoncés arithmétiques partiellement emboîtés³⁸⁹ : 1) L'ensemble A des théorèmes formels de PA, 2) L'ensemble hypothétique K, centre du débat, des énoncés arithmétiques humainement connaissables comme vrais dans le modèle standard de PA avec « certitude mathématique »³⁹⁰, 3) L'ensemble T des énoncés vrais dans le modèle standard de PA (en faisant l'hypothèse que PA est ω -consistant). Par rapport à cette circonscription le schéma général de l'argumentation est très simple [Shapiro 1998, p. 277-278] : par construction A est récursivement énumérable (ce qui est équivalent à l'exploitation de la machine de Turing) tandis que T n'est pas récursivement énumérable (par le théorème de Tarski de 1933-1936). Selon la thèse mécaniste : $K \equiv A$, ce qui entraîne qu'il existe des énoncés vrais dans le modèle standard de l'arithmétique³⁹¹ qui sont intrinsèquement non humainement

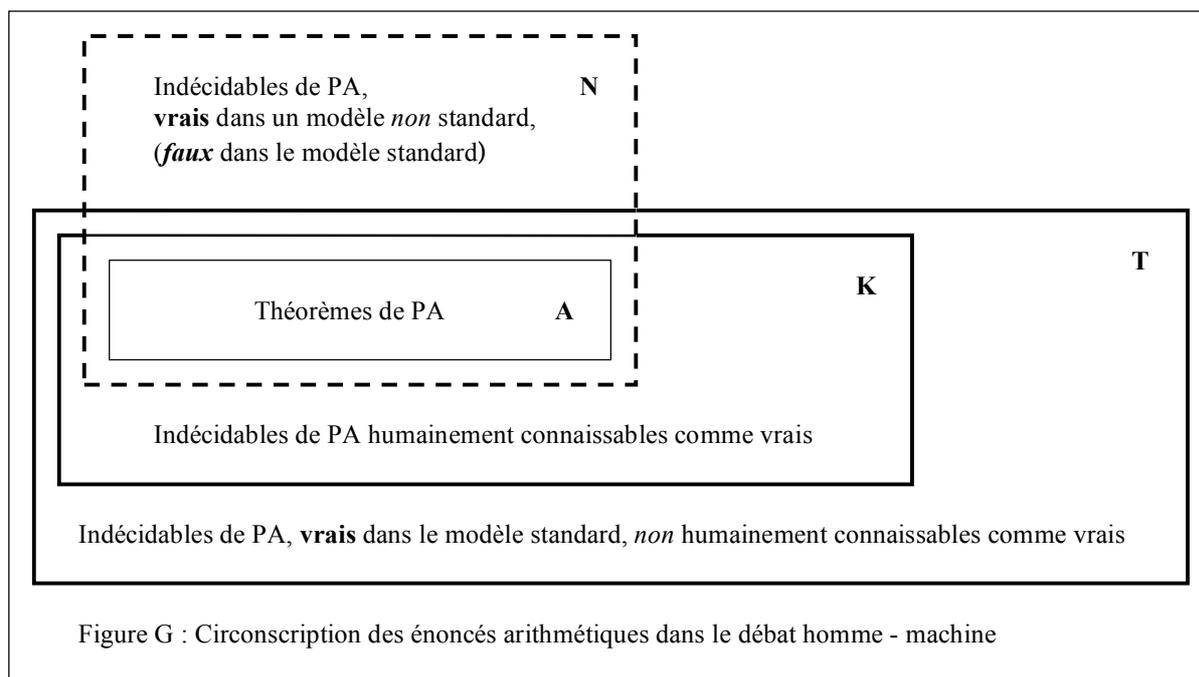
³⁸⁸ Dit autrement, il n'est pas possible de discriminer un être humain qui ferait pour sa part le choix d'un modèle non standard de l'arithmétique (qui reconnaîtrait ainsi la correction des axiomes du système formel sans pour autant asserter la vérité de G), des autres humains qui font un choix plus classique (en faveur du modèle standard) sur la base de ce qui pourra être publiquement communiqué par cet humain. L'argument est ici de nouveau que toute communication ne pourra porter, même à la limite, que sur la totalité (dénombrable) des entiers naturels. Il ne sera donc pas possible de savoir s'il faut tenir compte en plus, ou non, des entiers non-standard.

³⁸⁹ Sur cette figure, pour être complet et pour rendre compte de l'argument de [Dubucs 1992], nous avons ajouté un quatrième ensemble non proposé par Stewart Shapiro : l'ensemble N des énoncés vrais dans un modèle *non* standard de PA.

³⁹⁰ En reprenant la proposition de [Gödel 1951].

³⁹¹ On peut également envisager de prendre en compte des systèmes formels plus forts que PA (en se limitant alors aux seuls énoncés arithmétiques de ces systèmes formels), mais cela ne change pas fondamentalement la situation : pour tout système formel il y aura toujours des énoncés vrais dans le modèle standard et humainement inconnaisables.

connaissables comme tels³⁹². Selon la thèse anti-mécaniste : $K \equiv T$ ³⁹³, ce qui entraîne qu'il existe des procédés qui permettent de s'assurer qu'un énoncé est vrai dans le modèle standard de l'arithmétique et qui sont intrinsèquement non calculables. On retrouve ainsi, formulée de manière particulièrement claire, l'alternative de [Gödel 1951]. Il est tout de même nécessaire d'apporter une précision à une telle restitution : il est crucial d'ajouter que l'argument se rapporte au modèle standard de PA (ce que ne précise pas Stewart Shapiro). Il en résulte une contrainte supplémentaire pour l'anti-mécaniste : les procédés qu'il évoque lui permettent-ils de savoir si l'énoncé G appartient à T plutôt qu'à N (l'argument de Dubucs) ? Stewart Shapiro envisage également la solution de repli qui consiste à réduire la force du critère qui définit K, en limitant l'exigence à la requête « d'avoir un retour adéquat » (au lieu d'une connaissance avec certitude mathématique). Il montre alors facilement que ne pouvant définir un nouvel ensemble de manière suffisamment précise sur une telle base, il est impossible de lui appliquer les théorèmes de Gödel 1931 [Shapiro 1998, p. 295]. Ce résultat le conduit à sa propre conclusion générale : « Il n'y a pas de thèse mécaniste plausible proposée qui soit suffisamment précise pour être sapée par les théorèmes d'incomplétude » [Shapiro 1998, p. 275].



³⁹² La question de la pertinence *arithmétique* des énoncés indécidables de PA éventuellement laissés de côté (la thèse d'Isaacson : § 4.4) reste bien entendu posée, sachant qu'elle n'est pas prise en compte dans un tel débat.

³⁹³ A minima, K doit être plus grand que A et même que l'ensemble des théorèmes arithmétiques de tout système formel effectif.

Il apparaît finalement que la caractéristique essentielle de l'argument de Lucas est son caractère informel, celui-ci ne faisant que « s'inspirer » des théorèmes de Gödel de 1931. C'est sans doute ce qui explique sa persistance puisqu'il est toujours discuté cinquante ans après son émission ! Mais comme l'a souligné Stewart Shapiro [Shapiro 1998 p. 276-277], dans une telle controverse si l'on veut éviter un débat stérile il est indispensable de définir au préalable de manière précise la thèse que l'on veut défendre ou attaquer. Cela passe au moins par la clarification de trois points : 1) Sur quoi (sur quel contenu de la pratique mathématique) porte le débat : le contexte de la vérification uniquement ou bien également le contexte de la découverte ? 2) Quel est le niveau d'idéalisation retenu, d'une part pour les humains et d'autre part pour les machines ? 3) Quelle est la thèse (et donc le contenu du mécanisme) que l'on veut discuter ? Le point (1) a rarement été abordé par les protagonistes³⁹⁴. On peut cependant considérer que ceux-ci se sont implicitement accordés sur le fait que l'opération en jeu se limite à celle de la vérification, c'est-à-dire de l'établissement d'une preuve concernant les théorèmes (au sens informel) de l'arithmétique. L'étendue des idéalizations respectives adoptées pour l'être humain et pour la machine, qui correspond au point (2), est cruciale dans le débat. Le choix retenu par John Lucas, à savoir un être humain de durée de vie infinie d'une part et une machine de Turing d'autre part, paraît équilibré et raisonnable. En sens inverse, la solution de Judson Webb de faire appel à une machine de Turing *universelle*, outre le fait qu'elle change les règles du jeu, introduit une *dissymétrie* de l'idéalisation (en faveur de la machine), qui rend non pertinente la conclusion forte qu'il en déduit contre John Lucas. Pour définir le contenu de la thèse mécaniste en réponse au point (3), la solution la plus raisonnable est certainement celle que propose Stewart Shapiro en faisant intervenir l'ensemble des énoncés arithmétiques indécidables dans PA mais humainement connaissables comme vrais « avec certitude mathématique » (en reprenant en cela la proposition de [Gödel 1951]). Dans ces conditions, la clef de la défense de la position anti-mécaniste passe par la justification, non seulement de la consistance de PA, mais de plus de la correction de celui-ci dans le modèle standard. Comme nous l'avons vu, la justification par John Lucas de la consistance de PA est tout à fait indigente et de ce fait non recevable. Plus fondamentalement, l'appel à l'évidence de la vérité intuitive de l'énoncé G du premier théorème de Gödel de 1931 pour l'esprit humain le conduit en fait à un raisonnement circulaire, puisque c'est sur cette base qu'il défend finalement la supériorité

³⁹⁴ On peut toutefois signaler deux exceptions : Stewart Shapiro et Per Lindström [Shapiro 1994, p. 31] et [Lindström 2001, p. 242]. Le second se traduit par une mention implicite du point 1, mais l'article porte en fait sur les arguments proposés par Roger Penrose que nous allons analyser au § 6.3. Le premier se traduit par une mention explicite du point 1, tout en ne portant pas sur une critique des arguments de Lucas et de Penrose, mais sur une réflexion plus générale sur les liens entre calculabilité et raisonnement.

de l'esprit humain sur la machine. La conclusion à laquelle il parvient n'est donc que la conséquence de ses prémisses³⁹⁵, le premier théorème de Gödel de 1931 n'étant ici que le support (ou le travestissement : voir le titre de [Good 1969]) de ce transfert. Comme le montre [Shapiro 1998], une réduction de la force de la thèse mécaniste attaquée ne permet pas à John Lucas de sauver la mise. Cela justifie globalement la volée de bois vert dont l'argument a été l'objet de la part de l'ensemble de tous ses critiques. Si l'on envisage d'exprimer d'une manière plus technique l'argument de Lucas il apparaît pour cela deux solutions raisonnables, mais comme l'a établi Jacques Dubucs dans [Dubucs 1992] on est conduit dans les deux cas à la même limitation fondamentale : la caractéristique spécifiquement humaine dont la machine de Turing ne peut pas rendre compte n'est pas publiquement communicable. Eu égard à l'analyse que nous avons consacrée à la thèse de Church-Turing un tel résultat ne doit pas nous étonner, puisque nous avons montré que la machine de Turing est justement l'expression dans le contexte de la vérification formelle en mathématique, de l'exigence du caractère publiquement communicable garant de l'intersubjectivité recherchée (§ 5.4). A l'encontre de l'argument de Lucas selon de telles interprétations, son adversaire mécaniste aura alors beau jeu de dénoncer l'impossibilité d'en justifier l'intersubjectivité. Il lui restera cependant à expliquer l'origine de l'accord empirique que l'on peut néanmoins constater entre mathématiciens sur un choix (du modèle retenu pour un système formel donné) qui relève de la première personne, ce qui conduit bien à un argument anti-mécaniste, mais de forme très faible.

6.3 Les deux arguments de Penrose

Roger Penrose a historiquement présenté deux arguments successifs en faveur de la même conclusion anti-mécaniste que John Lucas, d'abord en 1989 : [Penrose 1989], puis en 1994 : [Penrose 1994]. Et comme John Lucas³⁹⁶, plus de vingt ans après son entrée en scène dans ce débat (dans [Penrose 2011]) Roger Penrose est resté sur ses positions, malgré les nombreuses objections qui lui ont été opposées. L'argument de 1989 est très proche de celui de John Lucas ; il est exprimé de manière tout aussi informelle (et vague) et il repose comme celui-ci sur l'évidence de la vérité (au sens naïf) du même énoncé G. La différence essentielle, soulignée

³⁹⁵ Ce point est également relevé par Paul Benacerraf [Benacerraf 1967, p. 21, 22], Judson Webb [Webb 1980, p. 235] et Stanislas Krajewski [Krajewski 1991, p. 44].

³⁹⁶ Jusqu'à considérer, comme celui-ci, que son argument est très proche de celui de [Gödel 1951] [Penrose 1993, p. 619], ce qui lui est explicitement contesté par Hao Wang [Wang 1993, p. 117].

par Roger Penrose, est le fait qu'il ne considère plus un être humain (idéalisé) unique mais la communauté de l'ensemble des mathématiciens [Penrose 1990, p. 696]. Cette position traduit donc le recours à une règle du jeu différente de celle de John Lucas. Avec le même souci d'un équilibre de l'idéalisation retenue, d'une part pour les humains, d'autre part pour les machines, c'est donc ici légitimement une machine de Turing *universelle* qu'il faut prendre en compte. Bien entendu, la critique de [Webb 1980] s'applique alors pleinement, même si elle doit être confrontée à la structure de l'argumentation de Roger Penrose qui diffère sensiblement de celle de John Lucas³⁹⁷. La justification introduite par Roger Penrose se ramène à un appel à l'évidence [Penrose 1989 (1992), p. 454-455] associée à un engagement réaliste au sens fort, accompagné ici d'une dénégation hautaine à l'égard de ceux qui ne partageraient pas une telle conviction (certes dans un ouvrage à destination d'un large public) :

[...] Quant au formaliste très dogmatique, impénétrable aux arguments de Gödel et qui prétend ne pas même admettre *l'existence* de quelque chose comme une vérité mathématique [au sens absolu], je l'ignorerai purement et simplement, puisqu'il ne possède apparemment pas la qualité de divination de la vérité sur laquelle porte toute la discussion !

[Penrose 1989 (1992), p. 509, n. 2 du § 10 ; le commentaire est nôtre]

Sans surprise, les très nombreuses critiques qui se sont élevées contre une telle position³⁹⁸ reprennent les trois principaux arguments dirigés contre [Lucas 1961]. Tout d'abord, le premier théorème de Gödel de 1931 ne nous dit pas que l'énoncé G est vrai, mais seulement que si le système formel pris en compte est (simplement) consistant (et même ω -consistant), alors l'énoncé G qui lui est associé est vrai (dans le modèle standard) sachant que nous n'avons a priori aucune certitude sur cette consistance (l'argument de [Putnam 1960]). Cela a été souligné à de multiples reprises [Boolos 1990, p. 655-656], [Davis 1990, p. 660], [Franzen 2005, p. 118-119], [Raatikainen 2005, p. 523] et [Berto 2009b, p. 180]. Ensuite, même si nous sommes une machine de Turing, nous ne pouvons pas savoir laquelle (l'argument de [Benacerraf 1967]) [Gaifman 2000, p. 469], [Franzen 2005, p. 118-119] et [Burgess 2010, p. 133]. Enfin, le théorème de Gödel ne nous fournit aucune indication quant à l'issue du débat sur le mécanisme [Davis 1993, p. 612]. Par contre, l'objection clef de [Dubucs 1992] sur le caractère non

³⁹⁷ Curieusement, aucun auteur à notre connaissance n'a soulevé ce point, et en particulier pas Judson Webb qui ne semble pas être revenu sur ce débat depuis la survenue de Roger Penrose.

³⁹⁸ Ces critiques ont été beaucoup plus nombreuses que celles formulées contre John Lucas (le seul n° 4, vol. 13, 1990, de la revue *Behavioral and Brain sciences* en compte pas moins d'une trentaine). Elles sont cependant de qualités très inégales et certains auteurs n'ont manifestement pas compris la teneur de l'argument, à l'instar de John Searle [Searle 1997 (1999), p. 63-102].

publiquement communicable des prémisses en jeu (via le choix du modèle standard) n'y apparaît jamais.

En 1994, sans doute conscient des faiblesses de sa position de 1989 (et de celle de [Lucas 1961]), Roger Penrose a affiné son argumentation au travers de ce qui a ensuite été appelé « le nouvel argument de Penrose ». Plus subtil et surtout un peu plus technique bien qu'encore informel, cet argument a donné lieu à nettement moins de critiques que le premier. La difficulté, renforcée par rapport à [Lucas 1961] et [Penrose 1989], est de devoir d'abord fournir une restitution plus claire que l'argument originel : [Penrose 1994] et plus détaillée que la version ramassée qu'en en ensuite proposé Roger Penrose lui-même dans [Penrose 1996]. Pour corser cette difficulté, les quelques critiques qui se sont attelés à la tâche en proposant un contre-argument ont également émis des critiques sur les contre-arguments des autres critiques. Si l'on tente de résumer une situation quelque peu embrouillée, il y a eu essentiellement trois critiques détaillées du second argument de Penrose³⁹⁹ : 1) La première est celle de [Chalmers 1995], contestée ensuite par [Lindström 2001], 2) La deuxième émane de [Lindström 2001] et a entraîné un commentaire de [Shapiro 2003], commentaire lui-même critiqué dans [Lindström 2006], 3) [Shapiro 2003] constitue la troisième, en restituant l'argument déjà proposée dans [Shapiro 1998] et en l'assortissant d'une critique plus technique, accompagnée de trois échappatoires possibles présentées avec leur discussion ; cette critique technique a elle-même été ensuite contestée dans [Lindström 2006]. Tout cela justifie un traitement spécifique du second argument de Penrose que nous allons effectuer, en commençant par présenter cet argument, puis en abordant successivement les trois critiques mentionnées accompagnées des contre-critiques auxquelles elles ont donné lieu. Par rapport au premier argument le second introduit trois innovations : 1) il ne s'appuie pas d'emblée sur un appel à l'évidence quant à la vérité de G ou la consistance du système formel pris en compte, 2) il minimise l'engagement nécessaire à la seule correction des énoncés Π_1 de PA⁴⁰⁰, 3) la démarche utilisée consiste à introduire l'hypothèse mécaniste, puis à montrer qu'elle conduit à une contradiction (ce qui

³⁹⁹ En excluant celle de Torkel Franzen, plus éloignée de l'original et surtout nettement moins fouillée que les trois autres [Franzen 2005, p. 119-123]. En mettant également à part celle très récente de Peter Koellner [Koellner 2016], particulièrement exotique puisque basée sur l'utilisation d'une théorie de la vérité non classique libre de type [type free]. Il est à noter qu'elle conduit à une réponse sceptique, proche de celles de Per Lindström et Stewart Shapiro.

⁴⁰⁰ Ce choix, qui semble légitime dans la mesure où G est un énoncé Π_1 , présente par ailleurs clairement deux avantages : 1) L'engagement demandé ne va pas au-delà de celui de la seule *consistance simple* du système formel (sachant que la Π_1 -correction : correction des énoncés Π_1 par rapport à un modèle, est coextensive avec la consistance simple : figure C du § 1.5), 2) la *vérité* des énoncés Π_1 est *exprimable* dans PA (ce qui n'est pas le cas des énoncés arithmétiques supérieurs de la hiérarchie de Kleene, par le théorème de Tarski de 1933-1936).

n'est pas très loin de la démarche « dialectique » de John Lucas). L'argument informel, tel que présenté par Roger Penrose, est le suivant :

- 1) selon l'hypothèse mécaniste, il existe un système formel F dont les théorèmes Π_1 sont précisément les énoncés Π_1 qui sont « démontrablement » vrais par la raison et l'intuition humaine ;
- 2) la correction de F ressort de l'accord empiriquement constaté de la communauté de l'ensemble des mathématiciens sur la vérité ou la fausseté (au sens naïf) des énoncés Π_1 ;
- 3) à partir de F , il est possible de construire un nouveau système formel plus fort F^* , constitué de l'ensemble des théorèmes Π_1 de F auxquels sont ajoutés ceux issus de l'hypothèse mécaniste (1) ;
- 4) de par l'hypothèse mécaniste (1), si F est correct alors F^* doit l'être aussi ;
- 5) en appliquant l'argument de Lucas à F^* , on obtient un énoncé $\Pi_1 : G^*$ qui est vrai et qui échappe à F , en contradiction avec l'hypothèse mécaniste de départ.

Dans cet argument, les deux étapes problématiques sont bien évidemment (3) et (4). En effet, comme nous le verrons dans l'analyse plus détaillée des critiques techniques formulées, d'une manière ou d'une autre l'interprétation formelle des hypothèses (1) et (2) nécessite de faire intervenir un prédicat de vérité et un prédicat de « connaissabilité » dans le système formel, ce qui va à minima contraindre fortement ce système formel. Par ailleurs, nous n'avons aucune assurance (et il est même peu vraisemblable) que le passage de F à F^* ne va pas introduire des énoncés non- Π_1 ; comment pouvons-nous alors être assurés de la vérité (dans le modèle standard) des énoncés Π_1 issus des énoncés non- Π_1 introduits (c'est une objection mentionnée par Stewart Shapiro [Shapiro 1998, p. 285]) ? Enfin, (4) n'est pas vrai dans le cas général (ce sera la critique principale de [Lindström 2001]). En fait, (4) se ramène à l'utilisation de la remarque de [Gödel 1951] sur « l'incomplémentabilité » des mathématiques, à l'instar de la démarche qui sera adoptée ultérieurement par Torkel Franzén pour le conduire à son concept « d'inépuisabilité » des mathématiques et dont nous avons déjà montré les limites (§ 3.4).

Comme nous l'avons dit, une critique technique de l'argument nécessite au préalable une traduction plus formelle que celle de son exposé par Roger Penrose. Le choix effectué pour cela par David Chalmers consiste à introduire un prédicat de croyance [Chalmers 1995, p. 5], ce qui lui sera reproché par Roger Penrose qui considère qu'il s'agit là d'une hypothèse inutilement plus forte que la sienne [Penrose 1996, p. 6]. Sur cette base, David Chalmers propose comme contre-argument une adaptation du théorème de Löb de 1955, à partir d'une diagonalisation sur ce prédicat de croyance ce qui conduit à une contradiction [Chalmers 1995, p. 6-7]. La réponse

de Roger Penrose est que la diagonalisation conduit à un énoncé qui n'est pas Π_1 [Penrose 1996, p. 6-7]. La critique beaucoup plus précise de Per Lindström est que l'opération de diagonalisation appliquée n'est pas licite, dans la mesure où l'utilisation du théorème de Löb de 1955 ou du second théorème de Gödel de 1931 nécessite que l'énoncé formalisé dans le système formel ait un sens bien défini, ce qui est le cas du prédicat $\text{Con}(S)$ qui traduit la consistance du système formel S^{401} , mais non de l'énoncé traduisant la croyance introduit ici [Lindström 2001, p. 247-249]. Comme le dit avec humour Per Lindström : « L'argument diagonal n'est pas le chapeau du magicien pour lequel il est quelquefois pris » [Lindström 2001, p. 248].

Dans sa propre restitution du second argument de Penrose, Per Lindström ne fait pas appel à un prédicat de « connaissabilité » ou de vérité spécifique. Du coup, il est amené à faire état du *sens*, d'une part de l'acte de reconnaître une démonstration comme vraie (dans le modèle standard) pour l'étape (3), d'autre part de la notion de correction (par rapport au modèle standard) pour l'étape (4). Son objection frontale concerne l'étape (4), qu'il considère comme non justifiée dans le cas général. Il prend pour cela l'exemple classique du système formel F' , constitué en ajoutant à un système formel F supposé consistant, l'énoncé formel traduisant la négation de la consistance de F : $\neg\text{Con}(F)$. Dans ce cas, F' est consistant, mais $F' + \text{Con}(F')$ n'est pas consistant [Lindström 2001, p. 244-245]. Per Lindström conclut cependant son analyse de manière très prudente : « Aussi, il est encore fondamentalement ouvert de savoir si le raisonnement mathématique humain peut ou non être capturé par un système formel calculable ; [...] » [Lindström 2001, p. 249]. Stewart Shapiro suggère que Roger Penrose pourrait répondre en demandant un engagement plus fort en faveur de l' ω -consistance (et plus seulement de la consistance) du système formel [Shapiro 2003, p. 29]. Bien entendu, cela enlèverait à Roger Penrose les avantages de son choix initial, en revenant à un engagement en faveur de la pleine correction de PA (par rapport au modèle standard) et surtout en ne permettant plus une définissabilité de tous les énoncés pertinents du système formel dans le système formel (par le théorème de Tarski de 1933-1936).

Comme nous l'avons indiqué, la restitution par Stewart Shapiro du second argument de Penrose est traitée en 1998 [Shapiro 1998, p. 284], tandis que sa critique technique détaillée se trouve dans [Shapiro 2003]. Dans cette restitution, l'auteur introduit un prédicat de « connaissabilité » $\kappa(x)$: l'énoncé codé par x est humainement connaissable comme vrai. Le contre-argument qu'il propose est alors basé sur l'application du point fixe à la fonction $\neg\kappa(x)$

⁴⁰¹ Nous avons évoqué cette contrainte (pour le second théorème de Gödel de 1931) au § 2.3.

[Shapiro 2003, p. 31], ce qui conduit à une contradiction (d'une manière analogue au contre-argument de [Chalmers 1995]⁴⁰²). L'auteur discute alors successivement trois stratégies possibles susceptibles de contourner son contre-argument. La première voie consiste en une restriction syntaxique du champ d'application de $\kappa(x)$ aux seuls énoncés Π_1 ; mais on ne peut alors plus passer à l'étape (4). La deuxième voie porte sur une stratification sémantique du prédicat $\kappa(x)$ à la Tarski ; mais on ne peut plus alors passer à l'étape (5). La troisième voie vise à éliminer les énoncés non désirés issus des extensions de $\kappa(x)$ sur la base de leur sens. Cela conduit alors à une complexification notable du système formel, puisqu'il est nécessaire d'introduire un prédicat de vérité avec la nécessité d'appliquer le prédicat de « connaissabilité » à des énoncés contenant ce prédicat de vérité. Cela implique de disposer d'un langage sémantiquement clos (Stewart Shapiro suggère pour cela l'appel à une théorie de la vérité de Kripke de 1975). Stewart Shapiro fait alors remarquer qu'en supposant même qu'une telle construction s'avère possible, celle-ci reviendrait à ajouter des contraintes extrêmement fortes au modèle initialement envisagé par le mécaniste et celui-ci pourrait légitimement défendre que ce sont ces hypothèses et non pas le modèle initial qui sont réfutées. D'une manière plus générale, il apparaît finalement que cette troisième voie nous ramène à peu près au point de départ du premier argument de Penrose, comme cela transparait dans une remarque conclusive de Stewart Shapiro [Shapiro 2003, p. 35].

Le deuxième argument de Penrose apparaît ainsi indéniablement plus sophistiqué que le premier (ou celui de Lucas). Comme il est présenté de manière informelle, sa critique détaillée nécessite au préalable une reconstruction, et pour cela l'introduction d'hypothèses sans être bien entendu assurée de couvrir toutes les éventualités. Cela peut expliquer les commentaires finaux relativement mesurés de Per Lindström et Stewart Shapiro. C'est sans doute également cela qui a conduit Roger Penrose à affirmer en 2011 que les critiques qui lui ont été opposées n'ont pas prouvé l'impossibilité d'une reconstruction qui permette à l'argument de fonctionner. On est alors renvoyé au problème général de la responsabilité de la charge de la preuve, bien illustré par le commentaire final de Stewart Shapiro en 2003 [Shapiro 2003, p. 39]. Au-delà de ces remarques, on peut constater que l'argument n'élimine pas les objections majeures qui ont été apportées à celui de Lucas (et le premier de Penrose), il ne fait (au mieux) qu'en réduire la portée. Un engagement, à minima sur la consistance de PA reste nécessaire, que l'auteur justifie

⁴⁰² Il n'est donc nullement surprenant que Per Lindström reconduise à l'encontre de ce contre-argument, la critique qu'il avait formulée contre celui de David Chalmers [Lindström 2006, p. 232-234]. Nous ignorons s'il y a eu une réponse de Stewart Shapiro ; Il n'y en a en tout cas pas trace dans les références mentionnées dans [Penrose 2011].

toujours par un engagement explicite réaliste fort⁴⁰³ [Penrose 1997, p. 18]. Au-delà de cette analyse, l'objection issue de [Dubucs 1992] contre l'argument de Lucas s'applique encore ici : souscrire à l'engagement demandé sur la correction (par rapport au modèle standard) d'un système formel tel que PA, même limité aux théorèmes Π_1 (étape1), suppose tout de même de faire le choix du modèle standard, ce *qui ne peut pas être publiquement communiqué*. Il est à noter que l'auteur fait explicitement allusion à cette difficulté, mais sans en tirer toutes les conséquences [Penrose 1994 (1995), p. 101]. Il apparaît ainsi que malgré le caractère indubitablement plus subtil du nouvel argument de Penrose par rapport à ceux de Penrose de 1989 et de Lucas de 1961, celui-ci achoppe finalement aux mêmes écueils : d'une part la nécessité d'un engagement ontologique en faveur d'une forme minimale de réalisme mathématique (concernant les entiers naturels), d'autre part l'appel à une convention qui pour empiriquement intersubjective qu'elle soit, n'est pas susceptible d'une communication publique.

6.4 L'argument de Wright

Dans le paysage somme toute assez uniforme du débat issu de [Lucas 1961] et [Penrose 1989], il convient de mentionner la situation tout à fait particulière d'une reprise du même argument par Crispin Wright en 1994 et 1995, mais dans le cadre (d'une restitution) de la logique *intuitionniste* : [Wright 1994 § 6] et surtout [Wright 1995]. Cette proposition n'a donné lieu, à notre connaissance, qu'à une seule critique, mais détaillée, émanant de Michael Detlefsen : [Detlefsen 1995]. Le projet de Crispin Wright est en fait très ambitieux, puisqu'il correspond à rien de moins que de reformuler le premier théorème de Gödel de 1931 au sein d'un système formel basé sur la logique intuitionniste, « qui n'implique aucune référence, ni explicite ni implicite, à la consistance de S [le système formel] » [Wright 1995, p. 92 ; le commentaire est nôtre]. Un tel résultat se traduirait par un renforcement notable de celui de Lucas, mais bien entendu, il vient en contradiction directe avec celui de Gödel de 1933 et l'on devine que, soit la restitution prise en compte de l'arithmétique associée à la logique intuitionniste est fortement différente de HA, soit la démonstration de Crispin Wright est défectueuse. A la suite d'une démonstration informelle peu détaillée, l'auteur prétend parvenir

⁴⁰³ Réalisme que l'auteur justifie par ce que Gödel aurait *démontré*, au travers des théorèmes de 1931 [Penrose 1994 (1995), p. 103].

au même résultat que Lucas et Penrose [Wright 1995, p. 99]. D'une manière assez originale dans ce débat, mais malheureusement non développée, il émet cependant une réserve sur la pertinence du choix d'une machine de Turing pour rendre compte du mécanisme [Wright 1995, p. 101]. Il estime par ailleurs qu'une véritable réfutation du mécanisme devrait faire intervenir le contexte de la découverte [Wright 1995, p. 101]. Il conclut enfin son argument par une réserve quant à l'issue du débat dans le sens indiqué par Lucas et Penrose [Wright 1995, p. 101].

La réponse de Michael Detlefsen est tout d'abord que même en supposant que la démonstration de Crispin Wright soit correcte, sa réfutation du mécanisme serait faible dans la mesure où elle conduit à l'affirmation : « Pour tout système formel F, il y a une théorie intuitionniste informelle I qui n'est pas formalisable dans F » et non pas l'affirmation converse que l'on pourrait plus raisonnablement opposer au mécanisme : « Pour toute théorie intuitionniste informelle I, il n'y a pas de système formel F qui peut formaliser I » [Detlefsen 1995, p. 114-115]. Mais Michael Detlefsen montre surtout, à partir d'une reconstruction technique détaillée de l'argument de Crispin Wright, que celui-ci contient en fait sans surprise, de manière cachée, une affirmation de consistance substantielle pour la théorie I [Detlefsen 1995, p. 114-115]. L'argument de Wright se ramène alors à ceux de Lucas et de Penrose de 1989 et amène bien entendu les mêmes critiques.

6.5 L'argument de la logique paraconsistante

Sur le plan historique, cet argument a été présenté par Graham Priest dans [Priest 1979], puis complété dans [Priest 1984] et [Priest 1987] à la suite des critiques qui lui ont été adressées par Charles Chihara dans [Chihara 1984]. Il est ensuite resté en sommeil durant plus de vingt ans, avant d'être exhumé sans modifications dans [Berto 2009a]. En dehors de celle de Charles Chihara, la principale critique qui en a été faite émane de Neil Tennant dans [Tennant 2004]. Du point de vue de son contenu, cet argument est extrêmement proche de ceux de Lucas (Graham Priest y fait d'ailleurs explicitement allusion [Priest 1984, p. 168]) et de Penrose (de 1989). L'interprétation en est toutefois radicalement différente, en ce sens qu'elle se traduit par l'affirmation que tout système formel rendant compte des mathématiques naïves doit être (simplement) *inconsistent*, retrouvant en cela l'échappatoire mécaniste suggérée par John Lucas [Lucas 1968, p. 157] et [Lucas 1999, p. 122]. La structure de l'argument est la suivante :

- 1) considérons le système formel P qui, par construction, formalise l'ensemble des moyens de preuves, donc tout ce qui nous permet d'affirmer qu'un théorème est vrai (au sens naïf) ;
- 2) par l'application du premier théorème de Gödel de 1931 à P, si P est simplement consistant, il existe au moins un énoncé vrai au sens naïf que P ne peut pas établir ;
- 3) ces deux résultats sont contradictoires, donc P est simplement inconsistant et la logique associée à l'établissement de la vérité au sens naïf est la logique paraconsistante et non la logique classique.

Cet argument se ramène à considérer que les mathématiques naïves peuvent être formalisées dans un système sémantiquement clos, auquel cas la contradiction s'ensuit directement (dans le cadre d'une théorie de la vérité classique à la Tarski).

La difficulté réside bien entendu dans la justification de la première étape, étape que Graham Priest a décomposée en deux points : 1a) les mathématiques naïves dans leur ensemble (les axiomes ainsi que les méthodes de preuves) peuvent être exprimées dans un système formel, 1b) un tel système formel présente les conditions d'effectivité requises pour l'application du premier théorème de Gödel de 1931. Les éléments présentés à l'appui de cette justification sont variés. Il y a d'abord un recours à la thèse de Church-Turing, au motif que nous savons empiriquement reconnaître une preuve naïve quand nous en voyons une [Berto 2009a, p. 203]. Nous avons suffisamment insisté sur les critères d'intersubjectivité qui accompagnent l'exploitation d'une machine de Turing (§ 5.4) pour être en mesure d'affirmer qu'un simple constat empirique général ne saurait autoriser son emploi. Toujours sur un plan technique, l'auteur aborde également frontalement le cas des systèmes formels non effectifs, comme ceux traités dans [Feferman 1962] qu'il cite explicitement [Priest 1984, p. 169]. Il répond alors que de tels systèmes formels restent régis par des métarègles, qui pourraient être elles-mêmes incluses dans un système formel plus large, qui serait alors récursif parce que ces règles peuvent être enseignées et apprises. Cette dernière remarque recouvre l'argument empirique qui énonce le caractère social d'une preuve, duquel la récursivité découlerait nécessairement parce que « si la preuve n'était pas une notion récursive, le processus par lequel elle est apprise serait inintelligible » [Priest 1984, p. 168]. Mais une remarque analogue à celle effectuée sur la machine de Turing s'applique également ici : les caractéristiques d'effectivité demandées à un système formel traduisent des exigences d'intersubjectivité beaucoup plus fortes qu'un simple constat empirique général, qui pourrait quant à lui tout aussi bien être expliqué par un accord d'ordre culturel (donc contingent) ou bien par un appel implicite à une intuition humainement partagée. L'appel au constat empirique de l'harmonie entre les mathématiciens sur la reconnaissance de la vérité (naïve) d'une preuve [Priest 1987, p. 41], qui rejoint, comme nous

l'avons vu, une position analogue de Roger Penrose, tombe évidemment dans le même travers. En dernier recours il est fait un appel réitéré à l'évidence ou au caractère de conjecture plausible de la thèse, dont l'auteur a l'honnêteté de reconnaître le caractère circulaire [Priest 1984, p. 166]. Le caractère non probant de tous les arguments présentés par Graham Priest à l'appui de sa thèse est bien relevé par Neil Tennant, qui en déduit quant à lui et à rebours, l'invalidation de cette thèse par l'application du théorème de Gödel de 1931 [Tennant 2004, p. 383]. Mais l'argument le plus fort que l'on peut opposer à l'argument dialéthéiste est sans aucun doute celui de [Dubucs 1992] que nous avons déjà évoqué contre John Lucas et contre Roger Penrose : les critères de reconnaissance de la vérité dans le modèle standard de l'arithmétique (expression rigoureuse de la notion de vérité naïve) ne sont pas publiquement communicables. Ces critères ne peuvent donc pas être traduits par un système formel effectif et ce n'est pas par ce biais que l'on peut justifier l'utilisation de la logique paraconsistante.

Si nous récapitulons l'ensemble de l'analyse effectuée, nous sommes partis de la suggestion générale de Hilbert selon laquelle la pensée humaine à l'œuvre en mathématiques peut être réduite à une activité mécanique⁴⁰⁴, pour en venir à la question plus modeste de savoir ce qu'il en est pour la seule part de cette pensée mathématique en jeu dans le contexte de la vérification⁴⁰⁵. De manière plus précise, nous nous sommes intéressés à la réponse négative affichée successivement par Gödel, sur la base du second théorème de 1931 et par John Lucas et Roger Penrose, sur la base du premier théorème de Gödel de 1931. Ces trois auteurs défendent en commun la position forte de l'existence de procédures humaines non mécanisables pour la vérification de théorèmes en arithmétique, révélées par les théorèmes de Gödel de 1931. Dans le cas de Gödel, celui-ci justifie cette position par un engagement ontologique en faveur d'un réalisme au sens fort, tout en la dissociant de ses théorèmes de 1931, ce qui n'est pas problématique. En sens inverse, John Lucas et Roger Penrose affirment déduire la même position d'une application du premier théorème de Gödel tout en ne reconnaissant aucun engagement philosophique associé, ce qui est évidemment beaucoup plus ambitieux. L'analyse détaillée que nous en avons faite montre cependant qu'une telle gageure n'est pas tenue. L'argument de John Lucas repose bien en fait sur un engagement philosophique sous-jacent, et

⁴⁰⁴ Dans ce débat, le terme « machine » que nous avons introduit en tout début est trompeur : ce ne sont pas des machines, au sens de quelque dispositif physique que ce soit, qui sont en jeu, mais bien le caractère « mécanique » d'une démarche (en l'occurrence de la vérification en mathématiques). On retrouve ici très exactement le point que nous avons souligné à propos des machines de Turing, qui sont en fait des êtres humains qui calculent (qui vérifient).

⁴⁰⁵ Cette limitation ne s'oppose pas à la possibilité d'existence d'algorithmes de *recherche* de preuve reposant sur la formalisation, au moyen d'une démarche heuristique, donc avec une autre finalité que la vérification (par exemple l'algorithme AProS décrit dans [Sieg 2010]).

la technicité introduite (en l'occurrence l'appel au premier théorème de Gödel) n'est qu'un déguisement. Son raisonnement s'avère circulaire, et sa position le reflet de prémisses anti-mécanistes cachées dans l'affirmation de l'évidence de la correction des théorèmes d'un système formel de l'arithmétique par rapport au modèle standard. La même critique s'applique immédiatement au premier argument de Penrose, l'engagement philosophique étant alors de plus étendu à un réalisme en un sens fort. Elle s'applique enfin également, mais de manière moins directe, aux différentes reconstitutions techniques que l'on peut raisonnablement imaginer du second argument de Penrose. Nous rejoignons ici une position très consensuelle des philosophes de la logique contre la position de John Lucas et Roger Penrose, mais sur la base d'une argumentation beaucoup plus précise que celle qui a été généralement proposée. Cette argumentation fait en particulier ressortir un argument très fort, émanant de Jacques Dubucs, mais qui est quasiment passé inaperçu dans ce débat pourtant nourri : les caractéristiques spécifiquement humaines dont une machine de Turing ne peut pas rendre compte ne sont pas publiquement communicables⁴⁰⁶. Nous retrouvons ainsi le critère d'intersubjectivité caractéristique du contexte de la vérification, dont nous avons déjà montré qu'il était inhérent à l'exploitation de la machine de Turing elle-même (§ 5.4). Cette analyse laisse tout de même en lice une forme affaiblie de l'argument de John Lucas et Roger Penrose : il reste à expliquer l'origine de l'accord empirique que l'on peut constater entre les mathématiciens, sur les choix *privés* qu'ils effectuent pour décider de la vérité d'un théorème (faisant suite à une démonstration informelle). Toute cette discussion nous conduit assez naturellement à une question plus large, qui est au cœur de l'ensemble de notre réflexion et que nous allons aborder maintenant : celle qui concerne les affirmations d'une limitation de la connaissance humaine introduite par les théorèmes « négatifs ».

⁴⁰⁶ Cet argument est tout aussi dévastateur pour Graham Priest, qui a utilisé une démarche analogue à celle de John Lucas avec une finalité différente : afin de justifier l'usage de la logique paraconsistante.

7 Les théorèmes « de limitation des formalismes » représentent-ils une limitation de la connaissance ?

Il s'avère que très peu d'auteurs et aucun récent se sont engagés dans le sens de limitations *humaines* profondes révélées par les théorèmes qui ont été dits « de limitation des formalismes ». De fait, ce n'est que durant les années 1940 à 1960 que successivement Emil Post, John Myhill et surtout Jean Ladrière se sont exprimés en faveur d'une limitation de la connaissance, voire de la raison, révélée par les théorèmes de Gödel de 1931 et apparentés. Enfin, sur un plan historique, on ne peut ignorer la position critique beaucoup plus large affichée par Wittgenstein sur les théorèmes de Gödel de 1931. Publiée dans les années 1950 : [Wittgenstein 1956] et unanimement condamnée à cette époque, il se trouve que celle-ci a été objet d'une tentative de réhabilitation, d'abord par Stuart Shanker : [Shanker 1988], puis plus récemment par Juliet Floyd et Hilary Putnam : [Floyd, Putnam 2000].

7.1 Les « Remarques sur les fondements des mathématiques » de Wittgenstein concernant les théorèmes de Gödel de 1931

Publiée en 1956 à titre posthume, les « Remarques sur les fondements des mathématiques » : [Wittgenstein 1956], qui ne sont qu'une compilation de notes choisies et ordonnées par l'éditeur, contiennent plusieurs remarques pouvant être rattachées aux théorèmes de Gödel de 1931. Il s'agit très précisément des § 5, 6, 7, 8, 11, 16 et 17 de l'annexe III de la partie 1 ainsi que les § 19, 21 et 22 de la partie 7⁴⁰⁷. Dès après leur parution ces remarques ont donné lieu à une critique sévère de tous les auteurs qui se sont exprimé à leur sujet [Anderson 1957, p. 457], [Goodstein 1957, p. 551], [Dummett 1959 (1978), p. 166] et [Bernays 1959, p. 156]⁴⁰⁸. Au début des années 1970, donc à la fin de sa vie, Gödel a lui-même formulé des critiques similaires :

⁴⁰⁷ A la suite de l'analyse historique détaillée effectuée par Victor Rodych, il est possible de dater ces notes plus précisément que l'indication donnée dans [Wittgenstein 1956] : septembre 1937 avec une reprise en 1938 pour la partie 1 et juillet 1941 pour la partie 7 [Rodych 2003, p. 294].

⁴⁰⁸ Il faut également mentionner [Kreisel 1958c], certes critique, mais d'une manière inhabituellement mesurée et charitable (compte tenu du caractère de l'auteur).

Dans la mesure où mon théorème sur les propositions indécidables est concerné, il est effet clair à propos du passage que vous citez que Wittgenstein ne l'a *pas* compris (ou prétendu ne pas le comprendre). Il l'interprète comme une sorte de paradoxe logique, tandis qu'il est en fait juste le contraire, nommément un théorème au sein d'une part absolument non controversée des mathématiques, à savoir la théorie finitaire ou combinatoire des nombres. Incidemment, le passage global que vous citez me paraît être du non-sens. Voir par exemple la remarque sur « la peur superstitieuse de la contradiction de la part des mathématiciens ». [Lettre de Gödel à Karl Menger du 20 avril 1972, Gödel 1972c (2003), p. 132-133]

Russell interprète évidemment mal mon résultat [de 1931] ; cependant, il le fait d'une manière très intéressante, [...]. En contraste, Wittgenstein dans son ouvrage posthume, avance une interprétation incorrecte complètement triviale et inintéressante.

[Lettre de Gödel à Abraham Robinson du 2 juillet 1973, Gödel 1973 (2003), p. 201]

L'affaire en était restée là jusqu'à la fin des années 1980, la communauté des philosophes ayant conclu qu'il s'agissait là « du produit insignifiant de manière surprenante, de la part d'un esprit étincelant », pour reprendre une formule de Georg Kreisel [Kreisel 1958c, p. 158]. Puis, en 1988, Stuart Shanker a publié un très long article (qui l'apparente à un ouvrage) : [Shanker 1988] dans lequel il propose une justification des propos incriminés de Wittgenstein, pour en montrer la cohérence *dans le cadre de pensée de Wittgenstein en philosophie des mathématiques*. Curieusement, à notre connaissance, cet article n'a suscité que très peu de réactions⁴⁰⁹. Enfin, à la fin des années 1990 et au début des années 2000, Juliet Floyd d'une part : [Floyd 1995] et [Floyd 2001] et Victor Rodych de l'autre : [Rodych 1999] et [Rodych 2002], ont entrepris chacun de leur côté une réhabilitation partielle de Wittgenstein sur ce sujet. Surtout, et cette fois de manière remarquée, Juliet Floyd et Hilary Putnam ont présenté une interprétation audacieuse en vue de le réhabiliter complètement : [Floyd, Putnam 2000]. Cette proposition a cependant été elle-même l'objet de critiques, de Mark Steiner : [Steiner 2001], et surtout de Victor Rodych : [Rodych 2003]⁴¹⁰. Nous allons d'abord présenter puis discuter les propos incriminés de Wittgenstein, afin de déterminer si la violente critique dont ils ont été l'objet jusqu'en 1970 était justifiée ou non. Nous confronterons ensuite notre analyse à la proposition de [Shanker 1988], pour analyser enfin les termes de la dernière controverse citée.

⁴⁰⁹ Gerhard Heinzmann l'évoque brièvement dans [Heinzman 1992], mais n'en propose pas une analyse détaillée.

⁴¹⁰ Il convient également de mentionner le cas particulier de [Berto 2009a], qui inclut une tentative de rendre compte de la position de Wittgenstein (en particulier concernant la contradiction en mathématiques) en faisant intervenir la logique paraconsistante, pour laquelle celui-ci est considéré comme un précurseur [Berto 2009a, p. 208]. Nous avons cependant suffisamment insisté sur les limites de la portée épistémologique de la logique paraconsistante elle-même (§ 3.5 et § 6.5) pour qu'il ne soit pas utile d'y revenir.

La première question qu'il convient de se poser est de savoir si les propos incriminés de Wittgenstein portaient bien sur les théorèmes de Gödel de 1931. Le § 7 est sans ambiguïté sur ce point, puisque Wittgenstein mentionne explicitement ces théorèmes (il se limite en fait manifestement au premier), à trois reprises successives, au travers de remarques quelque peu elliptiques :

Ma tâche n'est pas de parler de la démonstration de Gödel, par exemple, mais de l'esquiver dans mon discours. [Wittgenstein 1956 (1983), § 7 – 19, p. 307]

Le théorème de Gödel qui énonce quelque chose sur lui-même, ne se *mentionne* pas. [...]
[Wittgenstein 1956 (1983), § 7 – 21, p. 309]

On peut demander avec raison quelle est l'importance de la démonstration de Gödel pour notre travail. [...] Aussi étrange que cela paraisse, ma tâche concernant le théorème de Gödel consiste uniquement à mettre en évidence ce que signifie en mathématiques une proposition comme : « en admettant que l'on puisse prouver cela ». [Wittgenstein 1956 (1983), § 7 – 22, p. 310-311]

La seconde question qui a été posée est de savoir si Wittgenstein a même lu l'article de Gödel de 1931 au-delà de son introduction. On dispose ici d'un témoignage tardif (soixante ans plus tard) de Georg Kreisel, qui indiquerait que Wittgenstein n'avait en fait pas lu l'article au début des années 1940 (donc au moment de la rédaction des remarques incriminées) :

Je suis tombé par hasard sur des commentaires par Wittgenstein dans la note de bas de page 4, page 281 des *Foundations of Mathematics* (1950) concernant l'incomplétude formelle, [...]. L'article était une mise à jour de quelque chose que j'avais fait au début des années 1940, mais publié seulement après la guerre. Le fond de cette note de bas de page consistait en une conversation assez brève avec Wittgenstein, dans laquelle il voulait me parler de la preuve de Gödel (*Il me dit qu'il ne l'avait jamais lue, ayant été rebuté par l'introduction*). [Kreisel 1998, p. 119 ; souligné par nous]

Pourtant le § 7 – 22, qui semble avoir été oublié par les protagonistes⁴¹¹, contient une allusion explicite au principe de la démonstration elle-même (au travers du codage gödélien), qui indique que Wittgenstein ne s'est donc pas limité à une lecture de l'introduction. L'explication la plus probable est alors, en suivant la suggestion de Victor Rodych, que Wittgenstein s'est contenté de *parcourir* l'article de 1931 [Rodych 2003, p. 307]. Il n'en reste pas moins que le contenu des propos de Wittgenstein que nous venons de citer, mentionnant explicitement « le »

⁴¹¹ A l'exception de Victor Rodych qui le mentionne explicitement [Rodych 2003, p. 291-292].

théorème de Gödel ou « la » démonstration de Gödel, indique sans équivoque que Wittgenstein se limite à un usage *métaphorique* du premier théorème de Gödel de 1931, en cohérence avec la remarque de Kreisel que nous venons de citer. Il est alors clair que ce choix enlève toute pertinence à une telle analyse *pour ce qui concerne les théorèmes de Gödel de 1931 eux-mêmes*. Si l'on cherche tout de même à comprendre les propos de [Wittgenstein 1956] basés sur un usage métaphorique du premier théorème de Gödel de 1931, il convient de distinguer deux formes de réponses : 1) La contestation qu'il puisse y avoir une proposition « vraie mais indémontrable dans les *Principia Mathematica* (PM) », 2) La possibilité de faire avec des contradictions.

Le premier point est l'objet des § 1 – III – 5, 1 – III – 6, 1 – III – 7 et surtout 1 – III – 8, (qui a été baptisé « le paragraphe notoire » par Juliet Floyd, dans le cadre du débat en jeu) :

Y a-t-il dans le système de Russell [PM] des propositions vraies qui ne doivent pas être prouvées dans son système ? [...]. [Wittgenstein 1956 (1983), § 1 – III - 5, p. 112]

Si donc on demande en ce sens : « Dans quelles circonstances affirme-t-on une proposition dans le jeu russellien ? » la réponse est : A la fin de l'une de ses démonstrations, ou bien comme "principe" (Pp.). [...]. [Wittgenstein 1956 (1983), § 1 – III - 6, p. 112-113]

« Mais ne peut-il y avoir, écrites dans ce symbolisme, des propositions vraies mais indémontrables dans le système russellien ? » - "Propositions vraies", ce sont donc des propositions qui sont vraies dans un *autre* jeu, c'est-à-dire qui peuvent être affirmées en droit dans un autre jeu. [...] Bien sûr que non ; et une proposition que l'on ne peut prouver dans le système russellien est vraie ou fausse en un autre sens qu'une proposition des « Principia Mathematica ». [Wittgenstein 1956 (1983), § 1 – III - 7, p. 113]

J'imagine que quelqu'un me demande conseil ; il dit : « J'ai construit une proposition (appelons-la "P") en symboles russelliens, et au moyen de certaines définitions et transformations, on peut l'interpréter de telle façon qu'elle dise : "P n'est pas démontrable dans le système russellien". Ne dois-je pas dire maintenant de cette proposition que d'un côté elle est vraie, de l'autre indémontrable ? [...]. » 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 de russellien ; [...] Si tu admets que la proposition est démontrable dans le système russellien, de ce fait elle est vraie *dans le système russellien* et il te faut de nouveau renoncer à l'interprétation « P n'est pas démontrable ». [...] En outre : si la proposition doit être fausse dans un sens autre que le sens russellien cela n'est pas en contradiction avec le fait qu'elle soit démontrée dans le système de Russell. [...] [Wittgenstein 1956 (1983), § 1 – III - 8, p. 113-114]

Comme l'a souligné avec juste raison Stuart Shanker dans [Shanker 1988], pour comprendre la position de Wittgenstein il faut la placer dans le cadre de la conception particulière des mathématiques de celui-ci. Cette conception peut être formulée par la définition suivante (en utilisant un vocabulaire moderne) : la vérité (ainsi que la signification) d'un énoncé mathématique est *uniquement* définie par sa dérivabilité au sein d'un système formel ; elle est donc *relative à celui-ci et indépendante* de toute référence qui lui serait *extérieure*. De ce fait, vérité, prouvabilité et dérivabilité se confondent et tout système formel bien construit est nécessairement syntaxiquement complet (la notion de métalangage est évacuée). Par ailleurs, la notion de complétude sémantique n'a pas de sens (la notion de modèle est évacuée). Enfin, un système formel bien construit formant des énoncés doués de signification est nécessairement consistant. Une telle interprétation a le mérite de la cohérence avec deux idées fortes de Wittgenstein dont la paternité n'est contestée par aucun des protagonistes dans ce débat : 1) la vérité et la signification mathématiques sont définies par l'usage, 2) les preuves mathématiques n'ont aucun impact sur les réflexions en philosophie des mathématiques. On comprend alors que par rapport à cette conception, la proposition P (qui renvoie implicitement à l'interprétation heuristique que l'on trouve dans l'introduction de la démonstration de Gödel de 1931) ne puisse être considéré comme vraie mais non dérivable dans PM. D'une manière plus générale, toute la démarche de Hilbert se rapportant au problème des fondements n'a ici aucun sens, la consistance étant supposée assurée par un choix judicieux des systèmes formels retenus. Il en est de même du refus de l'*ignorabimus*, qui est *ipso facto* assuré. Bien entendu, le problème majeur d'une telle position est qu'elle n'a plus rien à voir avec l'orientation qu'a prise l'histoire du développement de la logique mathématique (théorie de la preuve et théorie des modèles), et cela indépendamment de toute prise de position philosophique sur les mathématiques elles-mêmes. Comme de plus cette position n'est pas explicitée dans les « Remarques sur les fondements des mathématiques » (qui ne sont qu'un recueil de notes), on peut comprendre qu'elle n'ait pas été perçue par les commentateurs non connaisseurs de la pensée de Wittgenstein, d'où l'incompréhension de ceux-ci à la lecture de ces remarques.

Le second point de [Wittgenstein 1956] basé sur un usage métaphorique du premier théorème de Gödel de 1931, porte sur la possibilité de faire avec des contradictions. C'est l'objet des § 1 – III – 11, 1 – III – 12, 1 – III – 16, 1 – III – 17 et 1 – III – 18 :

Admettons que je démontre l'indémontrabilité de P (dans le système russellien) ; par cette démonstration, j'ai démontré P. Si maintenant cette démonstration appartenait au système russellien, - j'aurais alors

démontré simultanément une appartenance et une non appartenance au système russellien. - [...] – Oui, il y a une contradiction. Mais est-elle vraiment gênante ?
[Wittgenstein 1956 (1983), § 1 – III - 11, p. 115]

La contradiction qui survient gêne-t-elle celui qui dit : « Je mens – Donc je ne mens pas. – Donc je mens, etc. » ? [...]. [Wittgenstein 1956 (1983), § 1 – III - 12, p. 115]

La proposition « P est indémontrable » n'a pas le même sens avant et après avoir été démontrée. Si elle est démontrée, elle est la conclusion de la preuve d'indémontrabilité. – Si elle n'est pas démontrée, *on ne sait ce qui* doit valoir comme critère de sa vérité, et son sens est – peut-on dire - encore voilé.
[Wittgenstein 1956 (1983), § 1 – III - 16, p. 116]

[...] Maintenant, pour voir *ce qui* est démontré, regarde la preuve ! [...] Si quelqu'un me demande : « Qu'est-ce qui est le cas – P ou non-P ? », je réponds : P se trouve à la fin d'une démonstration russellienne, donc tu écris dans le système russellien : P ; d'un autre côté il est précisément démontrable et cela s'exprime par non-P, mais cette proposition [non P] ne se trouve pas à la fin d'une démonstration russellienne ; donc n'appartient pas au système russellien. - [...] (La peur superstitieuse et l'adoration du mathématicien devant la contradiction.) [Wittgenstein 1956 (1983), § 1 – III - 17, p. 116-117]

[...] On peut très bien dire que le principe de non-contradiction est faux pour la raison que nous répondons fréquemment à bon escient « oui et non » à une question. Et de la même façon pour la proposition " $\sim\sim p=p$ " : parce que nous employons le doublement de la négation comme son *renforcement* et non pas simplement comme sa suppression.
[Wittgenstein 1956 (1983), § 1 – III - 18, p. 117]

Le lien avec l'introduction de la démonstration de Gödel de 1931 est ici implicite au travers de l'évocation du paradoxe du menteur au § 1 – III - 16, p. 115. Quant à la formule finale du § 1 – III - 17, p. 116-117, qui n'est accompagnée d'aucune explication, on peut comprendre qu'elle ait suscité la réaction de rejet de Gödel que nous avons mentionnée au début de ce paragraphe. D'une manière plus générale, à la lecture de l'ensemble de ces remarques on peut également comprendre que Wittgenstein ait pu être considéré comme un précurseur de la logique paraconsistante [Berto 2009a, p. 208]. Cependant, la justification qu'il donne de l'acceptation d'une non-contradiction, basée sur l'usage habituel du langage naturel avec toutes ses imperfections (§ 1 – III - 18, p. 117), a peut-être un sens dans son cadre de pensée, mais elle n'en a plus du tout dans le cas d'un système formel rigoureux (supposé consistant), qu'il s'agisse de PM ou de PA, ce qui rejoint la critique de Gödel. Un tel écart se retrouve dans l'affirmation de Wittgenstein selon laquelle la proposition P est démontrable dans PM, tandis que non-P ne l'est pas (§ 1 – III - 17, p. 116-117), si on la compare au contenu véritable du

premier théorème de Gödel de 1931⁴¹². On retrouve ici encore, que la position de Wittgenstein en mathématiques n'a rien à voir avec l'orientation qu'a prise l'histoire du développement de la logique mathématique.

Au-delà de son interprétation tout à fait pertinente du cadre de pensée de Wittgenstein en mathématiques, la tentative de Stuart Shanker⁴¹³ de défendre ce cadre dans l'absolu est manifestement désespérée⁴¹⁴. En essayant de justifier que la caractéristique de l'énoncé G d'être vrai (au sens naïf) est à la fois inhérente à la preuve du théorème et découle en fait de la conception platonicienne de Gödel, on peut se demander si l'auteur a lui-même réellement compris la démonstration de 1931 malgré l'analyse technique qu'il tente d'en restituer⁴¹⁵. Par ailleurs, la restitution de la démarche de Hilbert qu'il avance est effectuée à partir de fragments disparates de propos de celui-ci, issus de différents textes sans cohésion d'ensemble ; cela conduit en particulier à des confusions manifestes : entre le programme au sens restreint et son extension au programme au sens large, incluant le refus de *l'ignorabimus*, ainsi qu'entre l'insolubilité et l'incomplétude syntaxique ou l'indécidabilité. Enfin, le fait de se référer en 1988, aux échanges de Gödel à la suite de la publication de 1931, avec Perelman d'une part et Zermelo de l'autre, pour prétendre réfuter la portée du théorème, contre le premier et avec les seconds, se passe de commentaires.

L'interprétation de Juliet Floyd et Hilary Putnam consiste à voir dans les remarques de Wittgenstein sur les théorèmes de Gödel de 1931 (en fait dans le seul § 1 – III – 8, considéré comme le plus critique), une allusion à un point métamathématique (au sens classique) très particulier, effectivement en jeu dans le premier théorème de Gödel de 1931. Ce point concerne l'impact syntaxique et sémantique de la différence entre la consistance (hypothèse nécessaire pour montrer que l'on ne peut pas dériver l'énoncé G du premier théorème) et l' ω -consistance (hypothèse plus forte, nécessaire pour montrer que l'on ne peut pas dériver $\neg G$) [Floyd, Putnam 2000, p. 625-626 et 632]. Pour justifier une interprétation aussi spécifique, les auteurs se fondent sur deux éléments historiques précis :

⁴¹² Où il est montré dans la métalangue, au moyen d'une démonstration pouvant elle-même être formalisée en supposant la 1-consistance de PM, que ni G ni $\neg G$ n'y sont dérivables.

⁴¹³ On peut d'ailleurs noter par exemple que Francesco Berto [Berto 2009a, p. 190-200] reprend quasiment l'interprétation de [Shanker 1988].

⁴¹⁴ Elle est également quelquefois particulièrement prétentieuse, par exemple lorsqu'il est affirmé qu'Hilbert, contrairement à Wittgenstein, « n'a pas compris la nature profonde des mathématiques » [Shanker 1988, p. 211]. Cela rejoint l'affirmation, dans un autre texte et concernant Turing cette fois, selon laquelle les travaux de ce dernier, postérieurs à 1941, représentent une « mauvaise interprétation des résultats [de Turing] de 1936 », en opposition à l'interprétation de Wittgenstein évidemment [Shanker 1987, p. 632].

⁴¹⁵ Le fait qu'il s'appuie pour cela exclusivement sur deux ouvrages de vulgarisation : [Nagel, Newman 1958] et [Hofstadter 1979], permet d'en douter.

1) la mention en 1957 puis en 1972 (donc vingt puis quarante ans après), par Reuben Goodstein qui a côtoyé Wittgenstein dans les années 1930, que celui-ci aurait eu une telle réaction en 1935 [Goodstein 1957, p. 551] et [Goodstein 1972, p. 279] ;

2) la publication en 1938 par Alison Watson, élève de Wittgenstein, d'un article sur ce sujet précis : [Watson 1938], sur lequel Wittgenstein aurait porté une appréciation flatteuse [Floyd, Putnam 2000, p. 626-628].

Une telle interprétation ne résiste cependant pas à l'analyse historique détaillée réalisée par Victor Rodych sur l'ensemble des écrits de Wittgenstein accessibles non publiés. Celui-ci précise qu'il n'y a trouvé nulle part d'allusion à un tel point technique [Rodych 2003, p. 304]. Par ailleurs, le seul fait que Wittgenstein ait émis un avis positif sur le travail d'Alison Watson ne prouve pas qu'il ait compris ni même lu l'article en question. Quant au souvenir sans trace écrite de Reuben Goodstein vingt et quarante ans après, il est permis de douter de la stricte exactitude de son contenu. Globalement, la position défendue par Juliet Floyd apparaît ainsi à la fois artificielle et non plausible. Dans [Floyd 2001], Juliet Floyd propose d'autres éléments d'interprétation des propos de Wittgenstein, mais qui sont beaucoup moins incisifs et encore moins justifiés.

Comme nous l'avons indiqué, Victor Rodych a lui-même proposé une interprétation charitable des remarques de Wittgenstein sur les théorèmes de Gödel de 1931⁴¹⁶. Il accepte la critique classique selon laquelle Wittgenstein s'est borné à contester l'introduction de l'article de Gödel de 1931 (qui n'avait qu'une visée heuristique), en se méprenant sur l'importance de cette introduction sur la preuve elle-même et en imaginant en particulier que la signification de G (et sa « vérité » de par son sens) y joue un rôle crucial [Rodych 1999, p. 182]. Il conteste par contre la critique tout aussi classique (c'est par exemple celle de Bernays [Bernays 1959, p. 156]) selon laquelle Wittgenstein n'avait pas perçu le caractère conditionnel de la consistance (ou ω -consistance) du système formel dans la preuve de Gödel [Rodych 1999, p. 188-189]. L'argumentation qu'il présente à l'appui n'est cependant ni plus précise ni plus justifiée que celle de Juliet Floyd pour son propre compte. Il tente enfin de montrer que le cadre de pensée de Wittgenstein en philosophie des mathématiques n'entraîne pas nécessairement que la soi-disant vérité de l'énoncé G du premier théorème implique un engagement platonicien, qui est refusé par Wittgenstein. L'argument avancé (il considère que les énoncés Π_1 ne sont pas plein de sens [Rodych 1999, p. 174]) apparaît cependant spécieux et non recevable eu égard à toute la discussion que nous avons eue à ce sujet (§ 1.3).

⁴¹⁶ Dont le niveau de charité est intermédiaire entre ceux de [Shanker 1990] et de [Floyd, Putnam 2000].

Il ressort finalement de notre analyse que les remarques de Wittgenstein sur les théorèmes de Gödel de 1931 apparaissent induites par le cadre de pensée que celui-ci s'est forgé en philosophie des mathématiques. La rigidité de ce cadre de pensée lui a en fait interdit de comprendre la question posée par Hilbert dans le cadre de son Programme et partant la réponse apportée par les théorèmes de Gödel de 1931 et d'une manière plus générale, l'esprit de l'ensemble des travaux conduits en logique depuis les années 1930⁴¹⁷. Ce constat est inchangé après les tentatives de défense héroïques mais désespérées de Stuart Shanker et Juliet Floyd (ainsi que de Victor Rodych dans une moindre mesure), qui s'apparentent à un combat d'arrière-garde.

7.2 La position de Post

Les écrits d'Emil Post en philosophie des mathématiques publiés de son vivant se résument à deux commentaires souvent cités :

Mais, masquer cette identification [la thèse de Church-Turing] derrière une définition, cache le fait qu'une découverte fondamentale sur les limitations du pouvoir de mathématisation de l'Homo-Sapiens a été faite et nous aveugle sur la nécessité de sa vérification continue.

[Post 1936, p. 105, n. 8 ; le commentaire entre crochets est nôtre]

La conclusion est inévitable que même pour un tel corps fixé bien défini de propositions mathématiques, [l'arithmétique du 1^{er} ordre], *la pensée mathématique est, et doit rester, essentiellement créative.*

[Post 1944 (1965), p. 316 ; le commentaire entre crochets est nôtre]

Un seul texte plus conséquent existe, rédigé de son vivant⁴¹⁸ mais publié après sa mort : [Post 1941-1965]. C'est dans ce dernier document que l'on trouve l'affirmation explicite de l'auteur selon laquelle les théorèmes « négatifs » révèlent une limitation humaine, en l'occurrence de la « puissance mathématique » :

⁴¹⁷ Par rapport à l'idée générale de Wittgenstein selon laquelle une preuve mathématique fixe le sens de l'énoncé prouvé, on peut mentionner par exemple la remarque tout à fait pertinente de Stewart Shapiro, selon laquelle la démonstration par Wiles de la conjecture de Fermat n'a certainement rien changé au sens de l'énoncé concerné [Shapiro 2013, p. 172].

⁴¹⁸ Proposé pour publication en 1941 et refusé (sachant que la plupart des notes incluses dans l'annexe qu'il contient ont été rédigés dans les années 1920).

[...], la capacité créative des mathématiciens humains a comme inévitable contrepartie les limitations de cela ; en témoigne les problèmes combinatoires absolument insolubles. [Post 1941-1965, p. 345, n. 12]

L'insolubilité du problème de la finitude [de la décision] pour tous les systèmes normaux, et l'incomplétude essentielle de toutes les logiques symboliques [systèmes formels], sont des évidences *des limitations de la puissance mathématique de l'homme*, bien qu'elles soient créatives. Elles suggèrent qu'au royaume de la preuve, comme au royaume des processus, un problème peut être posé, dont nous ne pouvons pas vaincre les difficultés ; [...].

[Post 1941-1965, p. 418 ; les commentaires entre crochets sont nôtres ; souligné par nous]

L'objectif de présenter les arguments d'Emil Post à l'appui d'une telle position se heurte à plusieurs difficultés. Tout d'abord la part proprement philosophique de [Post 1941-1965] ne constitue qu'une petite partie (essentiellement l'annexe) d'un article principalement technique. Qui plus est, cette annexe se présente sous la forme d'une compilation de remarques et d'extraits du journal de l'auteur sans lien véritable⁴¹⁹. S'ajoute à cela l'absence de commentaires sur ces écrits. Si l'on tente malgré tout de restituer la pensée d'Emil Post en philosophie des mathématiques, celle-ci apparaît profondément influencée par Bergson de l'aveu même de son auteur [Post 1941-1965, p. 416-417]. Son idée est que la pensée en mathématiques résulte de la combinaison de deux mouvements : 1) une génération permanente (flux) de créativité (l'auteur parle de « germe créateur » et « d'éther psychique ») « à la Bergson » est continuellement présente, 2) cette génération est saisie et « cristallisée » sous la forme d'une représentation spatiale de symboles « à la Hilbert » (ce que l'auteur nomme « symbolisation »), à la suite d'une réflexion consciente de l'être humain sur le processus créatif lui-même. C'est alors l'impossibilité de saisir entièrement le premier (en particulier tout ce qui a trait au transfini⁴²⁰) par le second, qui est à l'origine des limitations rencontrées :

Mais il y a un petit nombre de choses qui ne peuvent pas être symbolisées, parce que dans la transformation produite par la symbolisation elles réapparaissent nécessairement. Ces choses sont la compréhension, la symbolisation, la manipulation des symboles, l'itération, la perception sensorielle ou la vérification directe, et peut-être quelques autres choses. Celles-ci constituent ainsi les éléments en dehors de quoi la description est construite par l'addition de simples symboles. [Post 1941-1965, p. 433]

⁴¹⁹ La situation est donc proche de celle des *Remarques sur les fondements des mathématiques* de Wittgenstein, à cela près que la compilation a été ici réalisée par Emil Post lui-même.

⁴²⁰ Le lien entre le germe créateur et l'aperception du transfini est explicite page 423.

Si une telle interprétation de la position de Post est correcte, aussi peu précise soit-elle, elle amène deux remarques. Tout d'abord l'origine du premier mouvement (la génération de créativité) reste totalement mystérieuse. On peut sans doute supposer qu'elle doit être considérée comme étant a priori (au sens kantien). Plus gravement, pour le sujet qui nous concerne, si la source des limitations rencontrées était aussi générale en portant sur le processus de « symbolisation » lui-même, elle devrait s'appliquer sans distinction à tout système formel non trivial. Elle devrait donc s'appliquer aussi bien à l'arithmétique du 1^{er} ordre monadique qu'à l'algèbre et la géométrie élémentaire. Une telle position est alors invalidée, dans le premier cas par le résultat de Presburger de 1929 et dans le second par celui de Tarski de 1940-1967, 1948 (voir le § 5.1). Il est peut-être possible d'envisager le renforcement d'une telle explication qui permettrait d'exclure ces deux situations, mais cela impliquerait un traitement beaucoup plus détaillé qui reste à faire et qui ne se trouve en tout cas pas dans les travaux publiés d'Emil Post.

7.3 La position de Myhill

En 1952, dans un article remarqué portant sur une analyse des théorèmes de Gödel de 1931 et de Church de 1936, John Myhill a proposé de manière informelle une circonscription possible des « propriétés » ou des « caractères » (en mathématique) en trois classes : « effective », « constructive », et « prospective » : [Myhill 1952b]⁴²¹. Il est alors assez naturel d'associer les deux premières citées à deux classes bien définies d'*énoncés* arithmétiques : la récursivité pour la propriété « d'effectivité » et l'énumérabilité récursive pour celle de « constructivité » (c'est l'interprétation que propose par exemple Jean-Paul Delahaye [Delahaye 1994, p. 65]). Pour la définition de la troisième, voici les définitions très larges que l'auteur en donne :

Un caractère prospectif est un caractère que nous ne pouvons ni reconnaître ni créer par une série d'actes raisonnés mais en général imprédictibles. (Par imprédictibles je ne veux pas dire, bien sûr, sans principe générateur, mais simplement d'une telle sorte que l'on ne peut pas dire à l'avance si un objet donné peut être produit.) [Myhill 1952b, p. 190]

Un caractère est prospectif s'il n'est ni effectif ni constructif, et si néanmoins quelque humain peut clairement envisager une idée de ce caractère. [Myhill 1952b, p. 191]

⁴²¹ Hao Wang proposera ultérieurement une circonscription très proche, mais sans développement et donc sans justification [Wang 1954, p. 263].

Sur la base de cette notion vague de caractère « prospectif », il est alors conduit à la double affirmation forte : 1) qu'il existe des limitations humaines essentielles introduites par les théorèmes de Gödel de 1931, de Church de 1936, et de Turing de 1936-1937, 2) qu'une telle limitation est une condition nécessaire à l'existence de la créativité en mathématique⁴²². En 1960, le même auteur propose un troisième mode d'appréhension en mathématiques, au-delà des modes syntaxique et sémantique, qu'il nomme « apodictique » et qui fait intervenir une notion de prouvabilité en un sens absolu : [Myhill 1960]. Une telle suggestion rejoint une affirmation similaire effectuée dans [Myhill 1952b] concernant une notion de vérité mathématique en un sens également absolu, avec pour exemple l'énoncé G du 1^{er} théorème de Gödel de 1931⁴²³ [Myhill 1952b, p. 184], sachant que l'auteur associe lui-même explicitement les deux notions de vérité mathématique en un sens absolu introduites respectivement en 1952 et 1960 [Myhill 1960, p. 467]. Le caractère absolu de la notion de vérité utilisée par John Myhill est formulé de manière non ambiguë :

La vérité que nous espérons être une propriété absolue des propositions ; on ne parle pas de vérité dans un système, mais simplement de vérité. Nous supposons que le métalangage est véridique, c'est-à-dire que tout ce qui est prouvable dans ce langage est vrai. [Myhill 1952b, p. 184-185]

Comme justification, l'auteur tente de montrer que la validité du théorème de Gödel de 1931 lui-même implique le recours à une certaine forme de platonisme [Myhill 1952b, p. 184]. Il est toutefois conscient des limites de la justification qu'il propose, ce qui le conduit à défendre sa thèse principale *sans* nécessiter un engagement réaliste en mathématiques [Myhill 1952b, p. 189]. Mais bien entendu, hors engagement ontologique réaliste, sans aucune autre justification fournie en sa faveur, l'appel à une notion de vérité (et de prouvabilité) *absolue* doit être considéré comme quelque chose de l'ordre de l'évidence, ce qui ne correspond pas à autre chose qu'à un *acte de foi*. John Myhill s'exprime alors sans ambiguïté en faveur d'une limitation humaine essentielle introduite par les théorèmes « négatifs » :

Cette indétermination [introduite par les théorèmes de Church de 1936 et de Turing de 1936-1937] ne provient pas d'un caractère objectif des séries, encore moins de l'irruption non envisagée de facteurs

⁴²² Une telle affirmation est donc présentée sous la forme de la converse de l'affirmation d'Emil Post (à qui l'auteur fait explicitement référence [Myhill 1952b, p. 165]), que nous avons déjà citée [Post 1941-1965, p. 345, n. 12].

⁴²³ Ce qui constitue ainsi une anticipation du fond de la justification qui sera proposée dans [Lucas 1961] et [Penrose 1989] (§ 6.2 et 6.3).

étrangers en son sein, mais plutôt d'une *limitation radicale du pouvoir humain* d'envisager les sorties totales de séries d'actes, dont le principe générateur est envisagé avec une complète précision.

[Myhill 1952b, p. 180 ; le commentaire entre crochets est nôtre ; souligné par nous]

C'est ici que nous avons d'abord eu l'évidence conclusive d'une *limitation de la connaissance essentielle et non accidentelle*, et du fait que cette ignorance est la contrepartie de la créativité.

[Myhill 1952b, p. 192 ; souligné par nous]

Pour l'auteur, qui ne fournit aucune autre justification, le caractère absolu d'une telle limitation est lié au caractère absolu de la vérité, par exemple de G. Il repose donc sur l'acte de foi associé à ce dernier et ne peut prétendre, de ce fait, à aucune universalité. L'affirmation du caractère positif d'une telle limitation, en tant que permettant la créativité en mathématique, est certainement le point le plus original de la position de John Myhill⁴²⁴. L'explication qu'il en donne est la suivante :

Si la puissance humaine de reconnaissance [associée à l'aptitude à percevoir la vérité en un sens absolu] devenait comparable à la puissance de production raisonnée [associée à la « constructivité » et à « l'effectivité »], soit par un accroissement de la première, soit par une réduction de la dernière, la possibilité de création mathématique cesserait sur le champ.

[Myhill 1952b, p. 180 ; les commentaires entre crochets sont nôtres]

L'idée n'est donc autre que celle que nous avons déjà évoquée au § 2.1, sous la forme de la crainte exprimée par un certain nombre de mathématiciens que la décidabilité (a fortiori la complétude syntaxique) d'un système formel associé à un domaine particulier des mathématiques, entraîne la fin de l'inventivité et de la créativité nécessaires pour découvrir (ou inventer) un théorème de ce domaine. L'originalité est ici de retourner cette crainte sous la forme à la fois d'une *vertu* et de la *signification* des théorèmes « négatifs » :

La substance du théorème de Church est que ces prémisses sont incomparables, c'est-à-dire que notre créativité distance notre capacité à anticiper l'issue de cette créativité. Dans la mesure où si ce n'était pas le cas, il serait inapproprié de parler de création, nous pouvons dire que ces théorèmes affirment la créativité même au niveau pré-mathématique de la logique : [...]. [Myhill 1952b, p. 180]

⁴²⁴ Comme nous l'avons vu, le lien entre la limitation introduite par les théorèmes « négatifs » et la créativité (en mathématique) a déjà été introduit antérieurement par Emil Post, mais il n'est apparu que dans un document publié bien après la parution de l'article correspondant de John Myhill.

Qu'il soit présenté comme une vertu ou comme une crainte, cet argument achoppe néanmoins au même problème lié à la double confusion entre : 1) l'établissement d'une démonstration et la vérification de cette démonstration, et 2) une démonstration informelle et une démonstration formelle (dérivation). Dans le cadre de la *vérification* d'une démonstration donnée, le souci de rigueur conduit assez naturellement à une exigence de formalisation. Mais dans le cadre de l'*établissement* d'une démonstration, le passage par la formalisation n'est pas nécessaire et n'est même pas souhaitable, justement pour laisser la plus grande latitude créative au mathématicien. Quant à la décidabilité d'un système formel, elle ne concerne que l'existence *en principe* d'une démonstration formelle, qui, même dans les cas où elle est vérifiée, ne donne aucune indication sur la démonstration informelle que devra utiliser le mathématicien *en pratique*⁴²⁵. La distinction méthodologique classique entre l'établissement et la vérification d'une démonstration mathématique, invalide donc le lien introduit par John Myhill entre créativité en mathématique et indécidabilité (ou incomplétude syntaxique) d'un système formel. Ajoutée à la critique précédente, montrant que l'affirmation selon laquelle les théorèmes « négatifs » constituent une limitation de la connaissance humaine ne repose que sur un acte de foi (en faveur de l'existence d'une vérité absolue en arithmétique), en dehors d'un tel engagement philosophique il ne reste à peu près rien de la force de la thèse proposée par John Myhill en 1952 et 1960.

7.4 La position de Ladrière

Auteur particulièrement prolifique (sa bibliographie comporte plus de cent pages), Jean Ladrière a publié durant toute sa carrière (de 1949 à 1990) sur les implications philosophiques des théorèmes de Gödel de 1931 et apparentés⁴²⁶, comme l'indique cette bibliographie : [Ladrière 2005b]. En y regardant de plus près, l'on s'aperçoit cependant qu'il a cessé de publier intensément (et dans les revues les plus sérieuses) sur ce sujet à partir de la fin des années 1960. Par ailleurs, alors qu'il y défendait une position philosophique très engagée dans les années 1960, cet engagement a totalement disparu trente ans plus tard dans les publications les plus

⁴²⁵ Il suffit de reprendre l'exemple que nous avons déjà utilisé du théorème de Tarski de 1940-1967, 1948 (par ailleurs explicitement mentionné par John Myhill [Myhill 1952b, p. 182]), dont le contenu particulièrement non effectuable (voir le § 5.1) ne fournit aucun moyen utilisable pour établir la preuve de tout théorème d'algèbre ou de géométrie élémentaire que l'on peut pourtant montrer tel au sein du système formel associé.

⁴²⁶ Nous avons déjà évoqué à deux reprises la position de Jean Ladrière en philosophie de la logique et des mathématiques : d'une part son opposition au principe même du programme de Hilbert au sens restreint (§ 2.1), d'autre part sa défense du caractère constructif de la démonstration du théorème de Gentzen de 1936 (§ 3.1).

tardives des années 1990. Nous ignorons la raison de ces deux évolutions⁴²⁷. Dans les années 1960, il a longuement exposé son point de vue principalement dans trois documents : sa thèse ([Ladrière 1957, plus particulièrement le § X]), un article spécifiquement consacré à cette position : [Ladrière 1960], et dans une moindre mesure la synthèse de ces deux documents qui est proposée dans [Ladrière 1967a]. C'est donc essentiellement sur la base de ces trois documents que nous allons présenter puis discuter ce que nous appelons : « la position » de Jean Ladrière sur les « faits de limitation » (pour reprendre sa propre terminologie). Pour cela, nous allons commencer par en présenter le cadre, à la fois limitatif (du point de vue mathématique) et particulier (sur le plan philosophique). Au préalable, il convient sans doute de préciser que si l'auteur a certainement marqué la philosophie de la logique d'expression française dans les années 1960⁴²⁸, il a été pratiquement ignoré de la philosophie de la logique d'expression anglaise⁴²⁹ au sein de laquelle a été publiée l'immense majorité des travaux consacrés à ce sujet depuis cette époque.

Du point de vue mathématique, il apparaît clairement une double réduction de ce sur quoi porte en fait l'analyse de Jean Ladrière. Tout d'abord, et il s'agit de la première réduction, les limitations caractéristiques des théorèmes « négatifs » sont associées aux systèmes formels en tant que tels⁴³⁰, du seul fait que ceux-ci présentent les caractéristiques classiques d'effectivité [Ladrière 1957, p. 439]. Pourtant, l'auteur a par ailleurs particulièrement bien mis en évidence les conditions techniques générales nécessaires à l'obtention de ces théorèmes. Au-delà des caractéristiques d'effectivité demandées au système formel, celles-ci consistent principalement en la représentation au sein du système formel de la suite *illimitée* des *entiers naturels*, permettant l'application cruciale de l'argument diagonal [Ladrière 1957, p. 418] et [Ladrière 1960, p. 302, 315]. L'omission dans le cas présent de cette double restriction, le conduit alors à négliger les contre-exemples importants pour notre discussion que sont les systèmes formels strictement finitistes (bornés) ainsi que le théorème de Tarski de 1940-1967, 1948⁴³¹ qui n'inclut pas les entiers naturels pris dans toute leur généralité (voir le § 5.1). Il s'est ensuite limité à la possibilité de disposer (de reflets) d'expressions autoréférentielles au sein du système

⁴²⁷ Nous avons prévu de poser la question à Jean Ladrière lui-même, malheureusement, son décès survenu en 2007 n'a pas permis de donner suite à cette idée.

⁴²⁸ Comme cela est illustré par exemple par le commentaire que lui consacre Roger Martin [Martin 1964, p. 191].

⁴²⁹ La seule exception que nous avons rencontrée est une brève mention en 1958 [Fraenkel, Bar-Hillel 1958, p. 299].

⁴³⁰ On retrouve tout de même cette position, mais discrètement, dans les publications des années 1990 où il est fait état « d'une certaine limitation de la méthode de formalisation » [Ladrière 1989, p. 490 ; souligné par nous] et « de la méthode de formalisation » [Ladrière 1990, p. 1484 ; souligné par nous].

⁴³¹ Jean Ladrière mentionne ce dernier résultat en passant [Ladrière 1957, p. 274], mais sans en tenir compte dans son analyse philosophique.

formel, ce qui constitue la seconde réduction [Ladrière 1967a, p. 330], comme si tous les énoncés arithmétiques indécidables étaient nécessairement de cette forme, en négligeant en particulier la forme de l'énoncé non dérivable caractéristique du second théorème de Gödel de 1931⁴³² (ainsi d'ailleurs que l'énoncé indécidable du théorème de Rosser de 1936).

Sur le plan philosophique, il convient d'abord de souligner en matière d'ontologie l'engagement réaliste de Jean Ladrière, qui n'est jamais discuté par lui, mais qui se traduit par la référence répétée à « l'être mathématique » [Ladrière 1955, p. 538, 571] et [Ladrière 1957, p. 405], sans que le contenu n'en soit explicité. Du point de vue épistémologique, la distinction méthodologique classique, importante dans le cadre de la présente discussion, entre les contextes de la découverte et de la vérification n'est pas reconnue, tout en étant pourtant suggérée [Ladrière 1951, p. 381] et [Ladrière 1955, p. 552]. Son analyse porte donc sur une notion unique de « pensée mathématique », rattachée à la raison [Ladrière 1960, p. 319] et recouvrant indistinctement ces deux aspects sous la forme d'une « dualité de l'intuitif et du formel » [Ladrière 1957, p. 410-411]⁴³³. Pour cette pensée mathématique, il a repris à son compte la classification proposée dans [Myhill 1952b] que nous venons d'évoquer (sans clarifier le contenu du niveau « prospectif »), en lui adjoignant un quatrième niveau, qu'il nomme « le transcendantal » (en faisant alors référence à Hao Wang [Wang 1954, p. 263]) [Ladrière 1960, p. 304]. Il associe ce quatrième niveau aux hiérarchies transfinies en lui conférant un caractère absolu [Ladrière 1960, p. 307].

Dans un tel cadre limitatif, aussi bien mathématique que philosophique, Jean Ladrière est alors conduit à une interprétation des « faits de limitation » en trois étapes : 1) le rejet de ce qu'il nomme le « projet de formalisation intégrale », 2) l'introduction d'un lien entre les « faits de limitation » et la raison, 3) la proposition d'une origine de ce lien dans la temporalité de la conscience. Voici d'abord comment il définit le « projet de formalisation intégrale » (1^{ère} étape) :

Ce que vise le projet de formalisation intégrale, c'est donc la constitution de l'objet mathématique total. [...] On se trouve donc ici devant l'idée-limite d'un système parfaitement clos, ne renvoyant à rien d'autre qu'à lui-même. [Ladrière 1957, p. 409-410]

⁴³² Le traitement, et a fortiori l'interprétation du *second* théorème de Gödel de 1931, est en effet quasiment absent de [Ladrière 1957].

⁴³³ Tout en évoquant par ailleurs la liberté (pour le mathématicien) du choix des règles (du système formel) que l'on peut naturellement associer à la seule intuition.

On comprend qu'il s'agit là d'une position formaliste extrême, où un système formel est supposé pouvoir rendre compte de l'ensemble des mathématiques, sur la base d'une justification purement interne en éliminant en particulier tout recours à l'intuition. C'est à une telle position qu'il oppose alors les « faits de limitation » :

Les faits de limitation nous indiquent avec précision pourquoi le système total ne peut être qu'une idée-limite. [...] Le système total n'est réalisable ni comme représentation adéquate du champ intuitif, ni comme structure formelle capable de se réfléchir entièrement en elle-même, ni comme ensemble de procédures canoniques susceptible de fournir une solution effective à tout problème mathématique. [Ladrière 1957, p. 410-411]

Il n'existe pas [du fait des limitations] de système fermé qui serait comme le paradigme de tout discours, la cour suprême de la raison. [Ladrière 1957, p. 412 ; le commentaire entre crochets est nôtre]

Le caractère extrême (presque caricatural) de cette position ne permet certainement pas de la rapprocher du programme de Hilbert, auquel sa description renvoie inévitablement⁴³⁴. En effet, comme nous l'avons vu, dans le cas du programme de Hilbert au sens restreint (lié à la problématique des fondements) l'intuition (finitiste) joue au contraire un rôle crucial. Quant au programme de Hilbert au sens large (lié à la problématique du refus de *l'ignorabimus*), comme nous l'avons également vu, il n'est pas possible d'utiliser simplement et directement les théorèmes « négatifs » (en l'occurrence le premier théorème de Gödel de 1931 appliqué à PA) pour s'y opposer. En sens inverse, il n'est certainement pas nécessaire de faire appel à des résultats aussi sophistiqués que les théorèmes de Gödel de 1931 pour s'opposer à une telle vision extrême (et abstraite). L'affirmation d'un lien entre les « faits de limitation » et la raison (2^{ème} étape), est certainement l'engagement de Jean Ladrière le plus fort sur notre sujet. Elle est brève mais explicite et répétée :

Mais la nature même de la raison est telle que son objet ne cesse de la déborder : [...]. Ces deux limitations [de l'intelligence mathématicienne] sont solidaires et elles sont indiquées simultanément par les limitations que contient, à son niveau propre, la théorie des systèmes formels. [Ladrière 1957, p. 9 ; le commentaire entre crochets est nôtre]

Les faits de limitation en particulier apparaissent sur le trajet d'un effort d'élucidation radical que la pensée mathématique a entrepris à partir de son propre acquis. [Ladrière 1957, p. 401]

⁴³⁴ Programme de Hilbert par ailleurs explicitement analysé par Jean Ladrière [Ladrière 1957, p. 6, 9-10].

Les propriétés des formalismes nous manifestent quelque chose du fonctionnement de la raison mathématique, et même de la raison en général. [Ladrière 1960, p. 302]

Pour la justifier, Jean Ladrière donne essentiellement deux arguments. Il indique tout d'abord qu'il n'est pas possible d'inclure l'intuition dans la démarche calculatoire, c'est le premier argument [Ladrière 1957, p. 395, 413]. Sans contester une affirmation aussi générale, nous faisons remarquer que sa portée dans ce débat est liée à la conception, qui est celle de l'auteur, de la raison comme un tout, englobant indistinctement la part intuitive et la part calculatoire de la pratique mathématique. En sens inverse, elle perd toute pertinence dès que l'on introduit la distinction en mathématiques entre les contextes, de la découverte (où intervient l'intuition⁴³⁵), et de la vérification (objet de la démarche axiomatique où l'intuition est rejetée et seule concernée par les théorèmes « négatifs »). Il considère ensuite que la possibilité de construire des expressions autoréférentielles est une caractéristique de la pensée, c'est le second argument [Ladrière 1967a, p. 331]. Là encore, la portée de cet argument est liée à la réduction par l'auteur des énoncés indécidables en arithmétique à des (codages d') expressions autoréférentielles. Or cette réduction est manifestement abusive et elle est techniquement doublement invalidée par l'existence : 1) d'énoncés arithmétiques indécidables qui n'ont pas une forme autoréférentielle (à l'instar de l'énoncé non dérivable caractéristique du second théorème de Gödel de 1931), 2) d'énoncés arithmétiques, codages d'expressions autoréférentielles et pourtant dérivables dans un système formel fort (à l'instar de l'énoncé de Henkin, dérivable dans PA, ainsi que le montre le théorème de Löb de 1955). La proposition précise qui s'ensuit, selon laquelle un tel lien trouve son origine dans la temporalité de la conscience (3^{ème} étape), est explicite à la fois dans les trois documents cités [Ladrière 1957, p. 437-438], [Ladrière 1960, p. 312-313, 316, 320] et [Ladrière 1967a, p. 331]. C'est toutefois dans ce dernier document que l'on en trouve sans doute la formulation la plus claire :

Nos opérations [mentales] s'effectuent dans le temps ; elles se déroulent nécessairement dans une succession, mais en même temps elles s'inscrivent dans un horizon infini. Toute opération effective est ancrée dans l'actualité du présent, mais en même temps elle renvoie à une infinité d'opérations possibles. [Ladrière 1967a, p. 331 ; le commentaire entre crochets est nôtre]

Cette explication de Jean Ladrière implique alors l'établissement d'une correspondance entre, d'une part la caractéristique technique des systèmes formels concernés de pouvoir représenter

⁴³⁵ Où l'on peut douter que la démarche axiomatique éclaire vraiment les processus à l'œuvre, comme le suggère Jean Ladrière [Ladrière 1957, p. 443-444].

la suite potentiellement infinie des entiers naturels (condition de possibilité des théorèmes « négatifs ») et d'autre part la succession *temporelle* des pensées humaines qui caractérisent la structure de la conscience. Une telle correspondance est cependant doublement contestable. Il est tout d'abord discutabile de considérer que les pensées humaines puissent être assimilées à un ensemble potentiellement *infini*, ce qui est la première contestation. Eu égard au caractère fini de tout ce qui peut leur être associé *pratiquement* (le nombre de mots et la longueur des phrases possibles en langage naturel, étendu si l'on veut d'une manière plus générale aux perceptions), la référence à l'infini ne peut être qu'une métaphore⁴³⁶. En d'autres termes, une telle interprétation s'applique seulement, au sens strict, à un système formel *borné* dont la borne est un nombre suffisamment grand (ce qui nous ramène à l'ultrafinitisme évoqué au § 4.5), sachant que les théorèmes « négatifs » ne s'appliquent *pas* à un tel système formel. En effet, la possibilité de rendre compte de l'infinité de la suite des entiers (bien illustrée par l'infinité de la longueur du ruban de la machine de Turing) est indispensable à l'obtention de théorèmes « négatifs », faute de quoi l'application de l'argument diagonal, et partant la démonstration de Gödel (ou celle de Turing), est impossible. Il est surtout arbitraire de considérer que le flux ininterrompu de la pensée consciente est *discret* (seul moyen d'établir un accord avec la suite des entiers naturels), c'est-à-dire composé d'éléments successifs bien distincts, ce qui est la seconde contestation. Or ce caractère discret (associé au caractère d'infinité potentielle) est indispensable à la possibilité de théorèmes « négatifs », comme le montre le contre-exemple du théorème « positif » de Tarski de 1940-1967, 1948.

Il apparaît ainsi que la position forte défendue par Jean Ladrière dans les années 1960 en faveur d'un lien entre les « théorèmes de limitation des formalismes » et la raison humaine⁴³⁷, est en fait le reflet d'un choix arbitraire qu'il a effectué. Ce choix consiste d'abord en une réduction du cadre d'analyse mathématique retenu, d'une part en ne prenant en compte que la forme (autoréférentielle) de l'énoncé indécidable du premier théorème de Gödel de 1931 et en supposant que tous les énoncés de cette forme sont indécidables, d'autre part en négligeant les caractéristiques spécifiques que doivent présenter les systèmes formels (au-delà des critères d'effectivité) pour donner lieu aux théorèmes « négatifs » en arithmétique. Ce choix s'exprime ensuite par une orientation philosophique particulière en faveur d'une forme forte de réalisme mathématique, et surtout par la prise en considération de la pensée mathématique comme un

⁴³⁶ Sauf à supposer qu'il est fait référence à la raison d'un être humain *idéalisé* (à l'instar de ce qu'il est nécessaire d'introduire pour définir les machines de Turing), mais en atténuant alors grandement la force du lien affiché par l'auteur entre la raison et les « faits de limitation ».

⁴³⁷ Sachant que l'on trouve aussi, à cette période, la formulation neutre de « limite au pouvoir de représentation d'un formalisme » [Ladrière 1957, p. VIII].

tout indissociable, sans reconnaître que seule la part mise en œuvre dans le contexte de la vérification est concernée par ces théorèmes. De fait, il s'avère qu'une telle position n'a pas survécu à l'évolution que l'on constate depuis un demi-siècle, aussi bien des résultats mathématiques obtenus que des idées en philosophie des mathématiques. Il est d'ailleurs possible que l'auteur soit lui-même revenu sur sa position des années 1960, qu'il n'a en tout cas plus défendue dans ses derniers écrits sur le même sujet, dans les années 1990⁴³⁸ [Ladrière 1989, p. 490] et [Ladrière 1990, p. 1484].

Si nous faisons maintenant le bilan de notre analyse, nous soulignons d'abord que les remarques de [Wittgenstein 1956] sur les théorèmes de Gödel de 1931 relèvent manifestement d'une incompréhension des théorèmes de Gödel de 1931 (et plus généralement de l'ensemble des travaux effectués en logique depuis les années 1930), malgré les tentatives de défense héroïques de Stuart Shanker et Juliet Flyod qui s'apparentent à un combat d'arrière-garde. Il ne reste par ailleurs aujourd'hui à peu près rien de la force des trois grandes positions philosophiques sur la portée des théorèmes de « limitation des formalismes » qui ont été affichées en leur temps, respectivement : 1) par Emil Post dans les années 1940 en faveur d'une limitation du « pouvoir mathématique humain », 2) par John Myhill dans les années 1950 en faveur d'une limitation de la connaissance, 3) par Jean Ladrière dans les années 1960 en faveur d'une limitation de la raison. En dehors d'objections techniques précises qui s'y opposent, celles-ci apparaissent liées à un engagement philosophique de leur auteur et ne peuvent donc être considérées comme le reflet des théorèmes « de limitation des formalismes ».

⁴³⁸ A une exception près, à la fin de sa vie, pour laquelle nous n'avons pas d'explication : [Ladrière 2002].

Conclusion

Nous allons maintenant essayer de répondre le plus directement possible à la question fondamentale à laquelle notre travail a été consacré et qui concerne la nature des limitations introduites par les théorèmes qui ont été dits « de limitation des formalismes », dont on rappelle qu'ils désignent principalement les deux théorèmes de Gödel de 1931, le théorème de Kleene de 1936 et celui de Turing de 1936-1937. Notre fil conducteur pour cela a été le programme de Hilbert au sens large, c'est-à-dire d'une part la réponse que Hilbert a apporté au problème des fondements (ce que nous avons appelé le programme au sens restreint) et d'autre part la justification qu'il espérait obtenir de l'absence de problèmes mathématiques insolubles (l'optimisme rationaliste), par rapport auquel et contre lequel ces théorèmes se sont imposés comme un élément majeur. Cette analyse a été étendue au débat relatif à l'impact des théorèmes de Gödel de 1931 dans le contexte de la question générale de savoir si la pensée à l'œuvre dans l'exercice des mathématiques peut être considérée comme « mécanique ».

L'objectif général de Hilbert avec le programme au sens restreint était d'établir définitivement la sûreté de l'ensemble des mathématiques sans en exclure aucune partie. La démarche qu'il a proposée pour cela consiste en la fourniture d'une preuve *finitiste* de non-contradiction d'un système formel formalisant l'ensemble des énoncés « réels » et des énoncés « idéaux », sur la base de l'évidence intuitive des énoncés « réels » vrais. Nous en avons retenu une interprétation technique précise qui repose essentiellement sur la « thèse de Tait » (qui assimile les fonctions « réelles » aux FRP et les énoncés « réels » aux énoncés Δ_0 et Π_1 de la hiérarchie de Kleene), que nous avons libéralisée (sans imposer que toutes les preuves finitistes soient des théorèmes de PRA). Le programme se ramène alors à la requête de montrer dans PRA le caractère conservatif de Z_2 (a fortiori de PA) par rapport à PRA, pour les énoncés Π_1 (compris comme des « jugements hypothétiques »). Sur cette base, nous retrouvons le résultat classique d'une *opposition* de chacun des deux théorèmes de Gödel de 1931 au programme au sens restreint, en particulier contre la position défendue par Michael Detlefsen, mais à la suite d'une analyse beaucoup plus précise que celle qui est habituellement proposée. Nous montrons en effet que dans le cadre de notre interprétation du programme de Hilbert au sens restreint, qui *n'est pas* celle de Michael Detlefsen, les objections soulevées par celui-ci ne sont pas recevables.

La première voie envisagée dans le but de sauver le programme malgré le constat précédent consiste à étendre le finitisme originel. Nous avons introduit pour cela deux exigences

minimales qui permettent d'en conserver les objectifs : 1) la base intuitive doit en être acceptée par tout mathématicien quelle que soient ses inclinations philosophiques, 2) l'extension de la partie « idéale » à partir de cette base doit être progressive, ce qui impose une démarche effective et non imprédicative (au sens de Poincaré). Nous montrons alors qu'aucune des solutions « traditionnelles » ([Gödel 1933a], [Gentzen 1936], [Gödel 1958]) ne répond à de telles exigences. Le résultat de Gentzen étendu par [Jerolow 1973] conduit cependant au résultat intéressant, qu'une démonstration de consistance de PRA est possible sur la base de l'extension du finitisme la plus extrême qu'il soit raisonnablement possible d'envisager (où l'ordinal limite est ω^{ω}). L'introduction par Hilbert de la règle- ω , successivement en 1930 et 1931, est encore moins satisfaisante, en perdant tout l'intérêt épistémologique de simplicité et d'évidence de la base finitiste qui constituaient l'attrait du finitisme originel et en s'interdisant d'atteindre les objectifs du programme. Dans le cadre d'une conception instrumentaliste particulière, Michael Detlefsen est quant à lui bien parvenu à proposer une interprétation cohérente du programme de Hilbert au sens restreint qui permette de s'affranchir des deux théorèmes de Gödel de 1931. Il ne reste cependant alors quasiment plus rien des traits fondamentaux du programme originel de Hilbert et les objectifs de Hilbert ne peuvent donc être atteints non plus. Il apparaît ainsi qu'aucune des différentes extensions du finitisme de Hilbert qui ont été proposées ne peut prétendre sauver le programme au sens restreint⁴³⁹.

La seconde voie susceptible de sauver le programme consiste à réduire le domaine des mathématiques concernées. Les résultats obtenus dans les années 1970 et 1980 dans le cadre des « reverse mathematics », établissent qu'une grande partie des mathématiques existantes peut être formalisée dans des systèmes formels qui sont montrés conservatifs par rapport à PRA (au moins) pour les énoncés Π_1 . Il s'agit donc bien *de fait* d'une *réalisation partielle* mais significative, du programme de Hilbert au sens restreint selon notre interprétation. Le caractère partiel de cette réalisation réside dans la limite de l'étendue du domaine des mathématiques concernées. Des solutions ont été proposées pour repousser cette limite, mais qui se rapportent alors à des systèmes formels d'autant plus forts que PRA (à commencer par PA), que les extensions obtenues sont importantes. Or pour ces systèmes formels, il n'est plus possible d'en établir la consistance sur une base finitiste même raisonnablement étendue. Dans une autre direction, Daniel Isaacson a proposé la solution radicale qui consiste à exclure *a priori* des mathématiques (en l'occurrence de l'arithmétique) les théorèmes de Gödel de 1931 sur la base

⁴³⁹ La solution radicale de la logique paraconsistante, qui consiste à considérer que PA et Q sont simplement inconsistants (mais absolument consistants), ne peut pas être sérieusement retenue à la suite de l'analyse dirimante qui en a été faite par Stewart Shapiro dans [Shapiro 2002].

du fait que ceux-ci font intervenir en fait des énoncés *métamathématiques*. Une objection naturelle est alors l'irruption également dans les années 1970 de plusieurs résultats montrant que certains énoncés combinatoires *finis* sont des théorèmes de Z_2 , indécidables dans PA. La réponse est que ces énoncés s'avèrent être encore des énoncés métamathématiques, qui sont donc également exclus par la thèse d'Isaacson, même s'ils se présentent sous une forme mathématique plus naturelle que l'énoncé G de Gödel, ou que le codage brut de l'affirmation de la consistance dans un système formel naturel de l'arithmétique. Une conjecture, étendant celle de Friedman, peut alors être proposée : « Tout énoncé exprimable dans PA, qui n'est pas le codage d'une expression métamathématique concernant un système formel naturel, est décidable dans PA ». Il va de soi que la thèse d'Isaacson et la conjecture qui en a été induite, pour intéressantes qu'elles soient, ne peuvent cependant prétendre valider le programme de Hilbert au sens restreint, dans la mesure où la question de la démonstration finitiste en un sens raisonnable de la consistance de PA, amène une réponse négative⁴⁴⁰.

Sur le programme de Hilbert au sens large, nous aboutissons à un résultat beaucoup plus nuancé que celui qui est généralement proposé en réponse à la question : que nous disent les résultats « négatifs » obtenus en logique et formelle et en informatique théorique, sur l'existence de problèmes mathématiques insolubles ? Pour pouvoir répondre à cette question à la suite d'une *démonstration*, comme le souhaitait Hilbert, il est d'abord nécessaire de donner à celle-ci un contenu technique précis, incluant en particulier le critère *d'intersubjectivité* de la résolution des problèmes, ce qui conduit alors à deux interrogations de force différente : à minima sur la décidabilité de tout problème mathématique (optimisme rationaliste₁) et à maxima sur la complétude syntaxique d'un système formel, supposé unique, en mesure de rendre compte de l'ensemble des mathématiques, ou au moins de l'arithmétique du 1^{er} ordre (optimisme rationaliste₂). La réponse classique à l'optimisme rationaliste₂ est que l'incomplétude syntaxique de PA le réfute. Cette réponse ne prend cependant pas en compte le point technique que les indécidables arithmétiques qui sont cités à son appui sont en fait toujours des codages d'énoncés *métamathématiques* (et non des énoncés authentiquement arithmétiques), qui de ce fait ne peuvent prétendre être opposés à l'optimisme rationaliste₂. Pour aller plus loin dans le sens d'une démonstration de l'optimisme rationaliste₂ on peut d'abord envisager de faire appel à la thèse d'Isaacson, qui conduit à l'affirmation de la complétude syntaxique de PA pour les énoncés authentiquement arithmétiques. Cette stratégie échoue cependant, dans la mesure où l'appel à une thèse philosophique n'est pas compatible avec le

⁴⁴⁰ La solution encore plus radicale de l'ultrafinitisme, envisagée également dans les années 1960 et 1970 n'y réussit pas d'avantage alors qu'elle est par ailleurs assaillie de difficultés spécifiques.

critère de neutralité philosophique qui constitue l'une des conditions du critère d'intersubjectivité. Une autre solution possible, consiste au recours à des solutions particulièrement non effectives (en particulier de la part de Gödel), mais qui de ce fait impliquent également un engagement philosophique, et par là échouent à répondre aux exigences de Hilbert exactement pour la même raison. L'espoir de Hilbert d'une *démonstration* de l'optimisme rationaliste s'avère donc vain. Quant à la réponse la plus connue (et tranchée) *contre* l'optimisme rationaliste₁ : celle de Grégory Chaitin, elle est clairement invalidée par l'analyse dévastatrice de Panu Raatikainen. Il ne s'avère donc possible ni de pouvoir démontrer ni de pouvoir réfuter l'optimisme rationaliste de Hilbert, qui reste donc à l'état d'une thèse métamathématique.

Une solution radicale en vue du dépassement des résultats précédents consiste à envisager un élargissement de la thèse de Church-Turing. A la suite d'une analyse détaillée des conditions d'application de celle-ci, nous avons été amené à faire état de deux caractéristiques qui ont été soit oubliées soit ignorées jusqu'à une date récente. Tout d'abord la thèse fait intervenir les seules FRG (ou les seules fonctions Turing-calculables) qui peuvent être établies *constructivement*. Ensuite, comme cela a été souligné par Wilfried Sieg, une machine de Turing caractérise en fait *un être humain qui calcule*. Il en résulte deux conséquences philosophiques. 1) Il apparaît une certaine relativité, d'une part de la récursivité qui y intervient : « pour les humains » et non « en soi », d'autre part du concept de calculabilité, qui est bien indépendant de tout système formel particulier, mais avec la restriction forte que celui-ci respecte les conditions d'effectivité modernes habituelles. 2) Nous avons été amenés à proposer une signification épistémologique, à la fois précise et de portée générale de la machine de Turing (en retrouvant alors une proposition similaire de Sieg en 1994) : celle-ci exprime en fait l'exigence *d'intersubjectivité d'une démonstration formelle en mathématiques*, qui doit être publiquement communicable et dont la vérification ne doit faire appel à aucun engagement invérifiable en principe par n'importe quel être humain. Il en résulte une réfutation de tous les arguments qui ont été proposés en faveur d'un dépassement de la machine de Turing et donc a fortiori de la thèse de Church-Turing, sur la base du critère d'intersubjectivité introduit.

A la suggestion générale de Hilbert selon laquelle la pensée humaine à l'œuvre en mathématiques peut être réduite à une activité mécanique, nous avons substitué la question plus modeste de savoir ce qu'il en est pour la seule part de cette pensée mathématique en jeu dans le contexte de la vérification (en éliminant celui de la découverte). Gödel lui-même, John Lucas et Roger Penrose ont défendu en commun la position forte de l'existence de procédures humaines non mécanisables pour la vérification de théorèmes en arithmétique, révélées par les

théorèmes de Gödel de 1931. Gödel justifie toutefois sa position par un engagement ontologique en faveur d'un réalisme au sens fort, tout en la dissociant de ses théorèmes de 1931. En sens inverse, John Lucas et Roger Penrose affirment déduire la même position d'une application du premier théorème de Gödel tout en ne reconnaissant aucun engagement philosophique associé, ce qui est évidemment beaucoup plus ambitieux. Nous montrons cependant qu'une telle gageure n'est pas tenue. L'argument de John Lucas repose bien en fait sur un engagement philosophique sous-jacent, et la technicité introduite (en l'occurrence l'appel au premier théorème de Gödel) n'est qu'un déguisement. Son raisonnement s'avère circulaire, et sa position le reflet de prémisses anti-mécanistes cachées dans l'affirmation de l'évidence de la correction des théorèmes d'un système formel de l'arithmétique par rapport au modèle standard. La même critique s'applique immédiatement au premier argument de Penrose, l'engagement philosophique étant alors de plus étendu à un réalisme en un sens fort. Elle s'applique enfin également, mais de manière moins directe, aux différentes reconstitutions techniques que l'on peut raisonnablement imaginer du second argument de Penrose. Nous rejoignons ici une position très consensuelle des philosophes de la logique contre la position de John Lucas et Roger Penrose, mais sur la base d'une argumentation beaucoup plus précise que celle qui a été généralement proposée. On retrouve alors un résultat déjà évoqué à propos des machines de Turing : les caractéristiques spécifiquement humaines dont une machine de Turing ne peut pas rendre compte ne sont pas publiquement communicables. C'est sans doute la seule forme (faible) d'anti-mécanisme que l'on puisse raisonnablement défendre.

En réponse directe à notre interrogation fondamentale, seuls trois auteurs se sont engagés dans le sens de limitations *humaines* profondes révélées par les théorèmes « de limitation des formalismes » : 1) Emil Post dans les années 1940, en faveur d'une limitation du « pouvoir mathématique humain », 2) John Myhill dans les années 1950, en faveur d'une limitation de la connaissance, 3) Jean Ladrière dans les années 1960, en faveur d'une limitation de la raison⁴⁴¹. En 1941, Emil Post affirme que la pensée en mathématiques résulte de la combinaison d'une génération permanente de créativité « à la Bergson » continuellement présente, saisie et « cristallisée » sous la forme d'une représentation spatiale de symboles « à la Hilbert » à la suite d'une réflexion consciente de l'être humain sur le processus créatif lui-même. Il considère que c'est alors l'impossibilité de saisir entièrement le premier mouvement par le second, qui est à

⁴⁴¹ Au-delà du cas particulier des remarques plus générales de [Wittgenstein 1956] sur les théorèmes de Gödel de 1931 qui relèvent manifestement d'une incompréhension de ces théorèmes (et plus généralement de l'ensemble des travaux effectués en logique depuis les années 1930), malgré les tentatives de défense de Stuart Shanker et Juliet Flyod qui s'apparentent à un combat d'arrière-garde.

l'origine des limitations rencontrées dans les théorèmes « négatifs » en arithmétique. Outre que l'origine d'une telle génération de créativité reste totalement mystérieuse, une explication aussi générale est cependant invalidée par l'existence de systèmes formels non triviaux conduisant à des théorèmes « positifs » en arithmétique, à l'instar du théorème de Tarski de 1940-1967, 1948. En 1952, John Myhill a défendu l'existence de limitations humaines essentielles révélées par les théorèmes de Gödel de 1931, de Church de 1936, et de Turing de 1936-1937, en s'appuyant sur l'affirmation d'une notion de vérité mathématique en un sens absolu, complétée en 1960 par celle d'une notion de prouvabilité en un sens absolu, avec comme seule justification dans les deux cas le recours à l'évidence, soit un acte de foi. La distinction méthodologique classique entre l'établissement et la vérification d'une démonstration mathématique, invalide par ailleurs le lien introduit par John Myhill entre créativité en mathématique et indécidabilité (ou incomplétude syntaxique) d'un système formel. La position forte défendue par Jean Ladrière dans les années 1960 d'un lien entre les « théorèmes de limitation des formalismes » et la raison humaine, s'avère quant à elle le reflet d'un choix arbitraire que cet auteur a effectué. Ce choix consiste d'abord en une réduction du cadre d'analyse mathématique retenu, d'une part en ne prenant en compte que la forme (autoréférentielle) de l'énoncé indécidable du premier théorème de Gödel de 1931 et en supposant que tous les énoncés de cette forme sont indécidables, d'autre part en négligeant les caractéristiques spécifiques que doivent présenter les systèmes formels (au-delà des critères d'effectivité) pour donner lieu aux théorèmes « négatifs » en arithmétique. Ce choix s'exprime ensuite par une orientation philosophique particulière en faveur d'une forme forte de réalisme mathématique, et surtout par la prise en considération de la pensée mathématique comme un tout indissociable, sans reconnaître que seule la part mise en œuvre dans le contexte de la vérification est concernée par ces théorèmes. Il apparaît ainsi globalement qu'en dehors d'objections techniques précises qui s'opposent aux trois positions fortes mentionnées, celles-ci sont liées à un engagement philosophique de leur auteur et ne peuvent donc être considérées comme le reflet des théorèmes « de limitation des formalismes ».

De fait, dans leur immense majorité, les philosophes de la logique et des mathématiques (y compris Jean Ladrière dans ses écrits des années 1990) se sont clairement prononcés dans le sens où les « limitations » révélées en particulier par le premier théorème de Gödel de 1931 sont interprétées par eux comme exprimant seulement celles des systèmes formels eux-mêmes, lorsque ceux-ci sont appliqués à l'arithmétique récursive avec une exigence d'effectivité. Selon cette position, ce n'est que par rapport à une idée préconçue sur l'ampleur de ce que devraient (ou auraient dû) permettre les systèmes formels quant à la formalisation de l'ensemble des mathématiques, qu'il est possible en l'occurrence de faire état de « limitations », ce qui est

illustré par la jolie formule attribuée à Quine par Roman Murawski : « De tels systèmes voulaient avaler un morceau d'ontologie plus gros que ce qu'ils sont capables de digérer » [Murawski 1999, p. 298]. Bernays d'abord et Morris Kline ensuite ont même suggéré que ces limitations sont en fait celles d'une idée préconçue de la part des mathématiciens de la fin du XIX^e siècle, sur l'aptitude à formaliser strictement et entièrement les mathématiques [Introduction de la discussion qui a suivi Ladrière 1960 (1986), p. 326] et [Kline 1980 (1989), p. 480]. Le parallèle avec l'avènement des nombres irrationnels chez les pythagoriciens, par rapport à une idée préconçue sur les possibilités offertes par les seules manipulations des entiers naturels, vient alors naturellement à l'esprit (il est suggéré par Judson Webb et Hourya Sinaceur [Webb 1980, p. 201] et [Sinaceur 1990d, p. 553]). L'ensemble de l'analyse détaillée que nous avons effectuée permet cependant d'aller un peu plus loin, en faisant ressortir les éléments qui contribuent à une certaine *relativité* du recours à la formalisation, mise en évidence par les différents résultats métamathématiques que nous avons été amenés à évoquer.

La forme de relativité la plus évidente associée au recours à la formalisation est sans doute celle du choix du système formel, même en se limitant à l'arithmétique comme nous l'avons fait. Les travaux des « reverse mathematics » ont montré l'ampleur d'un tel choix, depuis \mathbb{Q} jusqu'à \mathbb{Z}_2 , en passant par EA, RCA₀, WKL₀, PRA, PA, ACA₀ et ATR₀ pour ne citer que les principaux. Cette relativité, qui ne va tout de même pas jusqu'à produire des résultats contradictoires, se traduit alors de différentes manières. Elle se révèle d'abord de la manière la plus directe au travers des résultats précis obtenus sur les énoncés combinatoires finis, qui sont d'une part des théorèmes de \mathbb{Z}_2 et d'autre part des indécidables d'un système formel plus faible (variable selon l'énoncé en question : voir le § 4.3). Il est cependant clair que cela ne concerne que des énoncés très particuliers, dont nous avons de plus souligné la caractéristique essentielle que chacun d'entre eux est équivalent à un énoncé *métamathématique* de 1-consistance d'un certain système formel (variant selon l'énoncé en question). Cette relativité se reflète plus largement dans l'opposition entre les tenants du « programme de Gödel » (selon l'expression de Feferman), qui veulent se laisser toute latitude d'utiliser les systèmes formels les plus forts (ensemblistes compris) pour démontrer des énoncés arithmétiques et les tenants du « programme de Hilbert » (interprété alors en un sens très large), qui cherchent au contraire à contenir au juste nécessaire le système formel à utiliser en suivant les résultats obtenus par les « reverse mathematics ». La conséquence de cette alternative se limite cependant à une prise de risque plus importante dans le premier cas que dans le second d'utiliser un système formel (associé à une ontologie plus riche), qui s'avèrerait ultérieurement contradictoire. Dans le même esprit, l'élément de relativité le plus évident concernant le choix du système formel consiste

dans le fait de retenir une logique particulière, par exemple la logique classique plutôt que la logique intuitionniste et a fortiori une logique non classique plus exotique, qui ne conduira alors pas à la même ampleur de théorèmes accessibles. Il est clair qu'il s'agit ici d'un choix d'ordre pratique ou philosophique qui reflète la liberté laissée au mathématicien quant aux contraintes que celui-ci s'impose pour l'établissement d'une démonstration arithmétique. L'élément important est que dans tous les cas, le choix effectué n'entraîne aucune subjectivité, dans la mesure où ce choix étant explicité n'importe quel mathématicien peut vérifier la validité de toute preuve formelle donnée pour le système formel retenu, comme l'avait déjà souligné Roger Martin il y a plus de cinquante ans [Martin 1964, p. 4-5].

Une deuxième forme de relativité évidente associée au recours à la formalisation est bien entendu celle du modèle retenu pour un système formel donné. L'énoncé G du premier théorème de Gödel de 1931 en fournit une illustration simple, en étant vrai dans le modèle naturel et faux dans le modèle non standard d'un système formel de l'arithmétique⁴⁴², ce qui a conduit à une analogie avec la situation du cinquième postulat d'Euclide en géométrie [Martin 1964, p. 165], [Kleene 1967 (1971), p. 333]. Le point important, relevé par Jacques Dubucs dans [Dubucs 1992] est alors qu'un mathématicien ne peut fournir de garantie publique du choix du *modèle* qu'il a effectué (pour un système formel retenu), même si les mathématiciens ont de bonnes raisons *empiriques* pour s'entendre sur ce choix.

Un élément de relativité plus profond associé au recours à la formalisation se situe au niveau des critères qui sont retenus pour définir ce qui est *intuitivement* accepté a priori dans le cadre d'une démarche « fondationnaliste »⁴⁴³. En effet comme Hilbert lui-même a été contraint de l'admettre en définissant son programme (au sens restreint) dans les années 1920 (ce qui n'était pas le cas en 1904), même des exigences finitistes très sévères nécessitent un point de départ qui devra être intuitivement admis, qui correspond en l'occurrence au contenu du finitisme retenu. La longue discussion que nous avons eue à ce sujet illustre cependant la difficulté de s'accorder sur la définition précise d'un tel contenu. Au-delà de la solution que nous avons retenue pour la circonscription du finitisme de Hilbert (et de la discussion que nous avons eue sur les autres solutions qui ont été proposées en alternative), nous avons été amenés à proposer

⁴⁴² Nous avons également été amenés à évoquer une situation sans doute extrême (le cas d'une théorie T, consistante sans être 1-consistante, constituée de $PA + \neg \text{Con}(PA)$, ce dernier énoncé étant indécidable dans PA par le 2^{ème} théorème de Gödel de 1931) qui illustre la conséquence potentielle d'un tel choix : dans une telle théorie « pathologique » T, l'énoncé $\neg \text{Con}(T)$ est dérivable dans T, en étant vrai dans les seuls modèles non-standard de la théorie, tout en étant faux de par son sens.

⁴⁴³ Ce qui renvoie à la boutade de Michael Resnik : « [...] la consistance est suffisante pour toute la vérité mathématique sauf pour la vérité des principes utilisés dans les preuves de consistance » [Resnik 1974 (1990), p. 116].

deux exigences caractérisant le finitisme de Hilbert auxquelles toute proposition d'extension raisonnable de ce finitisme doit répondre pour en conserver les objectifs. Si la seconde est technique, la première est simplement que la base intuitive puisse être acceptée par tout mathématicien, quelles que soient ses inclinations philosophiques ; elle est donc d'ordre empirique. Si l'on prend l'exemple de la solution instrumentaliste proposée par Michael Detlefsen pour s'affranchir de l'effet des deux théorèmes de Gödel de 1931, le caractère non naturel très largement reconnu des critères qui doivent être introduits pour parvenir à un tel résultat montre qu'il existe un consensus *social* sur ce que de tels critères doivent raisonnablement être. D'une manière encore plus directe dans le cadre du programme de Hilbert au sens restreint, l'impossibilité qui résulte du second théorème de Gödel de 1931 d'établir la consistance de Z_2 et même de PA sur une base finitiste, et d'une manière plus générale l'impossibilité de « gravir les échelons » c'est-à-dire l'impossibilité d'établir progressivement la consistance de systèmes de plus en plus forts en partant par exemple de celle de PRA, fait que les mathématiciens doivent encore s'appuyer sur une base empirique pour se convaincre de la consistance des systèmes formels auxquels ils se rattachent (cette base empirique étant liée en l'occurrence à leur *pratique* mathématique, comme l'a souligné Georg Kreisel).

Si l'on s'intéresse maintenant à « l'optimisme rationaliste » de Hilbert (qui traduit son refus de l'existence en principe de problèmes mathématiques insolubles) en prenant en compte cette fois le programme de celui-ci au sens large, on retrouve encore un élément de relativité associé au recours à la formalisation, au travers de la fixation de critères précis en définissant justement le contenu, ce qui nous a conduit à distinguer ce que nous avons appelé l'optimisme rationaliste₁ (décidabilité de tout problème mathématique) et l'optimisme rationaliste₂ (complétude syntaxique du système formel dans lequel peut être formalisée au moins l'arithmétique). Sur les quatre exigences que nous avons introduites à cet effet, la quatrième qui est certainement la plus contraignante est encore d'ordre empirique (et proche de celle retenue pour définir une base finitiste raisonnable) : il s'agit de l'exigence épistémologique générale d'intersubjectivité de toute résolution d'un problème mathématique (la solution doit être publiquement communicable en un temps fini et doit pouvoir être vérifiée en principe par n'importe quel mathématicien). Il s'avère alors qu'une machine de Turing (qui intègre de fait les critères d'effectivité habituels appliqués à un système formel, comme cela a été souligné par Wilfried Sieg) exprime naturellement cette exigence d'intersubjectivité d'une démonstration mathématique. L'origine de cet accord est de nouveau d'ordre empirique et provient du fait que les êtres humains n'ont pas trouvé d'autre (ou de meilleure) solution que celle d'un système formel effectif ou d'une machine de Turing (caractérisés dans les deux cas par le caractère

mécanique de leur utilisation), pour assurer l'exigence en question⁴⁴⁴. Il convient tout de même de signaler une autre forme de relativité un peu subtile liée à la thèse de Church-Turing qui tient au fait que dans sa définition même, les FRG totales retenues (ou leur équivalent sous la forme de fonctions Turing-calculables) sont limitées à celles *définies de manière constructive*, la circonscription associée ne pouvant être définie que de manière empirique ; sachant que là encore, il existe un très bon accord des mathématiciens sur le sujet.

Il apparaît finalement qu'au-delà de refléter les limitations intrinsèques des systèmes formels lorsque ceux-ci incluent l'arithmétique récursive avec une exigence d'effectivité, les théorèmes « de limitation des formalismes » mettent en évidence une certaine relativité associée au recours à la formalisation elle-même, y compris lorsque celle-ci est utilisée au-delà de l'arithmétique. Cette relativité se caractérise par un ancrage nécessaire mais parfaitement circonscrit de tout recours à la formalisation dans la pratique empirique des mathématiques informelles, ancrage qui s'étend au-delà du finitisme originel envisagé par Hilbert tout en conservant le critère fondamental qu'il doit être accepté par l'ensemble des mathématiciens. Cette dernière requête conduit naturellement à l'adoption des critères d'effectivité habituels des systèmes formels en association avec la thèse de Church-Turing prise au sens classique. Elle permet également de répondre à trois exigences dont la réponse ne peut pas être formelle : d'une part s'entendre sur le modèle choisi en relation avec le système formel retenu (pour un résultat mathématique donné), d'autre part être convaincu de la consistance du système formel utilisé, enfin s'accorder sur le contenu des FRG (« constructives ») qui sont prises en compte pour définir la thèse de Church-Turing. Une telle relativité peut sans doute être associée à ce que Jean Ladrière a appelé « la dialectique entre l'intuition et le formel », mais en conférant alors un contenu précis à une formulation aussi générale.

⁴⁴⁴ Ce que l'on retrouve dans l'argumentation que l'on peut opposer à l'affirmation de Lucas-Penrose, qui entend montrer la supériorité de l'être humain sur toute solution purement mécanique en matière de vérification d'une démonstration mathématique.

Bibliographie

Ackermann Wilhelm :

- 1937 (1950) → Hilbert David, etc.

Amaral Furtado do :

- 1993 → Costa Newton da, etc.

Anderson Alan :

- 1957 « Mathematics and the "Language game" », *The Review of Metaphysics*, vol. XI, n° 1, p. 446-458.

Andler Daniel :

- 1998 « Turing : pensée du calcul, calcul de la pensée » dans Nef Frédéric, Vernant Denis (ed.), *Le formalisme en question*, Paris, Librairie Vrin, p. 13-35.

Arbib Michael :

- 1964 *Brains, machines and mathematics*, 2^{ème} édition (1987), New-York, Springer-Verlag.

Arkoudas Constantine :

- 2006 → Bringsjord Selmer, etc.

Atten Mark van :

- 2014 « Gödel and intuitionism » dans Dubucs Jacques, Bourdeau Michel (ed.), *Constructivity and computability in historical and philosophical perspective*, Dordrecht, Springer, p. 169-214 ;
- 2015 *Essays on Gödel's reception of Leibniz, Husserl, and Brouwer*, Cham, Springer.

Atten Mark van, Kennedy Juliette :

- 2003 « On the philosophical development of Kurt Gödel », *The Bulletin of Symbolic Logic*, n° 9, vol. 4, p. 425-476.

Auerbach David :

- 1985 « Intensionality and the Gödel theorems », *Philosophical Studies*, vol. 48, p. 337-351 ;
- 1992 « How to say things with formalisms » dans Detlefsen Michael (ed.), *Proof, logic and formalization*, London, Routledge & Kegan Paul, p. 77-93.

Avigad Jeremy :

- 1996 « Formalizing forcing arguments in subsystems of second order arithmetic », *Annals of Pure and Applied Logic*, vol. 82, p. 165-191 ;
- 2003 « Number theory and elementary arithmetic », *Philosophia Mathematica*, n° 11, p. 257-284 ;
- 2010 « Gödel and the metamathematical tradition », dans Feferman Solomon, Parsons Charles, Simpson Stephen (ed.), *Kurt Gödel ; Essays for his centennial*, Cambridge, Cambridge University Press, p. 45-60.

Avigad Jeremy, Feferman Solomon :

- 1998 « Gödel's functional ("Dialectica") interpretation », dans Buss Samuel (ed.), *Handbook of proof theory*, Amsterdam, Elsevier, p. 337-405.

Avron Arnon :

- 2009 « Review of Smith 2007 », *The Bulletin of Symbolic Logic*, vol. 15, n° 2, p. 218-222.

Awodey Steve, Reck Erich :

- 2002a « Completeness and categoricity. Part I : nineteenth century axiomatics to twentieth century metalogic », *History and Philosophy of Logic*, n° 23, vol. 1, p. 1-30 ;
- 2002b « Completeness and categoricity. Part II : twentieth century metalogic to twenty-first century semantics », *History and Philosophy of Logic*, n° 23, vol. 2, p. 77-94.

Bacon John :

- 1999 → Detlefsen Michael, etc.

Bar-Hillel Yehoshua :

- 1958 → Fraenkel Abraham, etc.

Barrow John :

- 1998 *Impossibility : The Limits of Science and the Science of Limits*, Oxford, Oxford University Press.

Barwise Jon, Schlipf John :

- 1975 « On recursively saturated models of arithmetic », dans Saracino D., Weispfenning V. (ed.), *Model theory and algebra*, Berlin, Springer-Verlag, p. 42-55.

Beeson Michael :

- 2009 *Foreward to Kleene 1952*, New-York, Iski, p. i-xi.

Beklemishev Lev :

- 2007 « Review of Franzen 2005 », *The Bulletin of Symbolic Logic*, vol. 13, n° 2, p. 241-243.

Benacerraf Paul :

- 1965 « What numbers could not be », *Philosophical Review*, n° 125, vol. 1, p. 17-73 ;
- 1967 « God, the devil and Gödel », *The Monist*, n° 51, p. 9-32 ;
- 1973 (1996) « Mathematical truth », dans Hart William (ed.), *The philosophy of mathematics*, Oxford, Oxford University Press, p. 14-30.

Bendegem Jean-Paul van :

- 1993 « The strong Hilbert's program », *Philosophical Studies*, n° 28, p. 357-362.

Bennett Charles :

- 1986 « On the nature and origin of complexity in discrete, homogeneous, locally-interacting systems », *Foundations of Physics*, vol. 16, n° 6, p. 585-592 ;
- 1988 (1988) « Dissipation, information, computational complexity and the definition of organisation » dans Pines David (ed.), *Emerging syntheses in science*, Redwood City, Addison-Wesley, p. 215-233 ;
- 1993 (1995) « Logical depth and physical complexity » dans Herken Rolf (ed.), *The universal Turing machine : a half-century survey*, Oxford, Oxford University Press, p. 103-134.

Berg Jonathan, Chihara Charles :

- 1975 « Church's thesis misconstrued », *Revue Internationale de Philosophie*, n° 4, p. 343-353.

Bernays Paul :

- 1922a (1998) « Hilbert's signifiacnce for the philosophy of mathematics », trad. angl. Paolo Mancosu, dans Mancosu Paolo (ed.), *From Brouwer to Hilbert, the debate on the foundation of mathematics in the 1920's*, New-York, Oxford University Press, p. 189-197 ;
- 1922b (1998) « On Hilbert's thoughts concerning the grounding of arithmetic », trad. angl. Paolo Mancosu, dans Mancosu Paolo (ed.) *From Brouwer to Hilbert, the debate on the foundation of mathematics in the 1920's*, New-York, Oxford University Press, p. 215-222 ;
- 1927 (1992) « Annexe à la conférence de Hilbert sur les fondements des mathématiques », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Intuitionnisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 172-174 ;
- 1930/1976 (2003) « La philosophie des mathématiques et la théorie de la démonstration de Hilbert », trad. fr. Hourya Benis Sinaceur, dans *Philosophie des mathématiques*, Paris, Librairie Vrin, p. 41-81 ;
- 1934 → Hilbert David, etc. ;
- 1935a « Sur le platonisme dans les mathématiques », *L'enseignement mathématique*, vol. 24, n° 1-2, p. 52-69 ;
- 1935b « Quelques points essentiels de la métamathématique », *L'enseignement mathématique*, vol. 24, n° 1-2, p. 70-95 ;
- 1937/1976 (2003) « Thèses et remarques sur les questions philosophiques et sur la situation de la recherche en fondement de la logique et des mathématiques », trad. fr. Hourya Benis Sinaceur, dans *Philosophie des mathématiques*, Paris, Librairie Vrin, p. 99-104 ;

- 1939 → Hilbert David, etc. ;
- 1941 « Sur les questions méthodologiques actuelles de la théorie hilbertienne de la démonstration », dans Gonseth Ferdinand (ed.) *Les entretiens de Zurich sur les fondements et la méthode des sciences mathématiques*, Zurich, Leemann frères et Cie, p. 144-152 ;
- 1944 « Review of Gentzen 1943 », *The Journal of Symbolic Logic*, vol. 9, p. 70-72 ;
- 1957/1976 (2003) « Considérations sur le paradoxe de Thoralf Skolem », trad. fr. Hourya Benis Sinaceur, dans *Philosophie des mathématiques*, Paris, Librairie Vrin, p. 135-140 ;
- 1959/1976 (2003) « Considérations sur les *Remarques sur les fondements des mathématiques* de Ludwig Wittgenstein », trad. fr. Hourya Benis Sinaceur, dans *Philosophie des mathématiques*, Paris, Librairie Vrin, p. 141-162 ;
- 1961/1976 (2003) « Sur le rôle de la langue du point de vue épistémologique », trad. fr. Hourya Benis Sinaceur, dans *Philosophie des mathématiques*, Paris, Librairie Vrin, p. 177-191 ;
- 1969/1976 (2003) « Remarques pour la philosophie des mathématiques », trad. fr. Hourya Benis Sinaceur, dans *Philosophie des mathématiques*, Paris, Librairie Vrin, p. 193-198 ;
- 1970/1976 (2003) « La correspondance schématique et les structures idéalisées », trad. fr. Hourya Benis Sinaceur, dans *Philosophie des mathématiques*, Paris, Librairie Vrin, p. 199-210 ;
- 1971/1976 (2003) « Symposium sur les fondements des mathématiques », trad. fr. Hourya Benis Sinaceur, dans *Philosophie des mathématiques*, Paris, Librairie Vrin, p. 211-233 ;
- 1972 (2003) « Letter to Gödel, march 16 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. IV, Correspondence : A - G*, New-York, Oxford University Press, p. 293-297 ;
- 1975 (2003) « Letter to Gödel, january 24 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. IV, Correspondence : A - G*, New-York, Oxford University Press, p. 310-313 ;
- 1976a (2003) « Avant-propos », trad. fr. Hourya Benis Sinaceur, dans *Philosophie des mathématiques*, Paris, Librairie Vrin, p. 21-24 ;
- 1976b (2003) « Appendice à Bernays 1930 », trad. fr. Hourya Benis Sinaceur, dans *Philosophie des mathématiques*, Paris, Librairie Vrin, p. 41-81.

Berto Francesco :

- 2007 *How to sell a contradiction*, London, College Publications ;
- 2009a « The Gödel paradox and Wittgenstein's reasons », *Philosophia Mathematica (III)*, vol. 17, n° 2, p. 208-219 ;
- 2009b *There's something about Gödel*, Malden, Wiley-Blackwell.

Beth Evert :

- 1937 « L'évidence intuitive dans les mathématiques modernes » dans *Travaux du IX^{ème} congrès International de Philosophie*, Paris, Hermann, VI, p. 161-165.

Bezboruah A., Shepherdson John :

- 1976 « Gödel's second incompleteness theorem for Q », *The Journal of Symbolic Logic*, n° 41, vol. 2, p. 503-512.

Bienvenu Laurent, Romashchenko Andrei, Shen Alexander, Taveneaux Antoine, Vermeeren Stijn :

- 2004 « The axiomatic power of Kolmogorov complexity », *Annals of Pure and Applied Logic*, n° 165, p. 1380-1402.

Black Robert :

- 2000 « Proving Church's thesis », *Philosophia Mathematica*, vol. 8, p. 244-258.

Blanché Robert, Dubucs Jacques :

- 1996 *La logique et son histoire*, Paris, Librairie Armand Colin.

Blass Andreas, Gurevich Yuri :

- 2003 (2006) « Algorithms : a quest for absolute definitions » dans Olszewski Adam, Wolenski Jan, Janusz Robert (ed.), *Church Thesis after 70 years*, Frankfurt, Ontos-Verlag, p. 24-57.

Boker Udi, Dershowitz Nachum :

- 2008 « The Church-Turing thesis over arbitrary domains » dans Avron Arnon (ed.), *Pillars of computer science*, Berlin, Springer, p. 199-229.

Boniface Jacqueline :

- 2004 *Hilbert et la notion d'existence en mathématiques*, Paris, Librairie Vrin.

Boolos George :

- 1968 « Review of Lucas 1961 and Benacerraf 1967 », *The Journal of Symbolic Logic*, vol. 33, n° 4, p. 655-656 ;
- 1987 « A curious inference », *The Journal of Philosophical Logic*, vol.16, p.1-12 ;
- 1990 « On "seeing" the truth of the Gödel sentence », *Behavioural and Brain Sciences*, vol. 13, p. 655-656 ;
- 1994 « Gödel's second incompleteness theorem explained in word of one syllable, *Mind* », n° 103, vol. 409, p. 1-3 ;
- 1995 « Introductory note to 1951 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. III, Unpublished essays and lectures*, New-York, Oxford University Press, p. 290-304.

Boolos George, Jeffrey Richard :

- 1989 *Computability and logic*, 3^{ème} édition, Cambridge, Cambridge University Press.

Boos William :

- 1998 « The *Transzendenz* of mathematical 'experience' », *Synthese*, vol. 114, n° 1, p. 49-98.

Borel Emile :

- 1952 *Les nombres inaccessibles*, Paris, Gauthiers-Villars.

Bouveresse Jacques :

- 1997 « La logique et le langage » dans *Encyclopédie Philosophique Universelle, Les Notions*, Paris, Presses Universitaires de France, p. 465-475.

Boyle Franklin :

- 1990 « Algorithms and physical laws », *Behavioural and Brain Sciences*, vol. 13, n° 4, p. 656-657.

Boyer Alain :

- 1986 (2001) « L'utopie unificatrice et le Cercle de Vienne » dans Sebestik Jan, Soulez Antonia (ed.), *Le Cercle de Vienne. Doctrine et controverses*, Paris, L'Harmattan, p. 253-269,

Bringsjord Selmer, Arkoudas Konstantine :

- 2006 « On the provability, veracity and AI-relevance of the Church-Turing thesis » dans Olszewski Adam, Wolenski Jan, Janusz Robert (ed.), *Church Thesis after 70 years*, Frankfurt, Ontos-Verlag, p. 66-118.

Brisson Luc, Meyerstein Walter :

- 1995 *Puissance et limite de la raison*, Paris, Les Belles Lettres.

Bromand Joachim :

- 2002 « Why paraconsistent logic can only tell half the truth ? », *Mind*, vol. 111, n° 444, p. 741-749.

Brouwer Luitzen :

- 1908 (1992) « Les principes logiques ne sont pas sûrs », trad. fr. Jacques Bouveresse, dans Rivenc François, Rouilhan Philippe de (ed.), *Logique et fondement des mathématiques, Anthologie 1850-1914*, Paris, Librairie Payot, p. 386-392 ;
- 1920 (1993) « Théorie des ensembles intuitioniste », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Intuition et intuitionisme*, Paris, Librairie Vrin, p. 183-189 ;
- 1928 (1992) « Remarques intuitionistes sur le formalisme », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Intuitionisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 247-251 ;
- 1929 (1992) « Mathématique, science et langage », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Intuitionisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 258-269 ;
- 1947 (1992) « Principes directeurs de la mathématique Intuitionniste », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Intuitionisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 416-417 ;
- 1948 (1992) « Conscience, philosophie et mathématique », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Intuitionisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 419-440 ;

- 1952 (1992) « Bases historiques, principes et méthodes de l'Intuitionnisme », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Intuitionisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 446-458 ;
- 1954 (1993) « Points et espaces », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Intuition et intuitionisme*, Paris, Librairie Vrin, p. 195-220 ;
- 1955 (1992) « L'effet de l'Intuitionnisme sur l'algèbre classique de la logique », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Intuitionisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 459-463.

Brown Douglas, Simpson Stephen :

- 1986 « Which set existence axioms are needed to prove the separable Hans-Banach theorem ? », *Annals of Pure and Applied Logic*, vol. 31, n° 2, p. 123-144.

Buchholz Wilfried, Feferman Solomon, Pohlers Wolfram, Sieg Wilfried :

- 1981 *Iterated inductive definitions and subsystems of analysis*, Berlin, Springer-Verlag.

Burgess John :

- 2010 « On the outside looking in : a caution about conservativeness », dans Feferman Solomon, Parsons Charles, Simpson Stephen (ed.), *Kurt Gödel ; Essays for his centennial*, Cambridge, Cambridge University Press, p. 128-141.

Calude Cristian :

- 1994 *Information and Randomness*, Berlin, Springer-Verlag ;
- 2002a « Chaitin Ω numbers, Solovay machines and incompleteness », *Theoretical Computational Science*, n° 284, p. 269-277 ;
- 2002b « Incompleteness, complexity, randomness and beyond », *Minds and machines*, vol. 12, n° 4, p. 503-517.

Calude Cristian, Calude Elena, Svozil Karl :

- 2010 « The complexity of proving chaoticity and the Church-Turing thesis », *Chaos*, p. 037103 1-5.

Calude Cristian, Campbell Karl, Svozil Karl, Stefanescu Doru :

- 1995 « Strong determinism versus computability » dans DePauli-Schimanovich Werner, Köhler Eckehart, Stadler Friedrich (ed.), *The foundational debate : complexity and constructivity in mathematics and physics*, Dordrecht, Kluwer Academic Publishers, p. 115-131.

Calude Cristian, Jürgensen Helmut :

- 2005 « Is complexity a source of incompleteness ? », *Advanced in Applied Mathematics*, p. 1-15.

Calude Cristian, Jürgensen Helmut, Zimand Marius :

- 1994 « Is independence an exception ? », *Applied Mathematics and Computation*, n° 66, p. 63-76,

Calude Elena :

- 2010 → Calude Cristian, etc.

Campbell Karl :

- 1995 → Calude Cristian, etc.

Carty David mac :

- 1999 → Detlefsen Michael, etc.

Cassou-Noguès Pierre :

- 2001 *Hilbert*, Paris, Les Belles Lettres ;
- 2004 *Gödel*, Paris, Les Belles Lettres ;
- 2005 « Gödel and the "objective existence" of mathematical objects », *History and Philosophy of Logic*, n° 26, p. 211-228 ;
- 2006 « Gödel et la thèse de Turing », *Revue d'Histoire des Mathématiques*, n° ? p. ?.

Cavallés Jean :

- 1937 « Réflexions sur le fondement des mathématiques » dans *Travaux du IX^{ème} congrès international de philosophie*, Paris, Hermann, VI, p. 136-139.

Chaitin Gregory :

- 1974 (1986) « Information-theoretic computational complexity » dans Tymoczko Thomas (ed.), *News directions in the philosophy of mathematics*, Boston, Birkhäuser Verlag, p. 287-299 ;
- 1982 (1986) « Gödel's theorem and information » dans Tymoczko Thomas (ed.), *News directions in the philosophy of mathematics*, Boston, Birkhäuser Verlag, p. 300-311 ;
- 1987a « Incompleteness theorems for random reals », *Advanced in Applied Mathematics*, vol. 8, p. 119-146 ;
- 1987b *Algorithmic information theory*, Cambridge, Cambridge University Press ;
- 1991 « Le hasard des nombres », *La Recherche*, n° 232, p. 610-615 ;
- 1992 *Information-theoretic incompleteness*, Singapour, World Scientific ;
- 1998 *The limits of mathematics*, Singapour, Springer ;
- 1999 *The unknowable*, Singapour, Springer ;
- 2005 (2009) *Hasard et complexité en mathématiques*, trad. fr. Laurence Decréau, Paris, Flammarion ;
- 2006 « Les limites de la raison », *Pour la Science*, n° 342, p. 70-76.

Chalmers David :

- 1990 « Computing the thinkable », *Behavioural and Brains Sciences*, vol. 13, n° 4, p. 658-659 ;
- 1995 « Minds, machines and mathematics ; A review of *Shadows of the mind* by Roger Penrose », *Psyche* 2, en ligne, n° 9, p. 1-10.

Chiffi Daniele :

- 2012 *Kurt Gödel : philosophical explorations : history and theory*, Roma, Aracne.

Chihara Charles :

- 1972 « On alleged refutations of mechanism using Gödel's incompleteness results », *The Journal of Philosophy*, vol. 59, n° 17, p. 507-526 ;
- 1973 *Ontology and the vicious circle principle*, Ithaca, Cornell University Press ;
- 1975 → Berg Jonathan, etc. ;
- 1984 « Priest, the liar and Gödel », *The Journal of Philosophical Logic*, n° 13, p. 117-124.

Church Alonzo :

- 1936a « An Unsolvable Problem of Elementary Number Theory », *American Journal of Mathematics*, n° 58, p. 345-363 ;
- 1936b « A note on the entscheidungsproblem », *The Journal of Symbolic Logic*, vol. 1, n° 1, p. 40-41 ;
- 1940 « On the concept of a random sequence », *Bulletin of the American Mathematical Society*, vol. 46, n° 2, p. 130-135 ;
- 1956 *Introduction to mathematical logic ; Vol. I*, Princeton, Princeton University Press.

Cleland Carol :

- 1993 « Is the Church-Turing thesis true ? », *Minds and Machines*, vol. 3, p. 283-312 ;
- 1995 « Effective procedures and computable functions », *Minds and Machines*, vol. 5, p. 9-23 ;
- 2001 « Recipes, algorithms and programs », *Minds and Machines*, vol. 11, p. 219-237 ;
- 2002 « On effective procedures », *Minds and Machines*, vol. 12, p. 159-179 ;
- 2004 « The concept of computability », *Theoretical Computer Science*, n° 317, p. 209-225 ;
- 2006 « The Church-Turing thesis. A last vestige of a failed mathematical program » dans Olszewski Adam, Wolenski Jan, Janusz Robert (ed.), *Church Thesis after 70 years*, Frankfurt, Ontos-Verlag, p. 119-146.

Coder David :

- 1969 « Goedel's theorem and mechanism », *Philosophy*, vol. 44, p. 234-237.

Cohen Paul :

- 1967 (1971) « Comments on the foundations of set theory », dans Scott Diana (ed.), *Axiomatic set theory, Proceedings of symposia in pure mathematics*, vol. XIII, part I ; *University of California 1967*, Los Angeles, p. 9-15.

Copeland Jack :

- 1998 « Turing's O-machines, Searle, Penrose and the brain », *Analysis*, vol. 58, n° 2, p. 128-138 ;
- 2000 « Narrow versus wide mechanism : including a re-examination of Turing's views on the mind-machine issue », *The Journal of Philosophy*, vol. 97, n° 01, p. 5-32 ;
- 2002a « Accelerating Turing machines », *Minds and Machines*, vol. 12, p. 281-301 ;
- 2002b « The Church-Turing thesis », *The Stanford Encyclopedia of Philosophy*, (En ligne), p. 1-15 ;
- 2002c « Hypercomputation », *Minds and Machines*, vol. 12, n° 4, p. 461-502 ;
- 2004 « Hypercomputation : philosophical issues », *Theoretical Computer Science*, n° 317, p. 251-267 ;
- 2006 « Turing's thesis » dans Olszewski Adam, Wolenski Jan, Janusz Robert (ed.), *Church Thesis after 70 years*, Frankfurt, Ontos-Verlag, p. 147-174.

Copeland Jack, Proudfoot Diane :

- 1999 « Alan Turing forgotten ideas in computer science », *Scientific American*, April, p. 76-81.

Copeland Jack, Shagrir Oron :

- 2013 « Turing versus Gödel on computability and the mind » dans Copeland Jack, Posy Carl, Shagrir Oron (ed.), *Computability : Turing, Gödel, Church and beyond*, Cambridge (USA), MIT Press, p. 1-33.

Copeland Jack, Sylvan Richard :

- 1999 « Beyond the universal Turing machine », *Australasian Journal of Philosophy*, vol. 77, n° 1, p. 46-66.

Coquand Thierry :

- 2014 « Recursive functions and constructive mathematics » dans Dubucs Jacques, Bourdeau Michel (ed.), *Constructivity and computability in historical and philosophical perspective*, Dordrecht, Springer, p. 159-167.

Cori René, Lascar Daniel :

- 1993 *Logique mathématique. Tome I*, Paris, Masson ;
- 1994 *Logique mathématique. Tome II*, Paris, Masson.

Costa Newton da, Doria Fransisco, Amaral Furtado do :

- 1993 « Dynamical system where proving chaos is equivalent to proving Fermat's conjecture », *International Journal of Theoretical Physics*, vol. 32, n° 11, p. 2187-2206.

Dantzig L. van :

- 1955 « Is 10^{10} a finite number ? », *Dialectica*, vol. 9, n° 3-4, p. 273-277.

Davis Martin :

- 1977 « Unsolvable problems » dans Barwise Jon (ed.), *Handbook of mathematical logic*, Amsterdam, North Holland, p. 567-594 ;
- 1978 « What is a computation ? » dans Steen Lynn Arthur (ed.), *Mathematics today ; Twelve informal essays*, New York, Springer-Verlag, p. 241-267 ;
- 1982 « Why Gödel didn't have Church's thesis ? », *Information and Control*, n° 54, p. 3-24 ;
- 1990 « Is mathematical insight algorithmic ? », *Behavioural and Brain Sciences*, vol. 13, n° 4, p. 659-660 ;
- 1993 « How subtle is Gödel's theorem ? More on Roger Penrose », *Behavioral and Brain Sciences*, vol. 16, n° 3, p. 611-612 ;
- 1995 « Introductory note to 193? » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. III, Unpublished essays and lectures*, New-York, Oxford University Press, p. 156-163 ;
- 2004 « The myth of hypercomputation » dans Teuscher Christof (ed.), *Alan Turing : life and legacy of a great thinker*, Berlin, Springer, p. 195-211 ;
- 2005 « What did Gödel believe and when did he believe it ? », *The Bulletin of Symbolic Logic*, vol. 11, n° 2, p. 194-206 ;
- 2006 « Why there is no discipline as hypercomputation ? », *Applied Mathematics and Computations*, vol. 178, n° 1, p. 4-7.

Davis Martin, Matijasevic Yuri, Robinson Julia :

- 1976 « Hilbert's tenth problem ; Diophantine equations : positive aspects of a negative solution », dans Browder Félix (ed.), *Mathematical developments arising from Hilbert's problems, Proceedings of symposia in pure mathematics*, vol. 28, Part 2, American Mathematical Society, p. 323-378.

Davis Martin, Sieg Wilfried :

- 2015 « Conceptual confluence in 1936 : Post and Turing », dans Sommaruga Giovanni, Strahm Thomas (ed.), *Turing's revolution : the impact of his ideas about computability*, Cham, Birkhäuser, p. 3-27.

Dawson John :

- 1984a « Kurt Gödel in Sharper Focus », *The Mathematical Intelligencer*, n° 6, vol. 4, p. 9-17 ;
- 1984b « Discussion on the foundation of mathematics », *History and Philosophy of Logic*, n° 5, p. 111-129 ;
- 1985 (1990) « The Reception of Gödel's Incompleteness Theorems » dans Shanker Stuart (ed.), *Gödel's Theorems in Focus*, London, Routledge, p. 75-95 ;
- 1997 *Logical dilemmas : the life and work of Kurt Gödel*, Wellesley, A.K. Peters ;
- 1998 « What hath Gödel wrought ? », *Synthese*, vol. 114, n° 1, p. 3-12.

De Mol Liesbeth :

- 2007 « Closing the circle : an analysis of Emil Post's early work », *The Bulletin of Symbolic Logic*, vol. 12, n° 2, p. 267-289 ;
- 2013 « Generating, solving and the mathematics of Homo Sapiens. Emil Post's views on computation », dans Zenil Hector (ed.), *A computable universe*, New Jersey, World Scientific, p. 1-16.

Dehornoy Patrick :

- 2007 « Au-delà du forcing : la notion de vérité essentielle en théorie des ensembles » dans Joinet Jean-Baptiste (ed.), *Logique, dynamique et cognition*, Paris, Publications de la Sorbonne, p. 147-169.

Delahaye Jean-Paul :

- 1994 *Information, complexité et hasard*, 2^{ème} édition (1999), Paris, Hermès ;
- 2006 *Complexités*, Paris, Belin – Pour la Science.

Delessert André :

- 2000 *Gödel, une révolution en mathématiques*, Lausanne, Presses polytechniques et universitaires Romandes.

Dennett Daniel :

- 1972 « Review of Lucas 1970 The freedom of the will », *The Journal of Philosophy*, vol. 59, n° 17, p. 527-531.

Dershowitz Nachum :

- 2008 → Boker Udi, etc.

Dershowitz Nachum, Gurevich Yuri :

- 2008 « A natural axiomatization of computability and proof of Church's Thesis », *The Bulletin of Symbolic Logic*, vol. 14, n° 3, p. 299-350.

Desanti Jean-Toussaint :

- 1990 « Fondements des mathématiques » dans *Encyclopaedia Universalis, Corpus*, p. 713-719.

Detlefsen Michael :

- 1979 (1990) « On Interpreting Gödel's second Theorem » dans Shanker Stuart (ed.), *Gödel's Theorems in Focus*, London, Routledge, p. 131-154 ;
- 1986 *Hilbert's program : an essay on mathematical instrumentalism*, Dordrecht, D. Reidel ;

- 1990 « On an alleged refutation of Hilbert's program using Gödel first incompleteness theorem », *The Journal of Philosophical logic*, n° 19, p. 343-377 ;
- 1993 « Hilbert's formalism : an essay on mathematical instrumentalism », *Revue Internationale de Philosophie*, n° 4, p. 285-304 ;
- 1995 « Wright on the non-mechanizability of intuitionist reasoning », *Philosophica Mathematica*, n° 3, vol. 3, p. 103-119 ;
- 1998 « Constructive existence claims » dans Schirn Matthias (ed.), *The Philosophy of Mathematics Today*, Oxford, Oxford University Press, p. 307-335 ;
- 2001 « What does Gödel's second theorem say ? », *Philosophica Mathematica*, n° 3, vol. 9, p. 37-71 ;
- 2002 « Löb's theorem as a limitation on mechanism », *Mind and Machines*, vol. 12, n° 3, p. 353-381 ;
- 2005 « Formalism » dans Shapiro Stewart (ed.), *The Oxford Handbook of Philosophy of Mathematics and Logic*, Oxford, Oxford University Press, p. 236-317 ;
- 2015 « Gentzen's anti-formalist views » dans Kahle Reinhard, Rathjen Michael (ed.), *Gentzen's centenary ; The quest for consistency*, Cham, Springer, p. 25-44.

Detlefsen Michael, Carty David mac, Bacon John :

- 1999 *Logic from A to Z*, London, Routledge.

Doyle Jon :

- 2002 « What is Church thesis ? An outline », *Mind and Machines*, vol. 12, n° 4, p. 519-520.

Dragalin A. :

- 1985 « Correctness of inconsistent theories with notions of feasibility », *Lectures Notes in Computer Science*, vol. 208, p. 58-79.

Drake Franck :

- 1989 « On the foundations of mathematics in 1987 » dans Ebbinghaus H., ... (ed.), *Logic Colloquium '87*, Amstardam, North Holland, p. 11-25.

Dreben Burton, van Heijenoort Jean :

- 1986 « Introductory note to 1929, 1930 and 1930a » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. I, Publications 1929-1936*, New-York, Oxford University Press, p. 44-59.

Dubucs Jacques :

- 1979 *Les recherches sur les fondements des mathématiques chez Hilbert et dans l'école hilbertienne jusqu'à 1930*, Thèse de 3^{ème} cycle, Paris V ;
- 1986 « Réalisme et anti-mécanisme chez K. Gödel », *Dialectica*, n° 40, vol. 4, p. 297-308,
- 1991 « L'œuvre logique de Kurt Gödel : 1929-1936 » dans Bouveresse Jacques (ed.), *L'âge de la science, n° 4, Philosophie de la logique et philosophie du langage, volume I*, Paris, Odile Jacob, p. 31-51 ;
- 1992 « Arguments Gödéliens contre la psychologie computationnelle », *Actes du Colloque Kurt Gödel : Neuchâtel, 13-14 juin 1991*, p. 73-89 ;
- 1995 « Logique de la connaissance : un panorama » dans Dubucs Jacques, Lepage François (ed.), *Méthodes logiques pour les sciences cognitives*, Paris, Hermès, p. 207-234 ;
- 1996 → Blanché Robert, etc. ;
- 1997 « Logique, effectivité et faisabilité », *Dialogue (Revue canadienne de philosophie)*, vol. 36, n° 1, p. 45-68 ;
- 2002 « Feasibility in logic », *Synthèse*, n° 132, p. 213-237.

Dubucs Jacques, Lepage François :

- 1995 « Introduction » dans Dubucs Jacques, Lepage François (ed.), *Méthodes logiques pour les sciences cognitives*, Paris, Hermès, p. 13-22.

Dummett Michael :

- 1959 (1978) « Wittgenstein's Philosophy of mathematics » dans Eardley Anthony (ed.), *Truth and Other enigmas*, London, Duckworth, p. 166-185 ;
- 1963 (1978) « The philosophical signifiacnce of Gödel's theorem » dans Eardley Anthony (ed.), *Truth and Other enigmas*, London, Duckworth, p. 186-201 ;

- 1975 « Wang's Paradox », *Synthese*, n° 30, p. 301-324 ;
- 1978 (1991) *Philosophie de la logique*, Paris, Les Editions de Minuit ;
- 1994 « Reply to Wright » dans McGuinness Brian, Gianluigi Oliveri (ed.), *The philosophy of Michael Dummett*, Dordrecht, Kluwer, p. 329-338 ;
- 1997 « Review of Pasquale Frasca : Wittgenstein's Philosophy of mathematics, and Cristoffer Gefwert : Wittgenstein on Philosophy and mathematics », *The Journal of Philosophy*, vol. 94, n° 7, p. 359-374.

Ehrenfeucht Andrezej :

- 1967 → Kreisel Georg, etc.

Enderton Herbert :

- 1977 « On Elements of recursion theory » dans Barwise Jon (ed.), *Handbook of mathematical logic*, Amsterdam, North Holland, p. 527-566.

Ewald William :

- 2013a « Introduction to the Bernays Habilitation thesis of 1918 » dans Ewald William, Sieg Wilfried (ed.), *David Hilbert's lectures on the foundations of arithmetic and logic 1917 - 1933*, Heidelberg, Springer, p. 222-230 ;
- 2013b « Introduction to "Ueber die Grundlagen des Denkens" » dans Ewald William, Sieg Wilfried (ed.), *David Hilbert's lectures on the foundations of arithmetic and logic 1917 - 1933*, Heidelberg, Springer, p. 761-764.

Ewald William, Sieg Wilfried :

- 2013a « Introduction to "Prinzipien der Mathematik" » dans Ewald William, Sieg Wilfried (ed.), *David Hilbert's lectures on the foundations of arithmetic and logic 1917 - 1933*, Heidelberg, Springer, p. 32-58 ;
- 2013b « Introduction to the 1920 lectures » dans Ewald William, Sieg Wilfried (ed.), *David Hilbert's lectures on the foundations of arithmetic and logic 1917 - 1933*, Heidelberg, Springer, p. 276-297.

Fallis Don :

- 1996 « The source of Chaitin's incorrectness », *Philosophia Mathematica*, vol. 4, n° 3, p. 261-269.

Feferman Solomon :

- 1960 « Arithmetization of metamathematics in a general setting », *Fundamenta Mathematicae*, n° 49, p. 35-92 ;
- 1962 « Transfinite recursive progressions of axiomatic theories », *The Journal of Symbolic Logic*, vol. 27, n° 3, p. 259-316 ;
- 1964 « Systems of predicative analysis ; Part I », *The Journal of Symbolic Logic*, vol. 29, n° 1, p. 1-30 ;
- 1968 « Systems of predicative analysis ; Part II », *The Journal of Symbolic Logic*, vol. 33, n° 2, p. 193-220 ;
- 1970 « Formal theories for transfinite iterations of generalized inductive definitions and some subsystems of analysis » dans Kino Akiko, Myhill John, Richard Eugène (ed.), *Intuitionism and proof theory*, Amsterdam, North Holland, p. 303-326 ;
- 1977 « Review of Takeuti 1975 », *The Bulletin of the American Mathematical Society*, n° 83, p. 351-361 ;
- 1979a « Review of Schütte 1977 », *The Bulletin of the American Mathematical Society (New Series)*, n° 1, p. 224-228 ;
- 1979b « Constructive theories of functions and classes » dans Boffa M. (ed.), *Logic Colloquium '78*, Amsterdam, North Holland, p. 159-224 ;
- 1979c (1998) « What does logic have to tell us about mathematical proofs ? The proof-theoretic analysis of mathematics » dans Feferman Solomon (ed.), *In the light of logic*, Oxford, Oxford University Press, p. 177-186 ;
- 1981 → Buchholz Wilfried, etc. ;
- 1982 « Inductively presented systems and the formalization of mathematics » dans Dalen D. van, Lascar D., Smiley T. (ed.), *Logic Colloquium '80*, North Holland, p. 95-128 ;
- 1984 (1998) « Kurt Gödel : conviction and caution » dans Feferman Solomon (ed.), *In the light of logic*, Oxford, Oxford University Press, p. 150-164 ;

- 1985 « Working foundations », *Synthese*, vol. 62, n° 2, p. 229-254 ;
- 1986 « Introductory note to 1931c » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. I, Publications 1929-1936*, New-York, Oxford University Press, p. 208-213 ;
- 1987 (1998) « Infinity in mathematics. Is Cantor necessary ? » dans Feferman Solomon (ed.), *In the light of logic*, Oxford, Oxford University Press, p. 28-73, 229-248 ;
- 1988a (1998) « Weyl vindicated : *Das Kontinuum* seventy years later » dans Feferman Solomon (ed.), *In the light of logic*, Oxford, Oxford University Press, p. 249-283 ;
- 1988b « Hilbert's program relativized : proof-theoretical and foundational reductions », *The Journal of Symbolic Logic*, n° 53, vol. 2, p. 364-384 ;
- 1989 « Finitary inductively presented logics » dans Ferro R. (ed.), *Logic Colloquium '88*, Amsterdam, North Holland, p. 191-220 ;
- 1990 « Introductory note to 1972a, Remark 1 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. II, Publications 1938-1974*, New-York, Oxford University Press, p. 281-287 ;
- 1991 « Reflecting on incompleteness », *The Journal of Symbolic Logic*, n° 56, vol. 1, p. 1-49 ;
- 1993a (1998) « Why a little bit goes a long way : logical foundations of scientifically applicable mathematics » dans Feferman Solomon (ed.), *In the light of logic*, Oxford, Oxford University Press, p. 284-298 ;
- 1993b (1998) « What rests on what ? The proof-theoretic analysis of mathematics » dans Feferman Solomon (ed.), *In the light of logic*, Oxford, Oxford University Press, p. 187-208 ;
- 1993c (1998) « Gödel's Dialectica interpretation and its two-way stretch » dans Feferman Solomon (ed.), *In the light of logic*, Oxford, Oxford University Press, p. 209-225 ;
- 1993d « Turing in the land of $O(z)$ » dans Herken Rolf (ed.), *The universal Turing machine : a half-century survey*, Oxford, Oxford University Press, p. 103-134 ;
- 1995a « Introductory note to 1933o » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. III, Unpublished essays and lectures*, New-York, Oxford University Press, p. 36-44 ;
- 1995b « Penrose's Gödelian argument ; A review of *Shadows of the mind* by Roger Penrose », *Psyche* 2, en ligne, n° 7, p. 1-12 ;
- 1996 « Gödel's program for new axioms : why, where, how and what ? » dans Hejek Petr (ed.), *Gödel's 96, Lecture notes in logic*, n° 6, Natick, A.K. Peters, p. 3-22 ;
- 1997 « My route to arithmetization », *Theoria*, vol. LXIII, part 3, p. 168-181 ;
- 1998a « Deciding the undecidable : wrestling with Hilbert's » dans Feferman Solomon (ed.), *In the light of logic*, Oxford, Oxford University Press, p. 3-27 ;
- 1998b → Avigad Jeremy, etc. ;
- 1999 « Does mathematics need new axioms ? », *American Mathematical Monthly*, vol. 106, n° 2, p. 99-111 ;
- 2000a « Why the programs for new axioms need to be questioned ? », *The Bulletin of Symbolic Logic*, n° 6, vol. 4, p. 401-413 ;
- 2000b « Does reductive proof theory have a viable rationale ? », *Erkenntnis*, vol. 53, n° 1, p. 53-96 ;
- 2003 « Paul Bernays : Introductory note » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. IV, Correspondence : A - G*, New-York, Oxford University Press, p. 41-79 ;
- 2005 « The Gödel editorial project : a synopsis », *The Bulletin of Symbolic Logic*, vol. 11, n° 2, p. 132-149 ;
- 2006 « Are there absolutely unsolvable problems ? Gödel's dichotomy », *Philosophia Mathematica*, n° 14, vol. 2, p. 134-152 ;
- 2011 « Liber Herr Bernays ! Lieber Herr Gödel ! Gödel on finitism, constructivity, and Hilbert's program » dans Baaz Matthias, Papadimitriou Christos, Putnam Hilary, Scott Dana, Harper Charles Jr (ed.), *Kurt Gödel and the foundations of mathematics*, New-York, Cambridge University Press, p. 111-133 ;
- 2015 « Theses for computation and recursion on concrete and abstract structures », dans Sommaruga Giovanni, Strahm Thomas (ed.), *Turing's revolution : the impact of his ideas about computability*, Cham, Birkhäuser, p. 105-126.

Feferman Solomon, Jäger Gerhard :

- 1983 « Choice principles, the bar rule and autonomously iterated comprehension schemes in analysis », *The Journal of Symbolic Logic*, vol. 48, n° 1, p. 63-70.

Feferman Solomon, Hellman Geoffrey :

- 1995 « Predicative foundations of arithmetic », *Journal of Philosophical Logic*, vol. 24, p. 1-17.

Feferman Solomon, Solovay Robert :

- 1990 « Introductory note to 1972a, Remark 2 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. II, Publications 1938-1974*, New-York, Oxford University Press, p. 287-292.

Feferman Solomon, Strahm Thomas :

- 2010 « Unfolding finitist arithmetic » dans Feferman Solomon (ed.), *Review of Symbolic Logic*, n° 3, vol. 4, p. 665-689.

Feys Robert, Fitch Frédéric :

- 1969 *Dictionnaire of symbols of mathematical logic*, Amsterdam, North Holland.

Field Hartry :

- 1984 (1996) « Is mathematical knowledge just logical knowledge ? », dans Hart William (ed.), *The philosophy of mathematics*, Oxford, Oxford University Press, p. 235-271 ;
- 1998 « Which undecidable arithmetical sentence have determinate truth values ? », dans Dales Harold, Oliveri Gianluigi (ed.), *Truth in mathematics*, Oxford, Clarendon Press, p. 291-310.

Fitch Frédéric :

- 1969 → Feys Robert, etc.

Fitz Harmut :

- 2006 « Church's thesis and physical computation » dans Olszewski Adam, Wolenski Jan, Janusz Robert (ed.), *Church Thesis after 70 years*, Frankfurt, Ontos-Verlag, p. 175-219.

Floyd Juliet :

- 1995 « On saying what you really want to say: Wittgenstein, Gödel and the trisection of the angle », dans Hintikka Jaakko (ed.), *From Dedekind to Gödel : essays on the development of mathematics*, Dordrecht, Kluwer, p. 373-425 ;
- 2001 « Prose versus proof : Wittgenstein on Gödel, Tarski and truth », *Philosophia Mathematica*, vol. 9, p. 280-307 ;
- 2012a « Wittgenstein's diagonal argument : a variation on Cantor and Turing », dans Dybjer Peter and al. (ed.), *Epistemology versus ontology*, Dordrecht, Springer, p. 25-44 ;
- 2012b « Wittgenstein, Carnap and Turing : contrasting notions of analysis », dans Wagner Pierre (ed.), *Carnap's ideal of explication and naturalism*, Basingstoke, Palgrave MacMillan, p. 34-46 ;
- 2013a « The varieties of rigorous experience », dans Beaney Michael (ed.), *The Oxford handbook of the history of analytic philosophy*, Oxford, Oxford University Press, p. 1003-1040 ;
- 2013b « Turing, Wittgenstein and types : philosophical aspects of Turing's 1944-5 », dans Barry Cooper S and Leeuwen Jan van (ed.), *Alan Turing : his work and impact*, Amsterdam, Elsevier, p. 250-253.

Floyd Juliet, Putnam Hilary :

- 2000 « A note on Wittgenstein's 'Notorious paragraph' about the Gödel theorem », *The Journal of Philosophy*, vol. 97, n° 11, p. 624-633.

Folina Janet :

- 1998 « Church's Thesis : Prelude to a Proof », *Philosophia Mathematica*, n° 3, vol. 6, p. 302-323 ;
- 2006 « Church's thesis and the variety of mathematical justifications » dans Olszewski Adam, Wolenski Jan, Janusz Robert (ed.), *Church Thesis after 70 years*, Frankfurt, Ontos-Verlag, p. 220-241.

Fraenkel Abraham, Bar-Hillel Yehoshua :

- 1958 *Foundations of set theory*, Amsterdam, North Holland.

Franks Curtis :

- 2009 *The autonomy of mathematical knowledge*, Cambridge, Cambridge University Press.

Franzen Torkel :

- 2004 *Inexhaustibility ; A non-exhaustive treatment*, Wellesley, A.K. Peters,
- 2005 *Gödel's Theorem : An incomplete guide to its use and abuse*, Natick, A.K. Peters.

Friedman Harvey :

- 1970 « Iterated inductive definitions and Σ^1_2 -AC » dans Kino Akiko, Myhill John, Richard Eugène (ed.), *Intuitionism and proof theory*, Amsterdam, North Holland, p. 435-442 ;
- 1975 « Some systems of second order arithmetic and their use » dans Duncan James (ed.), *Proceedings of the international congress of mathematicians ; Vancouver 1974*, Vancouver, Canadian Mathematical Congress, p. 235-242 ;
- 1976 « Systems of second order arithmetic with restricted induction », *The Journal of Symbolic Logic*, n° 41, p. 557-559 ;
- 1980 « A strong conservative extension of Peano Arithmetic » dans Keisler Howard, Barwise Jon, Kunen Kenneth (ed.), *The Kleene Symposium*, Amsterdam, North Holland, p. 113-122 ;
- 1986 « Necessary uses of abstract set theory in finite mathematics », *Advances in Mathematics*, n° 60, p. 92-122 ;
- 1998 « Finite functions and the necessary use of large cardinals », *Annals of Mathematics*, n° 148, p. 803-893 ;
- 2000 « Normal mathematics will need new axioms », *The Bulletin of Symbolic Logic*, n° 6, vol 4, p. 434-446.

Friedman Harvey, McAloon Kenneth, Simpson Stephen :

- 1982 « A finite combinatorial principle which is equivalent to the 1-consistency of predicative analysis » dans Metakides George (ed.), *Patras logic symposion*, Amsterdam, North-Holland, p. 197-230.

Friedman Harvey, Simpson Stephen, Smith Rick :

- 1983 « Countable algebra and set existence axioms » dans Metakides George (ed.), *Annals of Pure and Applied Logic*, vol. 25, p. 141-181.

Gaifman Haim :

- 2000 « What Gödel's incompleteness result does and does not show ? », *The Journal of Philosophy*, vol. 97, n° 8, p. 462-470.

Gallier Jean :

- 1991 « What's so special about Kruskal's theorem and the ordinal Γ_0 », *Annals of Pure and Applied Logic*, vol. 53, p. 199-260.

Galton Antony :

- 1999 « The Church-Turing thesis : its nature and status », dans Millican Peter, Clark Andy (ed.), *Machines and thought : The legacy of Alan Turing*, Oxford, Oxford University Press, p. 137-164.

Gandy Robin :

- 1980 « Church's Thesis and Principles for Mechanisms » dans Keisler Howard, Barwise Jon, Kunen Kenneth (ed.), *The Kleene Symposium*, Amsterdam, North Holland, p. 123-148 ;
- 1982 « Limitations to Mathematical Knowledge » dans Dalen D. van, Lascar D., Smiley T. (ed.), *Logic Colloquium '80*, North Holland, p. 129-146 ;
- 1988 « The confluence of ideas in 1936 » dans Herken Rolf (ed.), *The universal Turing machine : a half-century survey*, 2^{ème} édition (1995), Oxford, Oxford University Press, p. 51-102.

Garnham Alan :

- 1990 « Don't ask Plato about the emperor's mind », *Behavioural and Brains Sciences*, vol. 13, n° 4, p. 664-665.

Gauthier Yvon :

- 1989 « Les fondements des mathématiques » dans *Encyclopédie Philosophique Universelle, L'Univers Philosophique*, 3^{ème} édition (1997), Paris, Presses Universitaires de France, p. 541-549 ;
- 1991 *De la logique interne*, Paris, Librairie Vrin ;
- 1993 « Hilbert et la logique interne des mathématiques », *Revue Internationale de Philosophie*, n° 4, p. 305-318 ;
- 1997 *Logique et fondement des mathématiques*, Paris, Diderot.

Geiser James :

- 1974 « A formalization of the Essenin-Volpin's proof theoretical studies by means of nonstandard analysis », *The Journal of Symbolic Logic*, vol. 39, n° 1, p. 81-87.

Gentzen Gerhard :

- 1933 (1969) « On the relation between intuitionniste and classical arithmetic », trad. angl. M. Szabo, dans Szabo M. (ed.), *The collected papers of Gerhard Gentzen*, Amsterdam, North-Holland, p. 53-67 ;
- 1936a (1969) « The consistency of elementary number theory », trad. angl. M. Szabo, dans Szabo M. (ed.), *The collected papers of Gerhard Gentzen*, Amsterdam, North-Holland, p. 132-213 ;
- 1936b (1969) « The concept of infinity in mathematics », trad. angl. M. Szabo, dans Szabo M. (ed.), *The collected papers of Gerhard Gentzen*, Amsterdam, North-Holland, p. 223-233 ;
- 1938a (1969) « The present state of research into the foundations of mathematics », trad. angl. M. Szabo, dans Szabo M. (ed.), *The collected papers of Gerhard Gentzen*, Amsterdam, North-Holland, p. 234-251 ;
- 1938b (1969) « New version of the consistency proof for elementary number theory », trad. angl. M. Szabo, dans Szabo M. (ed.), *The collected papers of Gerhard Gentzen*, Amsterdam, North-Holland, p. 252-286.

Giaquinto Marcus :

- 1983 « Hilbert's philosophy of mathematics », *The British Journal of the Philosophy of Science*, vol. 34, n° 2, p. 119-132 ;
- 2002 *The Search for Certainty*, Oxford, Oxford University Press.

Girard Jean-Yves :

- 1983 « Le théorème d'incomplétude de Gödel » dans *Cinq conférences sur l'indécidabilité*, Paris, Presses de l'ENPC, p. 25-47 ;
- 1989 « Le champ du signe ou la faillite du réductionnisme » dans Nagel Ernest, Newman James, ..., *Le théorème de Gödel*, Paris, Le Seuil, p. 147-171 ;
- 1990 « Théorie de la démonstration » dans *Encyclopaedia Universalis, Corpus*, p. 171-175 ;
- 1995a « La machine de Turing : de la calculabilité à la complexité » dans *La Machine de Turing*, Seuil, p. 13-45 ;
- 1995b « Intelligence artificielle et logique naturelle » dans Girard Jean-Yves (ed.), *La Machine de Turing*, Paris, Seuil, p. 107-131 ;
- 2006 *Le point aveugle. Tome 1 : vers la perfection*, Paris, Hermann.

Glymour Clark, Kelly Kevin :

- 1990 « Why you'll never know whether Roger Penrose is a computer », *Behavioural and Brains Sciences*, vol. 13, n° 4, p. 666-667.

Gochet Paul, Gribomont Pascal :

- 1990 *Logique. Volume I*, 3^{ème} édition (1998), Paris, Hermès.

Gödel Kurt :

- 1929 (1986) « On the completeness of the calculus of logic », trad. angl. Stefan Bauer-Mengelberg, Jean van Heijenoort, dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. I, Publications 1929-1936*, New-York, Oxford University Press, p. 61-101 ;
- 1930 (1972) « La complétude des axiomes du calcul fonctionnel », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Logique mathématique, Textes*, Paris, Librairie Armand Colin, p. 175-185 ;
- 1931 (1989) « Sur les propositions formellement indécidables de Principia Mathematica et des systèmes apparentés I », trad. fr. Jean-Baptiste Scherrer, dans Nagel Ernest, Newman James, ..., *Le théorème de Gödel*, Paris, Le Seuil, p. 107-143 ;
- 1933a (1986) « On independence proofs in the propositional calculus », trad. angl. John Dawson, dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. I, Publications 1929-1936*, New-York, Oxford University Press, p. 269-271 ;
- 1933b (1986) « On intuitionistic arithmetic and number theory », trad. angl. Stefan Bauer-Mengelberg, Jean van Heijenoort, dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. I, Publications 1929-1936*, New-York, Oxford University Press, p. 287-295 ;
- 1933c (1995) « The present situation in the foundation of mathematics », trad. angl. Charles Parsons, dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. III, Unpublished essays and lectures*, New-York, Oxford University Press, p. 45-53 ;

- 1936 (1986) « On the length of proofs », trad. angl. Stefan Bauer-Mengelberg, Jean van Heijenoort, dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. I, Publications 1929-1936*, New-York, Oxford University Press, p. 396-399 ;
- 1938a (1990) « The consistency of the axiom of choice and the generalized continuum hypothesis » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. II, Publications 1938-1974*, New-York, Oxford University Press, p. 26-27 ;
- 1938b (1995) « Lecture at Zilsel's », trad. angl. Charles Parsons, dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. III, Unpublished essays and lectures*, New-York, Oxford University Press, p. 88-113 ;
- 193? (1995) « Undecidable diophantine propositions » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. III, Unpublished essays and lectures*, New-York, Oxford University Press, p. 164-175 ;
- 1941 (1995) « In what sense is intuitionistic logic constructive ? » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. III, Unpublished essays and lectures*, New-York, Oxford University Press, p. 189-200 ;
- 1942 (2003) « Letter to Paul Bernays, February 16 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. IV, Correspondence : A - G*, New-York, Oxford University Press, p. 133,
- 1944 (1969) « La logique mathématique de Russell », trad. fr. J. Miller et J. Milner, *Cahier pour l'Analyse*, vol. 10, p. 84-107 ;
- 1946 (1990) « Remarks before the Princeton bicentennial conference on problems in mathematics » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol II, Publications 1938-1974*, New-York, Oxford University Press, p. 150-153 ;
- 1951 (1995) « Some basic theorems on the foundation of mathematics and their implications » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. III, Unpublished essays and lectures*, New-York, Oxford University Press, p. 304-323 ;
- 1953 (1995) « Les mathématiques sont-elles une syntaxe logique du langage ? », trad. fr. Dominique Fagnot, Gehrard Heinzmann, *Dialogue (Revue Canadienne de philosophie)*, n° 34, vol. 1, p. 3-34 ;
- 1956 (2003) « Letter to John von Neumann, March 20 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol V, Correspondence : H - Z*, New-York, Oxford University Press, p. 373-377 ;
- 1958 (1992) « Sur une extension du point de vue finitiste non encore utilisée à ce jour », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Intuitionnisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 502-507 ;
- 1961/? (1995) « The modern development of the foundations of mathematics in the light of philosophy » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. III, Unpublished essays and lectures*, New-York, Oxford University Press, p. 374-387 ;
- 1962 (2003) « Letter to Leon Rappaport, August 2 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol V, Correspondence : H - Z*, New-York, Oxford University Press, p. 176-177 ;
- 1964a (1986) « Post Scriptum to Gödel 1934 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol I, Publications 1929-1936*, New-York, Oxford University Press, p. 369-371 ;
- 1964b (1992) « Sur la nature du problème du continu de Cantor », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Intuitionnisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 509-531 ;
- 1964c (2003) « Letter to Jean van Heijenoort, August 14 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. V, Correspondence : H - Z*, New-York, Oxford University Press, p. 315-316 ;
- 1966 (2003) « Letter to Constance Reid, March 22 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. V, Correspondence : H - Z*, New-York, Oxford University Press, p. 187-188 ;
- 1967 (2003) « Letter to David Plummer, July 31 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. V, Correspondence : H - Z*, New-York, Oxford University Press, p. 161-162 ;
- 1969 (2003) « Letter to George Brulian, December 10 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. V, Correspondence : H - Z*, New-York, Oxford University Press, p. 330 ;
- 1970 (2003) « Letter [non envoyée] to Yossef Balas, December 10 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. V, Correspondence : H - Z*, New-York, Oxford University Press, p. 9-11 ;
- 1972a (1990) « On an extension of finitary mathematics which has not got been used » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. II, Publications 1938-1974*, New-York, Oxford University Press, p. 271-280 ;

- 1972b (1990) « Some remarks on the undecidability results » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. II, Publications 1938-1974*, New-York, Oxford University Press, p. 305-306 ;
- 1972c (2003) « Letter to Karl Menger, April 20 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. V, Correspondence : H - Z*, New-York, Oxford University Press, p. 132-133 ;
- 1973 (2003) « Letter to Abraham Robinson, July 2 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. V, Correspondence : H - Z*, New-York, Oxford University Press, p. 201 ;
- 1974 (2003) « Letter [non envoyée] to Frederick Sawyer, [non datée] » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. V, Correspondence : H - Z*, New-York, Oxford University Press, p. 210-211 ;
- 1975 (2003) « Letter to Paul Bernays, January 12 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. IV, Correspondence : A - G*, New-York, Oxford University Press, p. 309.

Goldfarb Warren :

- 1979 « Logic in the twenties : the nature of the quantifier », *The Journal of Symbolic Logic*, vol. 44, n° 3, p. 351-368 ;
- 2005 « On Gödel way in : the influence of Rudolph Carnap », *The Bulletin of Symbolic Logic*, vol. 11, n° 2, p. 185-193.

Goldin Dina, Wegner Peter :

- 2008 « The interactive nature of computing : refuting the strong Church-Turing thesis », *Minds and Machines*, vol. 18, n° 1, p. 17-38.

Good I. :

- 1967 « Human and machine logic », *The British Journal for the Philosophy of Science*, vol. 18, p. 144-147 ;
- 1969 « Gödel's theorem is a red herring », *The British Journal for the Philosophy of Science*, vol. 19, p. 357-358.

Goodman Nicolas :

- 1979 (1986) « Mathematics as an objective science » dans Tymocko (ed.), *News directions in the philosophy of mathematics*, Boston, Birkhäuser Verlag, p. 79-94 ;
- 1984 « The knowing mathematician », *Synthèse*, n° 60, p. 21-38 ;
- 1987 « Intensions, Church's thesis and the formalization of mathematics », *Notre Dame Journal of Formal Logic*, vol. 28, n° 4, p. 473-489.

Goodstein Reuben :

- 1957 « Critical notice of Remarks on the foundations of mathematics », *Mind*, vol. VXXVI, p. 549-553 ;
- 1972 « Wittgenstein's philosophy of mathematics » dans Ambrose Alice, Lazerowitz Morris (ed.), *Ludwig Wittgenstein, philosophy and language*, London, Allen and Unwin, p. 271-286.

Goodstein Rebecca :

- 1963 « The signifiacnce of incompleteness theorems », *The British Journal for the Philosophy of Science*, vol. XIV, n° 55, p. 208-220.

Granger Gilles-Gaston :

- 1998 « Le problème du fondement selon Tarski » dans Nef Frédéric, Vernant Denis (ed.), *Le formalisme en question*, Paris, Librairie Vrin, p. 37-47.

Gribomont Pascal :

- 1990 → Gochet Paul, etc.

Guccione Salvatore :

- 1993 « Mind the truth : Penrose's new step in the Gödelian argument », *Behavioral and Brain Sciences*, vol. 16, n° 3, p. 612-613.

Gurevich Yuri :

- 1995 « Platonism, constructivism and computer proofs versus proofs by hands », *The Bulletin of European Association for Theoretical Computer Science*, n° 57, p. 145-166 ;

- 2003 → Blass Andreas, etc. ;
- 2008 → Dershowitz Nachum, etc.

Hajek Petr, Pudlak Pavel :

- 1993 *Metamathematics of first-order arithmetic*, Berlin, Springer.

Hallett Michael :

- 1990 « Physicalism, reductionism and Hilbert » dans Irvine Andrew (ed.), *Physicalism in mathematics*, Dordrecht, Kluwer, p. 183-257 ;
- 1994 « Hilbert 's axiomatic method and the laws of thought » dans Alexander George (ed.), *Mathematics and mind*, Oxford, Oxford University Press, p. 158-200.

Hanson William :

- 1971 « Mechanism and Gödel's theorems », *The British Journal for the Philosophy of Science*, vol. 22, p. 9-16.

Harrington Léo :

- 1977 → Paris Jeff, etc.

Hart William :

- 2008 « Review of Tennant 1997 », *Notre Dame Journal of Formal Logic*, vol. 49, n° 4, p. 447-451.

Hatcher William :

- 1968 *Foundations of mathematics*, Philadelphia, W.B. Saunders.

Heijenoort Jean van :

- 1967a (1985) « Logic as calculus and logic as language » dans *Selected Essay*, Napoli, Bibliopolis, p. 11-16 ;
- 1967b « Introduction to Skolem 1923 » dans Heijenoort Jean van (ed.), *From Frege to Gödel*, 3^{ème} édition (1977), Harvard University Press, p. 302-303 ;
- 1967c « Introduction to Herbrand 1930 » dans Heijenoort Jean van (ed.), *From Frege to Gödel*, 3^{ème} édition (1977), Harvard University Press, p. 525-529 ;
- 1967d « Introduction to Herbrand 1931b » dans Heijenoort Jean van (ed.), *From Frege to Gödel*, 3^{ème} édition (1977), Harvard University Press, p. 618-620 ;
- 1979 (1985) « Absolutivism and relativism in logic » dans *Selected Essay*, Napoli, Bibliopolis, p. 75-83 ;
- 1982 « L'œuvre logique de Jacques Herbrand et son contexte historique » dans Stern J. (ed.), *Proceedings of the Herbrand Symposium ; Logic Colloquium 81*, Amsterdam, North-Holland, p. 57-85 ;
- 1986 → Dreben Burton, etc.

Heinzmann Gerhard :

- 1992 « Wittgenstein et le théorème de Gödel », dans Sebestik Jan, Soulez Antonia (ed.), *Wittgenstein et la philosophie d'aujourd'hui*, Paris, Méridiens Klincksieck-L'Harmattan, p. 209-222.

Heinzmann Gerhard, Proust Jöelle :

- 1988 « Carnap et Gödel : échanges de lettres autour de la définition de l'analyticité », *Logique et Analyse*, n° 123-124, p. 257-291.

Hellman Geoffrey :

- 1995 → Feferman Solomon, etc.

Helmer Olaf :

- 1937 « The signifiacnce of undecidable sentences », *Journal of Philosophy*, vol. 34, p. 490-494.

Henkin Léon :

- 1949 (1972) « La complétude du calcul fonctionnel du premier ordre », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Logique mathématique ; Textes*, Paris, Librairie Armand Colin, p. 186-195.

Herbrand Jacques :

- 1930 (1967) « Investigations in proof theory : the properties of true propositions » dans Heijenoort Jean van (ed.), *From Frege to Gödel*, 3^{ème} édition (1977), Harward University Press, p. 529-581 ;
- 1931a (1968) « Note non signée sur la thèse de Herbrand écrite par Herbrand lui-même » dans Heijenoort Jean van (ed.), *Ecrits logiques*, Paris, Presses Universitaires de France, p. 209-214 ;
- 1931b (1968) « Sur la non-contradiction de l'arithmétique » dans Heijenoort Jean van (ed.), *Ecrits logiques*, Paris, Presses Universitaires de France, p. 221-232 ;
- 1931c (2003) « Letter to Kurt Gödel, april 7 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. V, Correspondence : H - Z*, New-York, Oxford University Press, p. 15-21.

Herzog Siegfried :

- 2014 → Oliveira E. Silva Tomas, etc.

Heyting Arend :

- 1958 « Intuitionism in mathematics » dans Klibansky Raymond (ed.), *La philosophie au milieu du vingtième siècle (Philosophy in the mid-century : a survey, vol. 1 : Logique et philosophie des sciences (Logic and philosophy of science)*, Firenze, La Nuova, p. 101-115 ;
- 1959 « Some remarks on intuitionism » dans Heyting Arend (ed.), *Colloquium : Constructivity in Mathematics*, Amsterdam, North Holland, p. 69-71.

Higginbotham James :

- 1990 « Penrose's platonism », *Behavioural and Brains Sciences*, vol. 13, n° 4, p. 667-668.

Hilbert David :

- 1900a (1996) « On the concept of number », trad. angl. William Ewald, dans Ewald William (ed.), *From Kant to Hilbert : a source book in the foundations of mathematics, vol 2*, Oxford, Clarendon Press, p. 1089-1095 ;
- 1900b « Mathematical problems », trad. angl. Mary Winston Newson, *Bulletin of the American Methemathical Society*, n° 8, p. 437-479 ;
- 1904 (1992) « Sur le fondement de la logique et de l'arithmétique », trad. fr. Hourya Sinaceur, dans Rivenc François, Rouilhan Philippe de (ed.), *Logique et fondement des mathématiques, Anthologie 1850-1914*, Paris, Librairie Payot, p. 255-270 ;
- 1918 (1996) « Axiomatic thought », trad. angl. William Ewald, dans Ewald William (ed.), *From Kant to Hilbert : a source book in the foundations of mathematics, vol 2*, Oxford, Clarendon Press, p. 1105-1115 ;
- 1922 (1992) « Nouvelle fondation des mathématiques ; première communication », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Intuitionisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 111-130 ;
- 1923 (1992) « Les fondements logiques des mathématiques », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Intuitionisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 131-144 ;
- 1926 (1972) « Sur l'infini », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Logique mathématique. Textes*, Paris, Librairie Armand Colin, p. 220-245 ;
- 1928 (1992) « Les fondements des mathématiques », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Intuitionisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 146-163 ;
- 1929 (1992) « Problèmes de fondation des mathématiques », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Intuitionisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 177-185 ;
- 1930 (1996) « Logic and the knowledge of nature », trad. angl. William Ewald, dans Ewald William (ed.), *From Kant to Hilbert : a source book in the foundations of mathematics, vol. 2*, Oxford, Clarendon Press, p. 1157-1165 ;
- 1931 (1992) « Le fondement de l'arithmétique élémentaire », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Intuitionisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 188-196.

Hilbert David, Ackermann Wilhelm :

- 1928 (1950) *Principles of mathematical logic*, 2^{ème} édition (1937), trad. angl. Lewis Hamard, George Lechie, F. Steinhardt, New-York, Chelsea.

Hilbert David, Bernays Paul :

- 1934 (2001) *Fondements des mathématiques. 1*, 2^{ème} édition (1968), trad. fr. François Gaillard, Marcel Guillaume, Paris, L'Harmattan ;

- 1939 (2001) *Fondements des mathématiques. 2*, 2^{ème} édition (1970), trad. fr. François Gaillard, Eugène Guillaume, Marcel Guillaume, Paris, L'Harmattan.

Hintikka Jaakko :

- 1989 « Comment je vois la philosophie » trad. fr. Jacques Riche, dans *Encyclopédie Philosophique Universelle. L'Univers Philosophique*, 3^{ème} édition (1997), Paris, Presses Universitaires de France, p. 758-765 ;
- 1996 (2007) *Les principes des mathématiques revisités*, trad. fr. Manuel Rebuschi, Paris, Librairie Vrin ;
- 1998 « On Gödel's philosophical assumptions », *Synthese*, vol. 114, n° 1, p. 13-23 ;
- 2000 *On Gödel*, Belmont, Wadsworth ;
- 2005 « Kurt Gödel : an introduction », *Revue Internationale de Philosophie*, n° 4, vol. 59, p. 451-458.

Hodges Andrew :

- 1983 (2014) *Alan Turing*, trad. fr. Nathalie Zimmermann et Sébastien Baert, Neuilly-sur-Seine, Michel Lafon ;
- 2004 « What would Alan Turing have done after 1954 ? » dans Teuscher Christof (ed.), *Alan Turing : life and legacy of a great thinker*, Berlin, Springer, p. 43-58 ;
- 2006 « Did Church and Turing have a thesis about machines ? » dans Olszewski Adam, Wolenski Jan, Janusz Robert (ed.), *Church Thesis after 70 years*, Frankfurt, Ontos-Verlag, p. 242-252.

Hofstadter Douglas :

- 1979 (1985) *Gödel, Escher et Bach : les brins d'une guirlande éternelle*, trad. fr. Jacqueline Henry, Robert French, Paris, InterEditions.

Hofweber Thomas :

- 2000 « Proof-theoretic reduction as a philosopher's tool », *Erkenntnis*, vol. 53, n° 1-2, p. 127-146.

Horsten Leon :

- 2006 « Formalizing Church's thesis » dans Olszewski Adam, Wolenski Jan, Janusz Robert (ed.), *Church Thesis after 70 years*, Frankfurt, Ontos-Verlag, p. 253-268.

Horsten Leon, Roelants Herman :

- 1995 « The Church-Turing thesis and effective mundane procedures », *Minds and Machines*, vol. 5, p. 1-8.

Hunter Geoffrey :

- 1971 *Metalogic : an introduction to the metatheory of standard first order logic*, Berkeley, University of California Press.

Ignjatovic Aleksandar :

- 1994 « Hilbert's program and the omega-rule », *The Journal of Formal Logic*, vol. 59, n° 1, p. 322-343.

Imbert Claude ~:

- 1998 « Les équivoques du formalisme » dans Nef Frédéric, Vernant Denis (ed.), *Le formalisme en question*, Paris, Librairie Vrin, p. 317-336.

Imbert Cyrille :

- 2008 *L'opacité intrinsèque de la Nature*, Thèse de Philosophie Paris I.

Isaacson Daniel :

- 1987 (1996) « Arithmetical truth and hidden higher-order concepts » dans Hart William (ed.), *The philosophy of mathematics*, Oxford, Oxford University Press, p. 203-224 ;
- 1992 « Some considerations on arithmetical truth and the Omega-rule » dans Detlefsen Michael (ed.), *Proof, logic and formalization*, London, Routledge & Kegan Paul, p. 94-138 ;
- 2007a « Sentences undecidable in Peano Arithmetic whose truth values are unknown », (En ligne), p. 1-13 ;

- 2007b « Sufficient conditions for undecidability of the Gödel sentence and its truth », (En ligne), p. 1-18.

Isles David :

- 1981 « Remarks on the notion of standard non-isomorphic natural number series » dans Richman Fred (ed.), *Constructive mathematics*, Berlin, Speinger-Varlag, p. 111-134 ;
- 1992 « What evidence is that 2^{65536} is a natural number ? », *Notre-Dame Journal of Formal Logic*, vol. 33, n° 4, p. 465-480.

Israel David :

- 2002 « Reflections on Gödel's and Gandy's reflections on Turing thesis », *Minds and Machines*, vol. 12, p. 181-201.

Iwan St. :

- 2000 « On the untenability of Nelson's predicativism », *Erkenntnis*, vol. 53, n° 1-2, p. 147-154.

Jäger Gerhard :

- 1983 → Feferman Solomon, etc.

Jeffrey Richard :

- 1989 → Boolos George, etc.

Jeroslow Robert :

- 1973 « Redundancies in the Hilbert-Bernays derivability conditions for Gödel second incompleteness theorem », *The Journal of Symbolic Logic*, vol. 38, n° 3, p. 359-367.

Jürgensen Helmut :

- 1994 → Calude Cristian, etc. ;
- 2005 → Calude Cristian, etc.

Kahle Reinhard :

- 2015 « Gentzen's consistency proof in context » dans Kahle Reinhard, Rathjen Michael (ed.), *Gentzen's centenary ; The quest for consistency*, Cham, Springer, p. 3-24.

Kalmar Laszlo :

- 1959 « An argument against the plausibility of Church's Thesis » dans Heyting Arend (ed.), *Constructivity in mathematics*, Amsterdam, North Holland P. C., p. 72-80 ;
- 1967 « Foundations of mathematics - whither now ? » dans Lakatos Imre (ed.), *Problems in the Philosophy of Mathematics*, Amsterdam, North Holland P. C., p. 187-207.

Kelly Kevin :

- 1990 → Glymour Clark, etc.

Kemeny John :

- 1948 « Models of logical systems », *The Journal of Symbolic Logic*, n° 13, vol. 1, p. 16-30 ;
- 1956 « A new approach to semantics. Part I », *The Journal of Symbolic Logic*, n° 21, vol. 1, p. 1-27.

Kennedy Juliette :

- 2003 → Atten Mark van, etc. ;
- 2011 « Review of Curtis 2009 », *The Bulletin of Symbolic Logic*, vol. 17, n° 1, p. 119-122 ;
- 2013 « On formalism freeness : implementing Gödel's 1946 Princeton bicentennial lecture », *The Bulletin of Symbolic Logic*, n° 19, vol. 3, p. 351-393 ;
- 2014 « Gödel's 1946 Princeton bicentennial lecture : an appreciation » dans Kennedy Juliette (ed.), *Interpreting Gödel : critical essays*, Cambridge, Cambridge University Press, p. 109-130 ;
- 2017 « Turing, Gödel and the "bright abyss" » dans Floyd Juliet, Bokulich Alisa (ed.), *Philosophical explorations of the legacy of Alan Turing*, Cham, Springer, p. 63-91.

Kentridge Robert :

- 1990 « Parallelism and patterns of thought », *Behavioural and Brains Sciences*, vol. 13, n° 4, p. 670-671.

Ketland Jeffrey :

- 1999 « Deflationism and Tarski's paradise », *Mind*, n° 429, vol. 108, p. 69-94.

Kirby Laurie, Paris Jeff :

- 1982 « Accessible independence results for Peano Arithmetic », *Bulletin of the London Mathematical Society*, n° 14, p. 285-293.

Kikuchi Makoto, Tanaka Kazuyuki :

- 1994 « On formalization of model-theoretic proofs of Gödel theorems », *Notre-Dame Journal of Formal Logic*, vol. 35, n° 3, p. 403-412.

Kitcher Philip :

- 1976 « Hilbert's epistemology », *Philosophy of Science*, vol. 43, p. 99-115.

Kleene Stephen :

- 1936 « General recursive functions of natural numbers », *Mathematische Annalen*, vol. 112, p. 727-742 ;
- 1938 « On notation for ordinal numbers », *The Journal of Symbolic Logic*, vol. 3, n° 4, p. 150-155 ;
- 1952 (2009) *Introduction to metamathematics*, New-York, Iski Press ;
- 1967 (1971) *Logique mathématique*, trad. fr. Jean Largeault, Paris, Librairie Armand Colin ;
- 1976 (1990) « The Work of Kurt Gödel » dans Shanker Stuart (ed.), *Gödel's Theorems in Focus*, London, Routledge, p. 48-71 ;
- 1978 (1990) « An addendum to "The Work of Kurt Gödel" » dans Shanker Stuart (ed.), *Gödel's Theorems in Focus*, London, Routledge, p. 72-73 ;
- 1986a « Introductory note to 1930b, 1931 and 1932b » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. I, Publications 1929-1936*, New-York, Oxford University Press, p. 126-141 ;
- 1986b « Introductory note to 1934 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. I, Publications 1929-1936*, New-York, Oxford University Press, p. 338-345 ;
- 1987 « Reflections on Church's Thesis », *Notre Dame Journal of Formal Logic*, vol. 28, n° 4, p. 490-498 ;
- 1988 « Turing's analysis of computability and major applications of it » dans Herken Rolf (ed.), *The universal Turing machine : a half-century survey*, 2^{ème} edition (1995), Oxford, Oxford University Press, p. 15-49.

Kline Morris :

- 1980 (1989) *Mathématiques : la fin de la certitude* trad. fr. Jean-Pierre Chrétien-Gori, Christian Lazzeri, Paris, Christian Bourgeois.

Koellner Peter :

- 2006 « On the question of absolute undecidability », *Philosophia Mathematica*, n° 14, vol. 2, p. 153-188 ;
- 2016 « Gödel's disjunction » dans Horsten Leon, Welch Philip (ed.), *Gödel's disjunction ; The scope and limits of mathematical knowledge*, Oxford, Oxford University Press, p. 148-188.

Kotlarski Henryk :

- 1994 « On the incompleteness theorems », *The Journal of Symbolic Logic*, vol. 59, n° 4, p. 1414-1419.

Krajewski Stanislaw :

- 1983 « Philosophical consequences of Gödel's theorem », *The Bulletin of the Section of Logic*, n° 12, p. 157-164 ;
- 1991 « Did Gödel prove we are not machines ? » dans Wolkowski Z. (ed.), *First international Symposium on Gödel's theorems*, Singapour, World Scientific, p. 40-49 ;
- 2006 « Remark on Church's thesis and Godel's theorem » dans Olszewski Adam, Wolenski Jan, Janusz Robert (ed.), *Church Thesis after 70 years*, Frankfurt, Ontos-Verlag, p. 269-280.

Kreisel Georg :

- 1951 « On the interpretation of non finitist proofs. Part I », *The Journal of Symbolic Logic*, n° 16, vol. 4, p. 241-267 ;
- 1952 « On the interpretation of non finitist proofs. Part II », *The Journal of Symbolic Logic*, n° 17, vol. 1, p. 43-58 ;
- 1955 « Models, Translations and Interpretations » dans *Symposium on Mathematical Interpretation of Formal Systems*, Amsterdam, North Holland, p. 26-50 ;
- 1958a « Mathematical significance of consistency proofs », *The Journal of Symbolic Logic*, n° 23, vol. 2, p. 155-182 ;
- 1958b (1992) « Le programme de Hilbert » trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Intuitionnisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 465-500 ;
- 1958c « Review : Wittgenstein's remarks on the foundations of mathematics », *The British Journal for the Philosophy of Science*, vol. IX, p. 135-158 ;
- 1967a « Informal rigour and completeness proofs » dans Lakatos Imre (ed.), *Problems in the Philosophy of Mathematics*, Amsterdam, North Holland, p. 138-186 ;
- 1967b « Mathematical logic : what has it done for the philosophy of mathematics ? » dans Schoenman Ralph (ed.), *Bertrand Russell Philosopher of the century*, London, G. Allen & Unwin, p. 201-272 ;
- 1968 « A survey of proof theory », *The Journal of Symbolic Logic*, n° 33, vol. 3, p. 321-388 ;
- 1970 « Hilbert's program and the search for automatic proof procedures », *Lectures Notes in Mathematics*, n° 125, p. 128-146 ;
- 1971 « The collected papers of Gerhard Gentzen (Book review) », *The Journal of Philosophy*, n° 68, vol. 8, p. 238-265 ;
- 1972 « Which number theoretic problems can be solved in recursive progressions on Π^1_1 path through 0 ? », *The Journal of Symbolic Logic*, n° 37, vol. 2, p. 311-334 ;
- 1974 « A notion of mechanistic theory », *Synthese*, vol. 29, n° 1, p. 11-26 ;
- 1976 « What have we learnt from Hilbert's second problem ? » dans Browder Félix (ed.), *Mathematical developments arising from Hilbert's problems ; Proceedings of symposia in pure mathematics*, vol. 28, part 1, American Mathematical Society, p. 93-130 ;
- 1982 « Review of Pour-El, Richards [1979] and [1981] », *The Journal of Symbolic Logic*, n° 47, vol. 4, p. 900-902 ;
- 1987a « Proof theory : some personal recollections » dans Takeuti Gaisi (ed.), *Proof theory*, Amsterdam, North Holland, p. 395-405 ;
- 1987b « Church's Thesis and the ideal of informal rigour », *Notre Dame Journal of Formal Logic*, vol. 28, n° 4, p. 499-520 ;
- 1989 « Logical aspects of the axiomatic method : on their significance in (traditional) foundations and in some (now) common or garden varieties of mathematics » dans Ebbinghaus H., ... (ed.), *Logic Colloquium '87*, Amsterdam, North Holland, p. 183-217 ;
- 1998 « Second thoughts around some of Gödel's writings : a non academic option », *Synthese*, vol. 114, n° 1, p. 99-160.

Kreisel Georg, Ehrenfeucht Andrezej :

- 1967 « Review of Esenin-Vol'pin 1961 », *The Journal of Symbolic Logic*, n° 32, vol. ?, p. 517.

Kreisel Georg, Krivine Jean-Louis :

- 1967 *Éléments de logiques mathématique*, Paris, Dunod.

Kreisel Georg, Takeuti Gaisi :

- 1974 « Formally self-referential propositions for cut free classical analysis and related systems », *Dissertationes Mathematicae*, n° 118, p. 5-50.

Kripke Saül :

- 2013 « The Church-Turing "thesis" as a special corollary of Gödel completeness theorem » dans Copeland Jack, Posy Carl, Shagrir Oron (ed.), *Computability : Turing, Gödel, Church and beyond*, Cambridge (USA), MIT Press, p. 77-104.

Krivine Jean-Louis :

- 1967 → Kreisel Georg, etc. ;
- 1996 « Une preuve formelle et intuitionniste du théorème de complétude de la logique classique », *The Bulletin of Symbolic Logic*, n° 2, vol. 4, p. 405-421.

Ladrière Jean :

- 1949 « Le rôle du théorème de Gödel dans le développement de la théorie de la démonstration », *Revue Philosophique de Louvain*, n° 47, p. 469-491 ;
- 1951 « Le théorème fondamental de Gentzen », *Revue Philosophique de Louvain*, n° 49, p. 357-384 ;
- 1955 « Mathématiques et formalisme », *Revue des Questions Scientifiques*, V^e série, tome XVI, p. 538-574 ;
- 1957 (1992) *Les limitations internes des formalismes*, réédition, Sceaux, Jacques Gabay ;
- 1958 (1973) « Le destin de la raison et les tâches de la philosophie » dans *Vie sociale et destinée*, J. Duculot, p. 19-37 ;
- 1959 « La philosophie des mathématiques et le problème du formalisme », *Revue Philosophique de Louvain*, N° 57, p. 600-622 ;
- 1960 « Les limitations des formalismes et leur signification philosophique », *Dialectica*, vol. 14, n° 4, p. 279-328 ;
- 1967a « Les limites de la formalisation » dans *Logique et connaissance scientifique*, Piaget Jean (ed.), Encyclopedie de la Pleïade, Paris, Gallimard, p. 312-333 ;
- 1967b « Sens et système », *Esprit*, n° 360, p. 822-824 ;
- 1976 « L'abîme » dans *Savoir, faire, espérer : les limites de la raison*, Bruxelles, Publication des Facultés Universitaires Saint-Louis, p. 171-191 ;
- 1989 « La forme et le sens » dans *Encyclopédie Philosophique Universelle. L'Univers philosophique*, 3^{ème} édition (1997), Paris, Presses Universitaires de France, p. 475-492 ;
- 1990 « Limitation (théorèmes de) » dans *Encyclopédie Philosophique Universelle. Les notions philosophiques*, 2^{ème} édition (1998), Paris, Presses Universitaires de France, p. 1484-1487 ;
- 1998 « La philosophie et le formalisme » dans *Encyclopédie Philosophique Universelle. Le discours philosophique*, Paris, Presses Universitaires de France, p. 2271-2280 ;
- 2002 Préface à Houziaux Alain *Dieu à la limite de l'infini, une légitimation du discours théologique*, Paris, Cerf, p. 11-38 ;
- 2005a « Les institutions de la raison » dans *Les défis de la rationalité*, Louvain, Peeters, p. 17-49 ;
- 2005b *Bibliographie de Jean Ladrière*, Louvain, Peeters.

Lakatos Imre :

- 1967 (1986) « A renaissance of empirism in the recent philosophy of mathematics » dans Tymoczko Thomas (ed.), *News directions in the philosophy of mathematics*, Boston, Birkhäuser Verlag, p. 30-48 ;
- 1976 (1984) *Preuves et réfutations*, trad. fr. Nicolas Balacheff, Jean-Pierre Laborde, Paris, Hermann.

Lambalgen Michiel van :

- 1989 « Algorithmic information theory », *The Journal of Symbolic Logic*, n° 54, vol. 4, p. 1389-1400.

Lambek Jim :

- 1994 « Are the traditionnal philosophies of mathematics really incompatible ? », *The Mathematical Intelligencer*, n° 16, vol. 1, p. 56-62 ;
- 1997 « Programms, grammars and arguments : a personnal view of some connections between computation, langage and logic », *The Bulletin of Symbolic Logic*, n° 3, vol. 3, p. 312-328.

Largeault Jean :

- 1972 Introduction à Hilbert 1926 dans Largeault Jean (ed.), *Logique mathématique. Textes*, Paris, Librairie Armand Colin, p. 215-220 ;
- 1992a Préface à Largeault Jean (ed.), *Intuitionisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 7-14 ;
- 1992b Introduction à Hilbert 1922 dans Largeault Jean (ed.), *Intuitionisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 107-111 ;
- 1992c Introduction à Hilbert 1923 dans Largeault Jean (ed.), *Intuitionisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 131 ;
- 1992d Introduction à Hilbert 1929 dans Largeault Jean (ed.), *Intuitionisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 175-177 ;
- 1992e Introduction à Brouwer 1928 dans Largeault Jean (ed.), *Intuitionisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 239-247 ;

- 1992f Introduction et commentaires à Gödel 1964b dans Largeault Jean (ed.), *Intuitionisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 509-531 ;
- 1993 *Intuition et Intuitionisme*, Paris, Librairie Vrin.

Lascar Daniel :

- 1993 → Cori René, etc. ;
- 1994 → Cori René, etc.

Lee Bowie G. :

- 1973 « An argument against Church's thesis », *The Journal of Philosophy*, vol. 70, n° 3, p. 70-76 ;
- 1982 « Lucas' number is finally up », *The Journal of Philosophical Logic*, vol. 11, p. 279-285.

Lepage François :

- 1995 → Dubucs Jacques, etc.

Levi Albert :

- 1964 « Wittgenstein as dialectician », *The Journal of Philosophy*, vol. 61, n° 4, p. 127-139.

Lewis David :

- 1969 « Lucas against mechanism », *Philosophy*, vol. 44, p. 231-233 ;
- 1979 « Lucas against mechanism II », *Canadian Journal of Philosophy* vol. 9, n° 3, p. 373-376.

Lindström Per :

- 2001 « Penrose's new argument », *The Journal of Philosophical Logic*, n° 30, p. 241-250 ;
- 2006 « Remarks on Penrose's "new argument" », *The Journal of Philosophical Logic*, n° 35, p. 231-237.

Löb Martin :

- 1955 « Solution of a problem of Leon Henkin », *The Journal of Symbolic Logic*, n° 20, vol. 2, p. 115-118 ;
- 1959 « Constructive truth » dans Heyting Arend (ed.), *Constructivity in mathematics*, Amsterdam, North Holland P. C., p. 159-168.

Loebl Martin, Matousek Jiri :

- 1987 « On undecidability of the weakened Kruskal theorem », dans Simpson Stephen (ed.), *Logic and combinatorics*, Providence, American Mathematical Society, p. 275-280.

Longo Giuseppe :

- 1999a « L'intelligence mathématique, l'infini et les machines », *Revue de synthèse*, n° 1, p. 111-138 ;
- 1999b « Mathematical intelligence, Infinity and Machines : Beyond Gödelitis », *The Journal of Consciousness Studies*, n° 6, vol. 11, p. 191-214 ;
- 2002 « On the proofs of some formally unprovable propositions and prototype proofs in Type theory », *Lecture notes in Computer Science*, n° 2277, p. 160-180.

Löwenheim Leopold :

- 1915 (1972) « Sur les possibilités du calcul des relations », trad. fr. Jean Largeault dans Largeault Jean (ed.), *Logique mathématique. Textes*, Paris, Librairie Armand Colin, p. 116-138.

Lucas John :

- 1961 (1983) « L'esprit humain, la machine et Gödel », trad. fr. Patrice Blanchard dans Anderson Alan (ed.), *Pensée et machine*, Seyssel, Champs-Vallon, p. 81-97 ;
- 1968a « Satan stultified : a rejoinder to Paul Benacerraf », *The Monist*, n° 52, vol. 1, p. 145-158 ;
- 1968b « Human and machine logic : a rejoinder », *The British Journal for the Philosophy of Science*, vol. 19, n° 1, p. 155-156 ;
- 1999 « Minds, machines and Gödel : a retrospect », dans Millican Peter, Clark Andy (ed.), *Machines and thought : The legacy of Alan Turing*, Oxford, Oxford University Press, p. 103-124.

Lutz Rudi :

- 1990 « Quantum AI », *Behavioural and Brains Sciences*, vol. 13, n° 4, p. 672-673.

MacAloon Kenneth :

- 1979 « Formes combinatoires du théorème d'incomplétude », *Séminaire Bourbaki ; Exposés 507-524*, Berlin, Springer-Verlag, p. 263-275 ;
- 1982 → Friedman Harvey, etc. ;
- 1985 « Paris-Harrington incompleteness and progressions of theories », dans Nerode Anil, Shore Richard (ed.), *Recursion theory*, Providence, American Mathematical Society, p. 447-460.

MacCall Storrs :

- 1999 « Can a Turing Machine know that the Gödel sentence is true ? », *The Journal of Philosophy*, vol. 96, n° 10, p. 525-532 ;
- 2001 « On "seeing" the truth of the Gödel sentence », *Facta Philosophica*, n° 3, p. 25-29.

MacCullough Daryl :

- 1995 « Can humans escape Gödel ? A review of *Shadows of the mind* by Roger Penrose », *Psyche* 2, en ligne, n° 4, p. 1-9.

MacIntyre Angus :

- 2011 « The impact of Gödel's incompleteness theorems on mathematics » dans Baaz Matthias, Papadimitriou Christos, Putnam Hilary, Scott Dana, Harper Charles Jr (ed.), *Kurt Gödel and the foundations of mathematics*, New-York, Cambridge University Press, p. 149-188.

MacLennan Bruce :

- 1990 « The discomforts of dualism », *Behavioural and Brain Sciences*, vol. 13, n° 4, p. 673-674.

Maddy Pénélope :

- 2000 « Does mathematics need new axioms ? », *The Bulletin of Symbolic Logic*, n° 6, vol. 4, p. 413-422 ;
- 2005 « Mathematical existence », *The Bulletin of Symbolic Logic*, n° 11, vol. 3, p. 351-376.

Manaster-Ramer Alexis, Savitch Walter, Zadrozny Wlodek :

- 1990 « Gödel redux », *Behavioural and Brain Sciences*, vol. 13, n° 4, p. 675-676.

Mancosu Paolo :

- 1998 « Hilbert and Bernays on metamathematics » dans Mancosu Paolo (ed.), *From Brouwer to Hilbert, the debate on the foundation of mathematics in the 1920's*, New-York, Oxford University Press, p. 149-188 ;
- 1999a « Between Russell and Hilbert : Behmann on the foundations of mathematics », *The Bulletin of Symbolic Logic*, vol. 5, n° 3, p. 303-330 ;
- 1999b « Between Vienna and Berlin : The immediate reception of Gödel's incompleteness theorems », *History and Philosophy of Logic*, vol. 20, p. 33-45 ;
- 2004 « Review of Gödel's Collected Works, vol. IV and V », *Notre-Dame Journal of Formal Logic*, vol. 45, n° 2, p. 109-125.

Mancosu Paolo, Zach Richard :

- 2015 « Heinrich Behmann's 1921 lecture on the decision problem and the algebra of logic », *The Bulletin of Symbolic Logic*, n° 21, vol. 2, p. 164-187.

Marchal Bruno :

- 1990 « Des fondements théoriques pour l'IA et la philosophie de l'esprit », *Revue Internationale de Philosophie*, n° 1, vol. 172, p. 104-117 ;
- 1998 *Calculabilité, Physique et Cognition*, Paris, Presses Universitaires de France.

Marion Mathieu :

- 1995 « Wittgenstein and finitism », *Synthese*, vol. 105, n° 2, p. 141-176.

Martin Roger :

- 1964 *Logique contemporaine et formalisation*, Paris, Presses Universitaires de France.

Martin-Löf Per :

- 1987 « Truth of a proposition, evidence of a judgement, validity of a proof », *Synthese*, vol. 73, p. 407-420.
- Matijasevic Yuri :
- 1976 → Davis Martin, etc. ;
 - 1993 (1995) *Le dixième problème de Hilbert*, trad. fr. Patrick Cegielski, Denis Richard, Paris, Masson.
- Matousek Jiri :
- 1987 → Loebel Martin, etc.
- Mendelson Elliott :
- 1963 « On some recent criticism of Church's thesis », *Notre-Dame Journal of Formal Logic*, vol. IV, n° 3, p. 201-205 ;
 - 1990 « Second thoughts about Church's thesis and mathematical proofs », *The Journal of Philosophy*, vol. 87, n° 5, p. 225-233 ;
 - 2006 « On the impossibility of proving the "hard-half" of Church's thesis » dans Olszewski Adam, Wolenski Jan, Janusz Robert (ed.), *Church Thesis after 70 years*, Frankfurt, Ontos-Verlag, p. 304-309.
- Meyerstein Walter :
- 1995 → Brisson Luc, etc.
- Meyer Robert :
- 1996 « Kurt Gödel and the consistency of R## » dans Hejek Petr (ed.), *Gödel's 96 ; Lecture notes in logic* n° 6, Natick, A.K. Peters, p. 247-256.
- Milne Peter :
- 2007 « On Gödel's sentences and what they say », *Philosophia Mathematica*, n° 14, vol. 2, p. 193-226.
- Monk James :
- 1976 *Mathematical logic*, New-York, Springer-Verlag.
- Montague Richard :
- 1960 « Towards a general theory of computability », *Synthese*, vol. 12, p. 429-438.
- Moore Gregory :
- 1991 « Sixty years after Gödel », *The Mathematical Intelligencer*, vol. 13, n° 3, p. 6-11.
- Moschovakis Yiannis :
- 1968 « Review of Porte 1960, Kalmar 1959, Peter 1959 and Mendelson 1963 », *The Journal of Symbolic Logic*, vol. 33, n° 3, p. 471-472.
- Mosconi Jean :
- 1998 « Calculabilité et formalisation » dans Nef Frédéric, Vernant Denis (ed.), *Le formalisme en question*, Paris, Librairie Vrin, p. 49-65 ;
 - 2014 « The developments of the concept of machine computability from 1936 to 1960s » dans Dubucs Jacques, Bourdeau Michel (ed.), *Constructivity and computability in historical and philosophical perspective*, Dordrecht, Springer, p. 37-56.
- Mostowski Andrzej :
- 1952 *Sentences undecidable in formalized arithmetic*, Amsterdam, North Holland ;
 - 1966 *Thirty years of foundational studies*, Oxford, Blackwell.
- Mundici Daniele, Sieg Wilfried :
- 1995 « Paper machines », *Philosophia Mathematica*, vol. 3, p. 5-30 ;
 - 2017 « Turing, the mathematician » dans Floyd Juliet, Bokulich Alisa (ed.), *Philosophical explorations of the legacy of Alan Turing*, Cham, Springer, p. 39-62.

Murawski Roman :

- 1999 *Recursive functions and metamathematics : problems of completeness and decidability*, Dordrecht, Kluwer Academic Press.

Murawski Roman, Wolenski Jan :

- 2006 « The status of Church's thesis » dans Olszewski Adam, Wolenski Jan, Janusz Robert (ed.), *Church Thesis after 70 years*, Frankfurt, Ontos-Verlag, p. 331-352.

Mycielski Jan :

- 1981 « Analysis without actual infinity », *The Journal of Symbolic Logic*, n° 46, vol. 3, p. 625-633.

Myhill John :

- 1950 « A complete theory of natural, rational and real numbers », *The Journal of Symbolic Logic*, vol. 15, p. 185-196 ;
- 1952a « A derivation of number theory from ancestral theory », *The Journal of Symbolic Logic*, vol. 17, n° 3, p. 192-197 ;
- 1952b « Some philosophical implications of mathematical logic », *The Review of Metaphysics*, n° 6, vol. 2, p. 165-198 ;
- 1960 « Some Remarks on the notion of proof », *The Journal of Philosophy*, n° 57, vol. 14, p. 461-471 ;
- 1966 « Notes towards an axiomatization of intuitionistic analysis », *Logique et Analyse*, n° 35, p. 280-297.

Nagel Ernest, Newman James :

- 1958 (1989) « La démonstration de Gödel », trad. fr. Jean-Baptiste Scherrer, dans Nagel Ernest, Newman James, ..., *Le théorème de Gödel*, Paris, Le Seuil, p. 70-103.

Nef Frédéric :

- 1998 « Trivalence, modalité, indéterminisme » dans Nef Frédéric, Vernant Denis (ed.), *Le formalisme en question*, Paris, Librairie Vrin, p. 67-83.

Nelson Edward :

- 1986 *Predicative arithmetic*, Princeton, Princeton University Press ;
- 1992 « Mathematical mythologies » dans Sinaceur Hourya, Salanskis Jean-Michel (ed.), *Le labyrinthe du continu*, Paris, Springer Verlag, p. 155-167 ;
- 1993 « Taking formalism seriously », *The Mathematical Intelligencer*, n° 15, vol. 3, p. 8-11.

Nelson R. :

- 1987 « Church's Thesis and cognitive science », *Notre Dame Journal of Formal Logic*, n° 28, vol. 4, p. 581-614.

Neumann John von :

- 1931 (2003) « Letter to Kurt Gödel, January 12 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. V, Correspondence : H - Z*, New-York, Oxford University Press, p. 341-345.

Newman James :

- 1958 (1989) → Nagel Ernest, etc.

Niebergall Karl-Georg :

- 2000 « On the logic of reductibility : axioms and examples », *Erkenntnis*, vol. 53, n° 1, p. 27-61.

Niebergall Karl-Georg, Schirn Matthias :

- 1998 « Hilbert's finitism and the notion of infinity » dans Schirn Matthias (ed.), *The Philosophy of Mathematics Today*, Oxford, Oxford University Press, p. 271-305.

Ochem Pascal, Rao Michael :

- 2012 « Odd perfect numbers are greater than 10^{1500} », *Mathematics of computation*, vol. 81, n° 79, p. 1869-1877.

Odifreddi Piergiorgio :

- 1989 *Classical recursion theory*, Amsterdam, North-Holland ;
- 2006 « Kreisel's Church » dans Olszewski Adam, Wolenski Jan, Janusz Robert (ed.), *Church Thesis after 70 years*, Frankfurt, Ontos-Verlag, p. 353-382.

Oliveira E. Silva Tomas, Herzog Siegfried, Pardi Silvio :

- 2014 « Empirical verification of the even Goldbach conjecture, and computation of prime gaps up to 4.10^8 », *Mathematics of computation*, vol. 83, n° 0, p. 2033-2060.

Olszewski Adam :

- 2006 « Church's thesis as formulated by Church ; An interpretation » dans Olszewski Adam, Wolenski Jan, Janusz Robert (ed.), *Church Thesis after 70 years*, Frankfurt, Ontos-Verlag, p. 383-392.

Panza Marco :

- 1995 « Platonisme et intentionnalité » dans Panza Marco, Salanskis Jean-Michel (ed.), *L'objectivité mathématique. Platonisme et structures formelles*, Paris, Masson, p. 85-132.

Pardi Silvio :

- 2014 → Oliveira E. Silva Tomas, etc.

Parikh Rohit :

- 1971 « Existence and feasibility in arithmetic », *The Journal of Symbolic Logic*, n° 36, vol. 3, p. 494-508 ;
- 1986 « Introductory note to 1936c » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. I, Publications 1929-1936*, New-York, Oxford University Press, p. 394-397.

Paris Jeff :

- 1982 → Kirby Laurie, etc. ;
- 1987 → Wilkie A., etc.

Paris Jeff, Harrington Léo :

- 1977 « A mathematical incompleteness in Peano arithmetic » dans Barwise Jon (ed.), *Handbook of mathematical logic*, Amsterdam, North Holland, p. 1133-1142.

Parsons Charles :

- 1970 « On a number theoretic choice schema and its relation to induction » dans Kino Akiko, Myhill John, Richard Eugène (ed.), *Intuitionism and proof theory*, Amsterdam, North Holland, p. 459-473 ;
- 1990 « Introductory note to 1946 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. II, Publications 1938-1974*, New-York, Oxford University Press, p. 144-150 ;
- 1995 → Sieg Wilfried, etc. ;
- 1998 « Finitism and intuitive knowledge » dans Schrin Matthias (ed.), *The Philosophy of Mathematics Today*, Oxford, Oxford University Press, p. 249-270 ;
- 2000 « Reason and intuition », *Synthèse*, vol. 125, p. 299-315 ;
- 2003a « Alonzo Church : Introductory note » dans Feferman Salomon (ed.), *Kurt Gödel Collected works, vol. IV, Correspondence : A - G*, Oxford, Clarendon Press, p. 361-366 ;
- 2003b « Arend Heyting : Introductory note » dans Feferman Salomon (ed.), *Kurt Gödel Collected works, vol. V, Correspondence : H - Z*, Oxford, Clarendon Press, p. 27-33 ;
- 2008a « Paul Bernay's later philosophy of mathematics », dans Dimitracopoulos Costas, Newelski Ludomir, Normann Dag, Steel John (ed.), *Logic Colloquium 2005*, Cambridge, Cambridge University Press, p. 129-150 ;
- 2008b *Mathematical thought and its objects*, Cambridge, Cambridge University Press ;
- 2009 « Review of Tait 2005 », *Philosophia Mathematica (III)*, vol. 17, n° 2, p. 220-247.

Pégny Maël :

- 2013 « Sur les limites empiriques du calcul », Thèse de l'Université Paris I.

Penrose Roger :

- 1989 (1992) *L'esprit, l'ordinateur et les lois de la physique*, trad. fr. Françoise Balibar, Claudine Tiercelin, Paris, InterEditions ;

- 1990 « The non algorithmic mind », *Behavioural and Brains Sciences*, vol. 13, n° 4, p. 692-705 ;
- 1993 « An emperor still without mind », *Behavioural and Brains Sciences*, vol. 16, n° 3, p. 616-622 ;
- 1994 (1995) *Les ombres de l'esprit*, trad. fr. Christian Jeanmougin, Paris, InterEditions ;
- 1996 « Beyond the doubting of a shadow ; A reply to commentaries on Shadows of the mind », *Psyche 2*, en ligne, n° 23, p. 1-40 ;
- 1997 « On understanding understanding », *International Studies in the Philosophy of Science*, vol. 11, n° 1, p. 7-20 ;
- 2011 « Gödel, the mind and the laws of physics » dans Baaz Matthias, Papadimitriou Christos, Putnam Hilary, Scott Dana, Harper Charles Jr (ed.), *Kurt Gödel and the foundations of mathematics*, New-York, Cambridge University Press, p. 339-358 ;
- 2016 « On attempting to model the mathematical mind » dans Cooper Barry, Hodges Andrew (ed.), *The once and future Turing*, Cambridge (GB), Cambridge University Press, p. 361-378.

Perlis Don :

- 1990 « The emperor's old hat », *Behavioural and Brains Sciences*, vol. 13, n° 4, p. 680-681.

Péter Rózsa :

- 1957 (1967) *Recursive functions*, 2^{ème} edition (1959), trad. angl. Istvan Földes, New York, Academic Press ;
- 1967 (1981) *Recursive functions in computer theory*, trad. angl. I. Juhasz, New York, Halsted Press.

Petitot Jean :

- 1992 « La bimodalité objective du continu et le platonisme transcendantal » dans Sinaceur Hourya, Salanskis Jean-Michel (ed.), *Le labyrinthe du continu*, Paris, Springer Verlag, p. 239-263.

Pohlers Wolfram :

- 1981 → Buchholz Wielfried, etc. ;
- 1998 « Subsystems of set theory and second-order number theory », dans Buss Samuel (ed.), *Handbook of proof theory*, Amsterdam, Elsevier, p. 209-335.

Post Emil :

- 1921 (1972) « Introduction à une théorie générale des propositions élémentaires », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Logique mathématique. Textes*, Paris, Librairie Armand Colin, p. 30-53 ;
- 1936 « Finite combinatory processes-formulation 1 », *The Journal of Symbolic Logic*, n° 1, vol. 3, p. 103-105 ;
- 1941-1965 « Absolutely unsolvable problems and relatively undecidable propositions » dans Davis Martin (ed.), *The undecidable*, Hewlett, Raven Press, p. 338-433 ;
- 1944 (1965) « Recursively enumerable sets of positive integers and their decision problems » dans Davis Martin (ed.), *The undecidable*, Hewlett, Raven Press, p. 305-337 ;
- 1947 « Recursive unsolvability of a problem of Thue », *The Journal of Symbolic Logic*, n° 1, vol. 12, p. 1-11.

Potter Michael :

- 2000 *Reason's nearest kin : Philosophies of Arithmetic from Kant to Carnap*, Oxford, Oxford University Press.

Pour-El Marian, Richards Ian :

- 1979 « A computable ordinary differential equation which possesses no computable solution », *Annals of Mathematical Logic*, n° 1, p. 209-230 ;
- 1981 « The wave equation with computable initial data such that its unique solution is not computable », *Advances in Mathematics*, n° 39, p. 215-239.

Prawitz Dag :

- 1979 « Proofs and the meaning and completeness of the logical constants » dans Hintikka Jaakko, Niiniluoto Ilkka, Saarinen Esa (ed.), *Essays on mathematical and philosophical logic*, Dordrecht, Reidel, p. 25-40 ;
- 1981 (1986) « Philosophical aspects of proof theory » dans Floistad G. (ed.), *Contemporary philosophy, a new survey. Vol. 1*, Dordrecht, Martinus Nijhoff, p. 235-277.

Priest Graham :

- 1979 « The logic of paradox », *The Journal of Philosophical Logic*, n° 8, p. 219-241 ;
- 1984 « The logic of paradox revisited », *The Journal of Philosophical Logic*, n° 12, p. 153-180 ;
- 1987 *In contradiction : A study of the transconsistent*, 2^{ème} édition (2006), Oxford, Clarendon Press ;
- 1994 « Is arithmetic consistent ? », *Mind*, vol. 103, n° 411, p. 337-349 ;
- 2000 « Motivations for Paraconsistency : the Slippery Slope from Classical Logic to Dialetheism » dans Batens Diderik, ... (Ed.), *Frontiers of Paraconsistent Logic*, Baldock, Research Studies Press, p. 223-232.

Proust Jöelle :

- 1988 → Heinzmann Gerhard, etc.

Pudlak Pavel :

- 1993 → Hajek Petr, etc. ;
- 1996 « On the light of proofs of consistency » dans *Collegium Logicum ; Annals of the Kurt Gödel Society*, vol. 2, Wien, Springer, p. 65-86 ;
- 1999 « A note on applicability of the incompleteness theorem to human mind », *Annals of pure and applied logic*, vol. 96, p. 335-342.

Putnam Hilary :

- 1960 (1983) « Pensée et machine », trad. fr. Patrice Blanchard, dans Anderson Alan (ed.), *Pensée et machine*, Seyssel, Champs-Vallon, p. 110-134 ;
- 1967 (1983) « Mathematics without foundations » dans Benacerraf Paul, Putnam Hilary (ed.), *Philosophy of Mathematics selected readings*, London, Cambridge University Press, p. 295-311 ;
- 1971 (1996) *Philosophie de la logique*, trad. fr. Patrick Peccatte, Combas, L'éclat ;
- 1980 (1983) « Models and reality » dans *Philosophy of Mathematics selected readings*, London, Cambridge University Press, p. 421-444 ;
- 1995 « Review of Shadows of the mind », *Bulletin of the American Mathematical Society*, vol. 32, n° 3, p. 370-373 ;
- 2000a → Floyd Juliet, etc. ;
- 2000b « Nonstandard models and Kripke's proof of the Gödel theorem », *Notre Dame Journal of Formal Logic*, vol 41, n° 1, p. 53-58 ;
- 2011 « The Gödel theorem and human nature » dans Baaz Matthias, Papadimitriou Christos, Putnam Hilary, Scott Dana, Harper Charles Jr (ed.), *Kurt Gödel and the foundations of mathematics*, New-York, Cambridge University Press, p. 325-337.

Quine Willard van Orman :

- 1953a (2003) *Du point de vue logique*, trad. fr. sous la dir. de Sandra Laugier, Paris, Librairie Vrin ;
- 1953b « On ω -inconsistency and a so-called axiom of infinity », *The Journal of Symbolic Logic*, vol. 18, n° 2, p. 119-124 ;
- 1972 (1984) *Méthodes de logique*, 3^{ème} édition, trad. fr. Maurice Clavelin, Paris, Armand Colin.

Raatikainen Panu :

- 1998 « On interpreting Chaitin's incompleteness theorem », *The Journal of Philosophical Logic*, vol. 27, p. 569-586 ;
- 2000 « Algorithmic information theory and undecidability », *Synthese*, vol. 123, p. 217-225 ;
- 2002 « Mc Call's Godelian argument is invalid », *Facta Philosophica*, vol. 4, n° 1, p. 167-169 ;
- 2003 « Hilbert's program revisited », *Synthese*, n° 137, p. 157-177 ;
- 2005 « On the philosophical relevance of Gödel's incompleteness theorem », *Revue Internationale de Philosophie*, n° 4, vol. 59, p. 513-534 ;
- 2013 « Gödel's incompleteness theorems », *The Stanford Encyclopedia of Philosophy*, (En ligne), p. 1-35.

Rao Michael :

- 2012 → Ochem Pascal, etc.

Rav Yehuda :

- 1998 « La philosophie de la logique » dans *Encyclopédie Philosophique Universelle, L'Univers philosophique*, Paris, Presses Universitaires de France, p. 2290-2309.
- Ravaglia Mark :
- 2005 → Sieg Wilfried, etc.
- Reck Erich :
- 2002a → Awodey Steve, etc. ;
 - 2002b → Awodey Steve, etc.
- Reid constance :
- 1970 *Hilbert*, Berlin, Springer Verlag.
- Reinhardt William :
- 1985 « Absolute versions of incompleteness Theorems », *Noûs*, n° 19, vol. 3, p. 317-346,
 - 1986 « Epistemic theories and the interpretation of Gödel's incompleteness theorems », *The Journal of Philosophical Logic*, n° 15, p. 427-474.
- Rescorla Michael :
- 2007 « Church's Thesis and the conceptual analysis of computability », *Notre Dame Journal of Formal Logic*, vol. 48, n° 2, p. 253-280.
- Resnik Michael :
- 1974 (1990) « On the Philosophical Significance of Consistency Proofs » dans Shanker Stuart (ed.), *Gödel's Theorems in Focus*, London, Routledge, p. 115-130.
- Rey Jean- François :
- 2004 *Calculabilité, complexité et approximation*, Paris, Vuibert.
- Richard Ian :
- 1979 → Pour-El Marian, etc. ;
 - 1981 → Pour-El Marian, etc.
- Richardson Daniel :
- 1968 « Some undecidable problems involving elementary functions of a real variable », *The Journal of Symbolic Logic*, vol. 33, n° 4, p. 514-520.
- Riche Jacques :
- 1998 « Finitization procedures and finite model property », *Logique et Analyse*, p. 155-165.
- Rivenc François :
- 1993 *Recherches sur l'universalisme logique*, Paris, Petite Bibliothèque Payot.
- Robinson Abraham :
- 1965 « Formalism 64 » dans Bar-Hillel Yehoshua (ed.), *Logic, methodology and philosophy of science*, Amsterdam, North-Holland Publishing Company, p. 228-246.
- Robinson Julia :
- 1976 → Davis Martin, etc.
- Robinson Raphael :
- 1958 « Review of Peter 1957 », *The Journal of Symbolic Logic*, vol. 23, n° 3, p. 362-363.
- Rodych Victor :
- 1999 « Wittgenstein's inversion of Gödel's theorem », *Erkenntnis*, vol. 51, p. 173-206 ;
 - 2002 « Wittgenstein on Gödel : the newly published remarks », *Erkenntnis*, vol. 56, p. 379-397 ;
 - 2003 « Misunderstanding Gödel : new arguments about Wittgenstein and new remarks by Wittgenstein », *Dialectica*, vol. 57, n° 3, p. 279-313.
- Roelants Herman :

- 1995 → Horsten Leon, etc.
- Roeper Peter :
- 2003 « Giving an account of provability within a theory », *Philosophia Mathematica*, vol. 11, p. 332-340.
- Romashchenko Andrei :
- 2004 → Bienvenu Laurent, etc.
- Rosen Robert :
- 1988 « Effective processes and natural law » dans Herken Rolf (ed.), *The universal Turing machine : a half-century survey*, 2^{ème} édition (1995), Oxford, Oxford University Press, p. 485-498.
- Roskies Adina :
- 1990 « Seeing truth or just seeming true ? », *Behavioural and Brains Sciences*, vol. 13, n° 4, p. 682-683.
- Ross David :
- 1974 « Church's Thesis : what its difficulties are and are not ? », *The Journal of Philosophy*, vol. 71, n° 15, p. 515-525.
- Rosser Barkley :
- 1936 « Extensions of some theorems of Gödel and Church », *The Journal of Symbolic Logic*, vol. 1, n° 3, p. 87-91 ;
 - 1937 « Gödel theorems for non-constructive logics », *The Journal of Symbolic Logic*, vol. 2, n° 3, p. 129-137 ;
 - 1939a « An informal exposition of proofs of Gödel's theorems and Church's theorem », *The Journal of Symbolic Logic*, vol. 4, n° 2, p. 53-60 ;
 - 1939b « Review of Gentzen 1938 », *Bulletin of the American Mathematical Society*, vol. 45, n° 7, p. 812-813.
- Rouilhan Philippe de :
- 1998 « Tarski et l'universalité de la logique » dans Nef Frédéric, Vernant Denis (ed.), *Le formalisme en question*, Paris, Librairie Vrin, p. 85-102.
- Rucker Rudy :
- 1995 *Infinity and the mind*, Princeton, Princeton University Press.
- Sabatier Xavier :
- 2009 *Les formes du réalisme mathématique*, Paris, Librairie Vrin.
- Salanskis Jean-Michel :
- 1995 « Platonisme et philosophie des mathématiques » dans Panza Marco, Salanskis Jean-Michel (ed.), *L'objectivité mathématique. Platonisme et structures formelles*, Paris, Masson, p. 179-212 ;
 - 1999 *Le constructivisme non standard*, Villeneuve-d'Ascq, Presses Universitaires du Septentrion.
- Savitch Walter :
- 1990 → Manaster-Ramer Alexis, etc.
- Scarpellini Bruno :
- 2003 « Comments on "Two undecidable problems of analysis" », *Minds and Machines*, vol. 13, n° 1, p. 79-85.
- Schirn Matthias :
- 1998 → Niebergall Karl-Georg, etc.
- Schlipf John :
- 1975 → Barwise Jon, etc.
- Schurz Gerhard :

- 2002 « McCall and Raatikainen on mechanism and incompleteness », *Facta Philosophica*, vol. 4, n° 1, p. 171-174.

Schütte Kurt :

- 1965 « Predicative well-orderings », *Formal systems and recursive functions*, Amsterdam, North-Holland, p. 280-303.

Searle John :

- 1997 (1999) *Le mystère de la conscience*, trad. fr. Claudine Tiercelin, Paris, Odile Jacob.

Seligman Jeremy :

- 2002 « The scope of Turing's analysis of effective procedures », *Minds and Machines*, vol. 12, n° 2, p. 203-220.

Setzer Anton :

- 2015 « The use of trustworthy principles in a revised Hilbert's program » dans Kahle Reinhard, Rathjen Michael (ed.), *Gentzen's centenary ; The quest for consistency*, Cham, Springer, p. 45-60.

Shagrir Oron :

- 2002 « Effective computations by humans and machines », *Minds and Machines*, vol. 12, n° 2, p. 221-240 ;
- 2003 → Copeland Jack, etc. ;
- 2006 « Gödel on Turing on computability » dans Olszewski Adam, Wolenski Jan, Janusz Robert (ed.), *Church Thesis after 70 years*, Frankfurt, Ontos-Verlag, p. 393-419.

Shanker Stuart :

- 1987 « Wittgenstein versus Turing on the nature of Church Thesis », *Notre Dame Journal of Formal Logic*, vol. 28, n° 4, p. 615-649 ;
- 1988 (1990) « Wittgenstein's remarks on the signifiacnce of Gödel's theorem » dans Shanker Stuart (ed.), *Gödel's Theorems in Focus*, London, Routledge, p. 155-256.

Shapiro Stewart :

- 1980 « On the Notion of Effectiveness », *History and Philosophy of Logic*, n° 1, p. 209-230 ;
- 1981 « Understanding Church's Thesis », *The Journal of Philosophical Logic*, n° 10, p. 353-365 ;
- 1983a « Remarks on the development of computability », *History and Philosophy of Logic*, n° 4, p. 203-220 ;
- 1983b (1996) « Conservativeness and incompleteness » dans Hart William (ed.), *The philosophy of mathematics*, Oxford, Oxford University Press, p. 225-234 ;
- 1990 « Review of Kleene 1981, Davis 1982 and Kleene 1987 », *The Journal of Symbolic Logic*, vol. 23, n° 1, p. 348-350 ;
- 1991 *Foundations without foundationalism : a case for second-order logic*, Oxford Logic Guides 17, Oxford, Clarendon Press ;
- 1994 « Reasoning, logic and computation », *Philosophia Mathematica*, vol. 2, p. 31-51 ;
- 1998 « Incompleteness, mechanism and optimism », *The Bulletin of Symbolic Logic*, n° 4, vol. 3, p. 273-302 ;
- 2002 « Incompleteness and inconsistency », *Mind*, vol. 111, n° 444, p. 817-832 ;
- 2003 « Mechanism, truth, and Penrose's new argument », *The Journal of Philosophical Logic*, n° 32, p. 19-42 ;
- 2006 « Review of Franzen 2005 », *Philosophia Mathematica*, vol. 14, n° 2, p. 262-264 ;
- 2013 « The open texture of computability » dans Copeland Jack, Posy Carl, Shagrir Oron (ed.), *Computability : Turing, Gödel, Church and beyond*, Cambridge (USA), MIT Press, p. 153-181 ;
- 2015 « Proving things about the informal », dans Sommaruga Giovanni, Strahm Thomas (ed.), *Turing's revolution : the impact of his ideas about computability*, Cham, Birkhäuser, p. 283-296.
- 2016 « Idealization, mechanism and knowability » dans Horsten Leon, Welch Philip (ed.), *Gödel's disjunction ; The scope and limits of mathematical knowledge*, Oxford, Oxford University Press, p. 189-207.

Shen Alexander :

- 2004 → Bienvenu Laurent, etc.

Shepherdson John :

- 1976 → Bezboruah A., etc.

Shieh Sanford :

- 1998 « Undecidability in anti-Realism », *Philosophia Mathematica*, n° 3, vol. 6, p. 324-333.

Shore Richard :

- 2010 « Reverse mathematics : the playground of logic », *The Bulletin of Symbolic Logic*, vol. 16, n° 3, p. 378-402.

Sieg Wilfried :

- 1981 → Buchholz Wilfried, etc. ;
- 1984 « Foundations for analysis and proof theory », *Synthese*, n° 60, p. 159-200 ;
- 1985a « Fragments of arithmetic », *Annals of Pure and Applied Logic*, n° 28, p. 33-72 ;
- 1985b « Reductions of theories for analysis » dans Dorn Georg, Weingartner Paul (ed.), *Foundations of logic and linguistics*, New York, Plenum Press, p. 199-231 ;
- 1988 « Hilbert's program sixty years later », *The Journal of Symbolic Logic*, n° 53, vol. 2, p. 338-348 ;
- 1991 « Herbrand analyses », *Archive for Mathematical Logic*, n° 30, p. 409-441 ;
- 1994 « Mechanical procedures and mathematical experience » dans Alexander George (ed.), *Mathematics and mind*, Oxford, Oxford University Press, p. 71-117 ;
- 1995 → Mundici Daniele, etc. ;
- 1997 « Step by Recursive Step : Church's Analysis of Effective Calculability », *The Bulletin of Symbolic Logic*, n° 3, vol. 2, p. 154-180 ;
- 1999 « Hilbert's programs : 1917-1922 », *The Bulletin of Symbolic Logic*, vol. 5, n° 1, p. 1-44 ;
- 2002 « Calculations by man and machine : conceptual analysis » dans Sieg Wilfried, Sommer Y., Talcott C. (ed.), *Reflections on the Foundations of Mathematics ; Essays in Honor of Solomon Feferman*, Urbana, A. K. Peters, p. 390-409 ;
- 2003a « Jacques Herbrand : Introductory note » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. V, Correspondence : H - Z*, New-York, Oxford University Press, p. 2-13 ;
- 2003b « Emil Post : Introductory note » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. V, Correspondence : H - Z*, New-York, Oxford University Press, p. 165-169 ;
- 2003c « John von Neumann : Introductory note » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. V, Correspondence : H - Z*, New-York, Oxford University Press, p. 327-335 ;
- 2005 « Only two letters : the correspondance between Herbrand and Gödel », *The Bulletin of Symbolic Logic*, n° 11, vol. 2, p. 172-184 ;
- 2006 « Gödel on Computability », *Philosophia Mathematica*, n° 14, vol. 2, p. 189-207 ;
- 2008 « Church without dogma : axioms for computability » dans S. Barry Cooper, Benedikt Löwe, Andrea Sorbi (ed.), *New computational paradigms*, New-York, Springer, p. 139-152 ;
- 2009 « On computability » dans Irvine Andrew (ed.), *Philosophy of mathematics*, Amsterdam, Elsevier, p. 535-630 ;
- 2010 « Searching for proofs (and uncovering of capacities of the mathematical mind) » dans Feferman Solomon, Sieg Wilfried (ed.), *Proofs, categories and computations*, London, College Publications, p. 189-215 ;
- 2013a « Gödel's philosophical challenge (to Turing) » dans Copeland Jack, Posy Carl, Shagrir Oron (ed.), *Computability : Turing, Gödel, Church and beyond*, Cambridge (USA), MIT Press, p. 183-201 ;
- 2013b « Normal forms for puzzles : a variant of Turing's thesis » dans Cooper Barry, Leeuwen Jan van (ed.), *Alan Turing : his work and impact*, Amsterdam, Elsevier, p. 332-339 ;
- 2013c *Hilbert's programs and beyond*, Oxford, Oxford University Press ;
- 2013d « Introduction to the lectures on proof theory » dans Ewald William, Sieg Wilfried (ed.), *David Hilbert's lectures on the foundations of arithmetic and logic 1917 - 1933*, Heidelberg, Springer, p. 418-430 ;
- 2013e « Axioms for computability : do they allow a proof of Church's thesis ? », dans Zenil Hector (ed.), *A computable universe*, New Jersey, World Scientific, p. 99-123 ;
- 2015 → Davis Martin ;
- 2017 → Mundici Daniele, etc.

Sieg Wilfried, Parsons Charles :

- 1995 « Introductory note to 1938a » dans Feferman Solomon (ed.), *Kurt Gödel Collected works. Vol. III, Unpublished essays and lectures*, New-York, Oxford University Press, p. 62-85.

Sieg Wilfried, Ravaglia Mark :

- 2005 « David Hilbert and Paul Bernays, Grundagen der Mathematik, first edition (1934, 1939) » dans Grattan-Guinness Ivor (ed.), *Landmark writings in western mathematics, 1640-1940*, Amsterdam, Elsevier, p. 981-999.

Siegelmann Hava :

- 1995 « Computation beyond the Turing limit », *Science*, vol. 268, p. 545-548.

Simpson Stephen :

- 1982 → Friedman Harvey, etc. ;
- 1983 → Friedman Harvey, etc. ;
- 1984 « Which set existence axioms are needed to prove Cauchy/Peano theorem for ordinary differential equations ? », *The Journal of Symbolic Logic*, vol. 49, n° 3, p. 783-802 ;
- 1985a « Reverse mathematics », dans Nerode Anil, Shore Richard (ed.), *Recursion theory*, Providence, American Mathematical Society, p. 461-471 ;
- 1985b « Non provability of certain combinatorial properties of finite trees », dans Feferman Solomon, Parsons Charles, Simpson Stephen (ed.), *Harvey Friedman's research on the foundations of mathematics*, Amsterdam, North-Holland, p. 87-117 ;
- 1985c « Friedman's research on subsystems of second order arithmetic », dans Feferman Solomon, Parsons Charles, Simpson Stephen (ed.), *Harvey Friedman's research on the foundations of mathematics*, Amsterdam, North-Holland, p. 137-160 ;
- 1986 → Brown Douglas, etc. ;
- 1987 « Unprovable theorems and fast-growing functions », dans Simpson Stephen (ed.), *Logic and combinatorics*, Providence, American Mathematical Society, p. 359-394 ;
- 1988 « Partial realizations of Hilbert's program », *The Journal of Symbolic Logic*, n° 53, vol. 2, p. 349-363 ;
- 2009 *Subsystems of second order arithmetic*, Cambridge, Cambridge University Press ;
- 2010 « The Gödel hierarchy and reverse mathematics », dans Feferman Solomon, Parsons Charles, Simpson Stephen (ed.), *Kurt Gödel ; Essays for his centennial*, Cambridge, Cambridge University Press, p. 109-127.

Simmons Keith :

- 2004 « Review of Giaquinto 2002 », *Philosophia Mathematica*, vol. 12, n° 2, p. 162-175.

Simpson Stephen, Smith Rick :

- 1986 « Factorization of polynomials and Σ^0_1 induction », *Annals of Pure and Applied Logic*, n° 31, p. 289-306.

Sinaceur Hourya :

- 1990a « Complétude » dans *Encyclopédie Philosophique Universelle, Les notions philosophiques*, 2^{ème} édition (1998), Paris, Presses Universitaires de France, p. 376-379 ;
- 1990b « Consistance » dans *Encyclopédie Philosophique Universelle, Les notions philosophiques*, 2^{ème} édition (1998), Paris, Presses Universitaires de France, p. 439-441 ;
- 1990c « Constructivité » dans *Encyclopédie Philosophique Universelle, Les notions philosophiques*, 2^{ème} édition (1998), Paris, Presses Universitaires de France, p. 454-457 ;
- 1990d « Décidabilité » dans *Encyclopédie Philosophique Universelle, Les notions philosophiques*, 2^{ème} édition (1998), Paris, Presses Universitaires de France, p. 549-553 ;
- 1991 *Corps et Modèles*, 2^{ème} édition (1999), Paris, Librairie Vrin ;
- 1992 Introduction à Hilbert 1904 dans Rivenc François, Rouilhan Philippe de (ed.), *Logique et fondement des mathématiques, Anthologie 1850-1914*, Paris, Librairie Payot, p. 245-255 ;
- 1993 « Du formalisme à la constructivité : le finitisme », *Revue Internationale de Philosophie*, n° 4, p. 251-283 ;
- 1997 « Logique et ontologie », *Bulletin de la Société Française de Philosophie*, n° 4, p. 97-130 ;
- 1998 « Différents aspects du formalisme » dans Nef Frédéric, Vernant Denis (ed.), *Le formalisme en question*, Paris, Librairie Vrin, p. 129-146 ;
- 2000 « Address at the Princeton University bicentennial conference on problems of mathematics (december 17-19, 1946), by Alfred Tarski », *The Bulletin of Symbolic Logic*, vol. 6, n° 1, p. 1-44 ;

- 2003 Introduction à *Philosophie des mathématiques* de Paul Bernays, Paris, Librairie Vrin, p. 7-17.
- Sipper Moshe :
- 2002 → Teuscher Christof, etc.
- Skolem Thoralf :
- 1923 (1977) « The foundations of elementary arithmetic established by means of the recursive mode of thought, without the use of apparent variables ranging over infinite domains » dans Heijenoort Jan van (ed.), *From Frege to Gödel*, Cambridge, Harvard University Press, p. 303-333 ;
 - 1928 (1972) « Sur la logique mathématique », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Logique mathématique, Textes*, Paris, Librairie Armand Colin, p. 93-108.
- Smart J. :
- 1961 « Gödel's theorem, Church's theorem and mechanism », *Synthese*, vol. 13, n° 1, p. 105-110.
- Smith Peter :
- 2007 *An introduction to Gödel's theorems*, New-York, Cambridge University Press.
- Smith Rick :
- 1983 → Friedman Harvey, etc. ;
 - 1985 « The consistency strengths of some finite forms of the Higman and Kruskal theorems », dans Feferman Solomon, Parsons Charles, Simpson Stephen (ed.), *Harvey Friedman's research on the foundations of mathematics*, Amsterdam, North-Holland, p. 119-136 ;
 - 1986 → Simpson Stephen, etc.
- Smorynski Craig :
- 1977 « The incompleteness theorems » dans Barwise Jon (ed.), *Handbook of mathematical logic*, Amsterdam, North Holland, p. 821-865 ;
 - 1982 « The varieties of arboreal experience », *The Mathematical Intelligencer*, vol. 4, n° 4, p. 182-189 ;
 - 1988 « Hilbert's program », *CWI Quarterly*, vol. 1, n° 4, p. 3-59.
- Smullyan Raymond :
- 1992 (1993) *Les théorèmes d'incomplétude de Gödel*, trad. fr. M. Morgenstern, Paris, Masson ;
 - 1993 (1995) *Théorie de la récursivité pour la métamathématique*, trad. fr. Philippe Ithier, Paris, Masson.
- Soare Robert :
- 1996 « Computability and recursion », *The Bulletin of Symbolic Logic*, vol. 2, n° 3, p. 284-321 ;
 - 2013 « Interactive computing and relativized computability » dans Copeland Jack, Posy Carl, Shagrir Oron (ed.), *Computability : Turing, Gödel, Church and beyond*, Cambridge (USA), MIT Press, p. 203-260 ;
 - 2014 « Turing and the discovery of computability » dans Downey Rod (ed.), *Turing's legacy : developments froms Turing's ideas in logic*, Lecture notes in logic, p. 467-492 ;
 - 2015 « Why Turing's thesis is not a thesis », dans Sommaruga Giovanni, Strahm Thomas (ed.), *Turing's revolution : the impact of his ideas about computability*, Cham, Birkhäuser, p. 297-310.
- Solovay Robert :
- 1990 → Feferman Solomon, etc.
- Soulez Antonia :
- 1986 (2001) « "Que reste-t-il" de la philosophie après le dépassement de "toute la métaphysique" ? Carnap et Heidegger » dans Sebestik Jan, Soulez Antonia (ed.), *Le Cercle de Vienne. Doctrine et controverses*, Paris, L'Harmattan, p. 149-168 ;
 - 1998 « Désaffirmation du formalisme » dans Nef Frédéric, Vernant Denis (ed.), *Le formalisme en question*, Paris, Librairie Vrin, p. 359-403.
- Steel John :
- 2000 « Mathematics need new axioms », *The Bulletin of Symbolic Logic*, n° 6, vol. 4, p. 422-433.

Stefanescu Doru :

- 1995 → Calude Cristian, etc.

Steiner Mark :

- 1991 « Review of *Hilbert's program : an essay on mathematical instrumentalism* », *Philosophia Mathematica*, Michael Detlefsen 1986 », *The Journal of Philosophy*, vol. 88, n° 6, p. 331-336 ;
- 2001 « Wittgenstein as his own worst enemy : the case of Gödel's theorem », *Philosophia Mathematica*, vol. 9, p. 257-279.

Stengers Isabelle :

- 1998 « Qu'est-ce que la science a décidé de ne pas savoir ? » dans Cazenave Michel (ed.), *Dictionnaire de l'ignorance*, Paris, Albin Michel, p. 278-286.

Stewart Ian :

- 1991 « Deciding the undecidable », *Nature*, vol. 352, p. 664-665.

Strahm Thomas :

- 2010 → Feferman Solomon, etc.

Sundholm Göran :

- 2014 « Constructive recursive functions, Church's thesis, and Brouwer's theory of the creating subject : afterthoughts on a parisian joint session » dans Dubucs Jacques, Bourdeau Michel (ed.), *Constructivity and computability in historical and philosophical perspective*, Dordrecht, Springer, p. 1-35.

Svozil Karl :

- 1995 → Calude Cristian, etc. ;
- 2010 → Calude Cristian, etc.

Sylvan Richard :

- 1999 → Copeland Jack, etc.

Szabo M. :

- 1969 « Introduction » dans Szabo M. (ed.), *The collected papers of Gerhard Gentzen*, Amsterdam, North-Holland, p. 1-28.

Tait William :

- 1981 « Finitism », *The Journal of Philosophy*, n° 78, vol. 9, p. 524-546 ;
- 1986 (1996) « Truth and proof : the platonism of mathematics » dans Hart William (ed.), *The philosophy of mathematics*, Oxford, Oxford University Press, p. 142-167 ;
- 2002 « Remarks on finitism » dans Sieg Wilfried, Sommer Y., Talcott C. (ed.), *Reflections on the Foundations of Mathematics ; Essays in Honor of Solomon Feferman*, Urbana, A. K. Peters, p. 410-419 ;
- 2005a *The Provenance of Pure Reason. Essays in the Philosophy of Mathematics and its history*, Oxford, Oxford University Press ;
- 2005b « Gödel's reformulation of Gentzen's first consistency proof for arithmetic : the no-counterexample interpretation », *The Bulletin of Symbolic Logic*, vol. 11, n° 2, p. 225-236 ;
- 2006a « Gödel's interpretation of intuitionism », *Philosophia Mathematica*, vol. 14, n° 2, p. 208-228 ;
- 2006b « Gödel's correspondence on proof theory and constructive mathematics », *Philosophia Mathematica*, vol. 14, n° 2, p. 76-111 ;
- 2010 « Gödel on intuition and on Hilbert's finitism », dans Feferman Solomon, Parsons Charles, Simpson Stephen (ed.), *Kurt Gödel ; Essays for his centennial*, Cambridge, Cambridge University Press, p. 88-108 ;
- 2012 « Primitive recursive arithmetic and its role in the foundations of arithmetic : historical and philosophical reflections », dans Dybjer Peter and al. (ed.), *Epistemology versus ontology*, Dordrecht, Springer, p. 161-180.

Tanaka Kazuyuki :

- 1994 → Kikuchi Makoto, etc.

Takeuti Gaisi :

- 1974 → Kreisel Georg, etc.

Tarski Alfred :

- 1933 (1974) « Quelques remarques sur les concepts d' ω -consistance et d' ω -complétude », trad. fr. sous la dir. de Gilles-Gaston Granger, dans Granger Gilles-Gaston (ed.), *Logique, sémantique et métamathématique : 1923-1944, vol. II*, Paris, Armand Colin, p. 7-21 ;
- 1933-1936 (1972) « Le concept de vérité dans les langages formalisés », trad. fr. sous la dir. de Gilles-Gaston Granger, dans Granger Gilles-Gaston (ed.), *Logique, sémantique et métamathématique : 1923-1944, vol. I*, Paris, Armand Colin, p. 157-269 ;
- 1939 (1974) « Sur des énoncés indécidables dans des systèmes logiques élargis et sur le concept de vérité », trad. fr. sous la dir. de Gilles-Gaston Granger, dans Granger Gilles-Gaston (ed.), *Logique, sémantique et métamathématique : 1923-1944, vol. II*, Paris, Armand Colin, p. 191-202 ;
- 1940-1967 (1974) « La complétude de l'algèbre et la géométrie élémentaires », trad. fr. sous la dir. de Gilles-Gaston Granger, dans Granger Gilles-Gaston (ed.), *Logique, sémantique et métamathématique : 1923-1944, vol. II*, Paris, Armand Colin, p. 203-242 ;
- 1948 (1986) « A decision method for elementary algebra and geometry », 2^{ème} édition (1951), dans Stevan Givant, Ralph MacKenzie (ed.), *Alfred Tarski collected papers, vol. 3*, Basel, Birkhauser, p. 301-367.

Taveneaux Antoine :

- 2004 → Bienvenu Laurent, etc.

Tennant Neil :

- 2002 « Deflationism and the Gödel's phenomena », *Mind*, n° 443, vol. 111, p. 551-582 ;
- 2004 « An anti-realist critique of dialetheism » dans Priest Graham, Beall J.C., Bradley Armour-Grab (ed.), *The law of non-contradiction*, Oxford, Clarendon Press, p. 355-384.

Teuscher Christof, Sipper Moshe :

- 2002 « Hypercomputation : hype or computation ? », *Communications of the Association for Computing Machinery*, n° 45, vol. 8, p. 23-24.

Thomas William :

- 1973 « Doubts about some standard arguments for Church's thesis » dans Bogdan Radu, Niiniluoto Ilkka (ed.), *Papers of the fourth international congress for logic, methodology and philosophy of science*, Dordrecht, Reidel, p. 13-22.

Tieszen Richard :

- 2006 « After Gödel : mechanism, reason and realism in the philosophy of mathematics », *Philosophia Mathematica*, n° 14, vol. 2, p. 229-254 ;
- 2011 *After Gödel*, Oxford, Oxford University Press.

Timpson Christopher :

- 2004 « Quantum computers : the Church-Turing hypothesis versus the Turing principle » dans Teuscher Christof (ed.), *Alan Turing : life and legacy of a great thinker*, Berlin, Springer, p. 213-240.

Tipler Franck :

- 1986 « Cosmological limits in computation », *International Journal of Theoretical Physics*, n° 25, vol. 6, p. 617-661.

Tragesser Robert :

- 1992 « Three insufficiently attended to aspects of most mathematical proofs : phenomenological studies » dans Detlefsen Michael (ed.), *Proof, logic and formalization*, London, Routledge & Kegan Paul, p. 162-198.

Troelstra Anne :

- 1977 « Aspects of constructive mathematics » dans Barwise Jan, *Handbook of mathematical logic*, Amsterdam, North Holland, p. 973-1052 ;

- 1986 « Introductory note to 1933e » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. I, Publications 1929-1936*, New-York, Oxford University Press, p. 282-286 ;
- 1990 « Introductory note to 1958 and 1972 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. II, Publications 1938-1974*, New-York, Oxford University Press, p. 217-241 ;
- 1995 « Introductory note to 1941 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, Vol. III, Unpublished essays and lectures*, New-York, Oxford University Press, p. 186-189.

Turing Alan :

- 1936-1937 (1995) « Théorie des nombres calculables, suivie d'une application au problème de la décision », trad. fr. Julien Basch, dans *La Machine de Turing*, Paris, Seuil, p. 47-104 ;
- 1937 « Computability and λ -definability », *The Journal of Symbolic Logic*, vol. 2, n° 4, p. 153-164 ;
- 1939 (1965) « Systems of logic based on ordinals » dans Davis Martin (ed.), *The undecidable*, Hewlett, Raven Press, p. 155-222 ;
- 1947 (2004) « Lecture on the Automatic Computing Engine » dans Copeland Jack (ed.), *The essential Turing*, Oxford, Oxford University Press, p. 378-394 ;
- 1950 (1995) « Les ordinateurs et l'intelligence », trad. Patrice Blanchard, dans *La Machine de Turing*, Paris, Seuil, p. 135-175 ;
- 1951 (2004) « Can digital computers think ? » dans Copeland Jack (ed.), *The essential Turing*, Oxford, Oxford University Press, p. 482-486 ;
- 1954 « Solvable and unsolvable problems », *Science News*, n° 31, p. 7-23.

Vaught Robert :

- 1954 « Applications of the Löwenheim-Skolem-Tarski theorem to problems of completeness and decidability », *Indagationes Mathematicae*, n° 16, p. 467-472.

Vermeeren Stijn :

- 2004 → Bienvenu Laurent, etc.

Vinnikov Victor :

- 1999 « We shall know : Hilbert's apology », *The Mathematical Intelligencer*, vol. 21, n° 1, p. 42-46.

Visser Albert :

- 1989 « Peano's smart children : a provability logical study of systems with built-in consistency », *Notre-Dame Journal of Formal Logic*, vol. 30, n° 2, p. 161-196,
- 2016 « The second incompleteness theorem : reflections and ruminations » dans Horsten Leon, Welch Philip (ed.), *Gödel's disjunction ; The scope and limits of mathematical knowledge*, Oxford, Oxford University Press, p. 67-91.

Vuillemin Jules :

- 1998 « La philosophie des mathématiques et les théorèmes négatifs de Gödel » dans Nef Frédéric, Vernant Denis (ed.), *Le formalisme en question*, Paris, Librairie Vrin, p. 169-189.

Wagner Pierre :

- 1998 *La machine en logique*, Paris, Presses Universitaires de France ;
- 2014 *Logique et philosophie*, Paris, Ellipses.

Wang Hao :

- 1954 « The formalization of mathematics », *The Journal of Symbolic Logic*, vol. 19, n° 4, p. 241-266 ;
- 1958 « Eighty years of foundational studies », *Dialectica*, vol. 12, p. 466-497 ;
- 1974 *From Mathematics to Philosophy*, London, Routledge & Kegan Paul ;
- 1987 (1990) *Kurt Gödel*, trad. fr. Laura Ovion, Michel Mérieux, Paris, Librairie Armand Colin ;
- 1993 « On physicalism and algorithmism : can machines think ? », *Philosophia Mathematica*, vol. 1, p. 97-138 ;
- 1996 *A logical journey : from Gödel to philosophy*, Cambridge (USA), The MIT Press.

Watson Alister :

- 1938 « Mathematics and its foundations », *Mind*, vol. XLVII, n° 185, p. 440-451.

Webb Judson :

- 1968 « Metamathematics and the philosophy of mind », *Philosophy of Science*, vol. 35, p. 156-178 ;
- 1980 *Mechanism, Mentalism and Metamathematics*, Dordrecht, D. Reidel ;
- 1990 « Introductory note to 1972a, Remark 3 » dans Feferman Solomon (ed.), *Kurt Gödel Collected works, vol. II, Publications 1938-1974*, New-York, Oxford University Press, p. 292-304 ;
- 1997 « Hilbert's formalism and arithmetization of mathematics », *Synthèse*, vol. 110, p. 1-14 ;
- 2005 « Gödel's encounters with formalism, intuition and Kant », *Revue Internationale de Philosophie*, n° 4, vol. 59, p. 491-512.

Wegner Peter :

- 2008 → Goldin Dina, etc.

Wells Benjamin :

- 2002 « Is there a nonrecursive decidable equational theory ? », *Minds and Machines*, vol. 12, n° 2, p. 303-326.

Weyl Hermann :

- 1921 (1992) « Sur la crise contemporaine des fondements des mathématiques », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Intuitionnisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 55-105 ;
- 1926 (1994) « L'état présent de la connaissance en mathématiques », trad. fr. Jean Largeault, dans *Le continu et autres écrits*, Paris, Librairie Vrin, p. 137-161 ;
- 1927 *Philosophy of mathematics and natural science*, Edition revue et augmentée (1949), trad. angl. Olaf Helmer, Princeton, Princeton University Press ;
- 1928 (1992) « Remarques et discussion à propos du second exposé de Hilbert sur les fondements des mathématiques », trad. fr. Jean Largeault, dans Largeault Jean (ed.), *Intuitionnisme et théorie de la démonstration*, Paris, Librairie Vrin, p. 166-169 ;
- 1929 (1994) « La consistance en mathématiques », trad. fr. Jean Largeault, dans *Le continu et autres écrits*, Paris, Librairie Vrin, p. 165-179 ;
- 1946 « Mathematics and Logic. A brief survey serving as a preface to a review of "The philosophy of Bertrand Russell" », *American Mathematical Monthly*, n° 53, p. 2-13.

Wette Eduard :

- 1974 « Contradiction within pure number theory because of a system internal "consistency-deduction" », *International Logic Review*, n° 9, p. 51-62.

Wilkie A, Paris Jeff :

- 1987 « On the scheme of induction for bounded arithmetic formulas », *Annals of Pure and Applied Logic*, n° 35, p. 261-302.

Wilder Raymond :

- 1950 (1986) « The cultural basis of mathematics » dans Tymoczko Thomas (ed.), *News directions in the philosophy of mathematics*, Boston, Birkhäuser Verlag, p. 185-199.

Wille Matthias :

- 2011 « Review of Feferman et al. 2010 », *The Bulletin of Symbolic Logic*, vol. 17, n° 1, p. 125-126.

Willard Dan :

- 2001 « Self-verifying axiom systems. The incompleteness theorem, and related reflection principles », *The Journal of Symbolic Logic*, n° 66, vol. 2, p. 536-596.

Wittgenstein Ludwig :

- 1956 (1983) *Remarques sur les fondements des mathématiques*, 2^{ème} édition (1974), trad. fr. Marie-Anne Lescourret, Anscombe G., Rhees Rush, Wright G. von (ed.), Paris, Gallimard.

Wolenski Jan :

- 2005 « Gödel, Tarsky and truth », *Revue Internationale de Philosophie*, n° 4, vol. 59, p. 459-490 ;
- 2006 → Murawski Roman, etc.

Wright Crispin :

- 1994 « About "The philosophical significance of Gödel's theorem" : some issues » dans McGuinness Brian, Gianluigi Oliveri (ed.), *The philosophy of Michael Dummett*, Dordrecht, Kluwer, p. 167-202 ;
- 1995 « Intuitionists are not (Turing) machines », *Philosophia Mathematica*, vol. 3, p. 86-102.

Yessenin-Volpin Alexander :

- 1961 « Le programme ultra-intuitionniste des fondements des mathématiques » dans Union mathématique internationale (ed.), *Infinitistic methods*, Oxford, Pergamon Press, p. 201-223 ;
- 1970 « The ultra-intuitionistic criticism and the antitraditionnal program for foundations of mathematics » dans Kino Akiko, Myhill John, Richard Eugène (ed.), *Intuitionism and proof theory*, Amsterdam, North Holland, p. 3-45.

Young Paul :

- 1985 « Gödel's theorems, exponential difficulty and undecidability of arithmetic theories : an exposition », dans Nerode Anil, Shore Richard (ed.), *Recursion theory*, Providence, American Mathematical Society, p. 503-522.

Zach Richard :

- 1999 « Completeness before Post : Bernays, Hilbert and the development of propositionnel logic », *The Bulletin of Symbolic Logic*, n° 5, vol. 3, p. 331-366 ;
- 2003a « The practice of finitism : epsilon calculus and consistency proofs in Hilbert's program », *Synthese*, n° 137, p. 211-259 ;
- 2003b « Hilbert's program », *The Stanford Encyclopedia of Philosophy*, (En ligne), p. 1-19 ;
- 2004 « Hilbert's "Verunglückter Beweis", the first epsilon theorem, and consistency proofs », *History and Philosophy of Logic*, n° 25, vol. 2, p. 79-94 ;
- 2005a « Review of Potter 2000 », *Notre Dame Journal of Philosophical Logic*, n° 46, vol. 4, p. 503-513 ;
- 2005b « Kurt Gödel, paper on the incompleteness theorems (1931) » dans Grattan-Guinness Ivor (ed.), *Landmark writings in western mathematics, 1640- 1940*, Amsterdam, Elsevier, p. 917-925 ;
- 2007 « Hilbert's program then and now » dans Jaquette Dale (ed.), *Philosophy of Logic*, Amsterdam, Elsevier-North Holland, p. 411-447 ;
- 2015 → Mancosu Paolo.

Zadrozny Wlodek :

- 1990 → Manaster-Ramer Alexis, etc.

Zimand Marius :

- 1994 → Calude Cristian, etc.

Zwirn Hervé :

- 2000 *Les limites de la connaissance*, Paris, Odile Jacob.