Revue d'Histoire des Mathématiques

Gödel et la thèse de Turing

Pierre Cassou-Noguès

Tome 14 Fascicule 1

2008

SOCIÉTÉ MATHÉMATIQUE DE FRANCE

Publiée avec le concours du Ministère de la culture et de la communication (DGLFLF) et du Centre national de la recherche scientifique

REVUE D'HISTOIRE DES MATHÉMATIQUES

Périodicité : La *Revue* publie deux fascicules par an, de 150 pages chacun environ.

Tarifs 2008 : prix public Europe : 65 €; prix public hors Europe : 74 €;
prix au numéro : 36 €.
Des conditions spéciales sont accordées aux membres de la SMF.

Diffusion : SMF, Maison de la SMF, B.P. 67, 13274 Marseille Cedex 9
AMS, P.O. Box 6248, Providence, Rhode Island 02940 USA

Maquette couverture : Armelle Stosskopf

Revue d'histoire des mathématiques
14 (2008), p. 77–111

GÖDEL ET LA THÈSE DE TURING

Pierre Cassou-Noguès

Résumé. — Cet article porte sur la discussion par Gödel de la thèse de Turing. Pour l'essentiel, nous présentons des notes inédites conservées dans les Archives Gödel, qui apportent des éléments nouveaux sur la relation ambiguë de Gödel à Turing. La première section examine la position qu'avait Gödel avant 1937 sur la possibilité d'une définition de la calculabilité. La deuxième concerne directement l'interprétation par Gödel de la thèse de Turing. Dans plusieurs passages, antérieurs à 1937, Gödel qualifie de « mécaniques » les procédures définies par des règles qui font abstraction du sens des symboles et ne portent que sur leur forme extérieure. Plusieurs notes montrent ensuite que Gödel identifie la thèse de Turing comme posant que ces procédures « mécaniques » (au sens où Gödel l'entendait avant Turing) sont représentables par une machine de Turing. Ce n'est pas en toute rigueur la thèse de Turing, puisque l'article de 1937, pris à la lettre, entend définir les procédures « finies ». Ce déplacement laisse Gödel libre de critiquer, après 1964, le texte de Turing, et la définition des procédures « finies » par des machines de Turing. La dernière section est consacrée à l'analyse d'un argument élaboré contre Turing par lequel Gödel entend justifier la possibilité de procédures finies mais non mécaniques.

Abstract (Gödel and Turing's thesis). — This paper concerns Gödel's remarks on Turing's thesis. Of fundamental importance to this analysis are unpublished notes kept among Gödel's papers. The first section concerns Gödel's position on the possibility of a definition of computability before

Texte reçu le 29 mars 2006, révisé le 7 décembre 2007.

P. Cassou-Noguès, CNRS, UMR, Savoirs, Textes, Langage, Université Lille III, B. P. 149, 59653 Villeneuve d'Ascq.
Courrier électronique : pierre.cassou-nogues@univ-lille3.fr

Classification mathématique par sujets (2000) : 03D10, 03F30.
Mots clefs : Turing, Gödel, machines, calculabilité, incomplétude.
Key words and phrases. — Turing, Gödel, machines, computability, incompleteness.

1937. The second and main section presents different notes on Turing's famous paper of 1937. Before 1937, Gödel qualified as "mechanical" procedures defined by rules that ignore the meaning of symbols and only consider their exterior form. Several notes then show that Gödel interpreted Turing's thesis as the claim that mechanical procedures in that sense can be represented by Turing machines. But Turing himself intended to define "finite" procedures. This shift enabled Gödel after 1964 to criticize Turing's paper. The third and last section deals with Gödel's argument against Turing, in which Gödel aimed to establish the existence of finite but non-mechanical procedures.

Ni dans les articles et conférences rassemblés dans les *Collected Works*, ni dans les notes, qui restent inédites, de la bibliothèque de Princeton, il n'y a véritablement de texte, où Kurt Gödel présente de façon circonstanciée sa position vis-à-vis de la thèse de Turing. Ce ne sont que des remarques fragmentaires, presque des aphorismes. Et celles-ci posent de nombreux problèmes. Le problème le plus marquant, et que signale en particulier J. C. Webb [1990, p. 293] est le suivant. On le sait, dans l'article présenté en 1936 et publié en 1937, Alan Turing utilise la notion de machine pour donner une définition de la calculabilité. Une fonction calculable est une fonction susceptible d'être calculée par une machine d'une certaine sorte, un dispositif donné par une description d'une certaine forme, disons une machine de Turing. Il y a d'autres définitions de la calculabilité. La première (au moyen du λ-calcul) est proposée par Alonzo Church autour de 1934. Une seconde définition est formulée par Gödel lui-même au printemps 1934. Ou, plus exactement, Gödel donne à partir d'une suggestion de Jacques Herbrand un énoncé que l'on peut considérer comme une définition de la calculabilité mais que lui-même ne présente pas comme tel. Nous y reviendrons. Enfin, à la suite de Turing, Emil Post et Stephen Kleene en 1936 proposeront deux autres définitions. Toutes ces définitions sont équivalentes. Mais, pour Gödel, c'est seulement l'analyse de Turing qui saisit le contenu intuitif de la notion de calculabilité et montre alors l'adéquation de ces définitions. C'est pourquoi, du reste, nous parlons de la thèse de Turing plutôt que de la thèse de Church-Turing comme on le fait habituellement. La définition de Turing justifie les autres, celle de Church par exemple, qui lui est équivalente. En particulier, c'est toujours sur la définition de Turing que s'appuie Gödel pour reformuler son théorème d'incomplétude et lui donner sa plus grande généralité. En même temps, Gödel [1972b, p. 306], dénonce « une erreur philosophique dans le travail de Turing ». Il soutient l'existence de « procédures finies non mécaniques » ou, par conséquent, des « procédures calculables » mais calculables de façon non mécanique, ce qui revient à nier la thèse de Turing. Turing, en effet, dans son article de 1937, visait à définir les procédures *finies* ou « calculables par des moyens finis » grâce à sa notion de machine. Parler de procédures finies non mécaniques comme le fait Gödel suppose donc une critique de Turing. Pourquoi, comment alors se référer à cette même thèse de Turing pour généraliser le théorème d'incomplétude ?

Nous nous appuierons sur des notes inédites conservées à la bibliothèque de l'université de Princeton, pour essayer de donner quelques éléments nouveaux sur la relation, ambiguë donc, de Gödel à Turing. Nous n'évoquerons pas les relations de Gödel aux autres acteurs de cette période, Church, Post ou même von Neumann, ce qui nous entraînerait trop loin. Nous ne tenterons pas non plus de mesurer la justesse des arguments que Gödel élabore contre Turing. Nous voulons seulement analyser le développement de la pensée de Gödel sur la question de la calculabilité et le sens de son interprétation de la thèse de Turing.

Sur le problème que nous venons d'évoquer — Gödel utilisant et critiquant à la fois la thèse de Turing — nous voudrions esquisser dès maintenant notre argument. La question est de savoir ce que Gödel accepte dans la thèse de Turing et ce qu'il refuse. Ou, en d'autres termes, ce que, selon Gödel, Turing définit avec sa notion de machine. Comme J. C. Webb et W. Sieg l'ont exactement formulé, à partir de plusieurs passages des textes de Gödel, le logicien considère la thèse de Turing comme une définition de la notion de « procédure mécanique ». Cependant, il reste à savoir ce que Gödel entend par « procédure mécanique » quand il affirme que c'est cette notion que définit Turing. Or nous voudrions montrer, que, avant même la thèse de Turing, Gödel utilise le terme de « procédure mécanique » pour désigner des opérations à l'intérieur d'un système formel, déterminées par des règles qui ne mettent en jeu que la forme des symboles et font abstraction du sens des symboles. Gödel ajoute même, dans des brouillons de 1933 et 1934 (citations (14) et (15) ci-dessous), que l'on pourrait imaginer une machine qui mette en œuvre ces règles. Il ne donne à cette assertion qu'un sens vague. Mais, dans son esprit, la notion de machine de Turing vient lui donner un sens précis. Nous interpréterons en particulier une note (citation (19)), où Gödel définit et oppose les procédures mécaniques et les procédures non mécaniques, comme l'énoncé pour Gödel de la thèse de Turing : les procédures mécaniques, au sens où Gödel l'a toujours entendu, c'est-à-dire les procédures qui ne considèrent que la forme des symboles et font abstraction de leur sens, peuvent être représentées par des machines de Turing. Cette thèse n'est pas celle de Turing, qui entend définir les procédures « finies ». Ce déplacement passe d'abord inaperçu, parce que, jusqu'en 1964, Gödel appelle également « finies » ces procédures « mécaniques ». Il n'en reste pas moins que, pour Gödel, l'objet de la thèse de Turing, ce sont les procédures qui font abstraction du sens des symboles et que l'on appelait déjà, avant Turing, en un sens vague, « mécaniques ». Turing, pour Gödel, a rendu précise cette idée vague. Cette interprétation de Turing a plusieurs conséquences.

D'une part, la thèse de Turing vue par Gödel laisse ouverte la possibilité qu'il existe des procédures non mécaniques, c'est-à-dire qui font référence au sens des symboles, et qui ne soient pas réductibles à des procédures de Turing. Gödel a toujours été convaincu de l'existence de procédures (de suites d'opérations déterminées par des règles univoques) irréductibles à celles que définissent les machines de Turing. Cela est indépendant de son acceptation de la thèse de Turing telle qu'il l'interprète. Néanmoins, cette thèse, qui, pour Gödel, identifie « procédures mécaniques » (qui ne concernent que la forme des symboles) et machines de Turing, implique que, s'il existe des procédures non représentables par des machines de Turing, celles-ci doivent s'appuyer sur le sens des symboles. Le renvoi à la notion de sens (que nous verrons dans de nombreux passages de Gödel) pour la définition d'une procédure non mécanique, n'est qu'une conséquence immédiate de la thèse de Turing vue par Gödel, et non l'indication même vague d'une méthode pour obtenir une procédure non mécanique.

D'autre part, si la thèse de Turing porte, en intention, sur les procédures mécaniques, la question reste de savoir dans quelle mesure des procédures non mécaniques peuvent être dites « finies ». Alors que Gödel énonce de façon explicite l'existence de procédures non mécaniques en 1946, ce n'est que bien après (à partir de l'automne 1963) qu'il critique de façon explicite le texte de Turing pour alors soutenir l'existence de procédures non mécaniques finies. C'est un pas supplémentaire qui ne suit pas forcément du précédent. Il faudra du reste examiner en quel sens ces procédures sont « finies ». Mais on comprend que Gödel puisse critiquer le texte de Turing, en tant que définition des procédures « finies », tout en s'appuyant sur ce qu'il considère comme la « thèse de Turing », la définition des procédures « mécaniques », pour généraliser le théorème de 1931.

Les textes inédits sur lesquels nous appuierons notre analyse nous semblent éclairer certains aspects de la relation de Gödel à Turing. Nous commencerons (dans la section 1) par rappeler le contexte de la thèse de Turing en discutant de la position de Gödel vis-à-vis des définitions antérieures proposées pour la notion de calculabilité. La principale question est de savoir si Gödel a pu, comme Church, puis Turing, énoncer une thèse pour définir la notion de calculabilité. Cette question a été bien entendu abordée depuis le début des années 1980, mais nous présenterons quelques notes inédites (notamment citations (3) et (6)). Nous discuterons ensuite (section 2) de l'interprétation par Gödel de la thèse de Turing comme définition de la notion de procédure mécanique. Enfin, la dernière section concerne l'argument de Gödel en faveur de l'existence de procédures non mécaniques.

1. AVANT TURING : Y A-T-IL UNE THÈSE DE GÖDEL ?

L'histoire de cette période, disons 1926-1937, au cours de laquelle sont données différentes définitions de la calculabilité, est très précisément décrite par M. Davis [1982], S. Kleene [1981 ; 1987] ou W. Sieg [1997]. Nous ne ferons qu'en rappeler certains points et présenter quelques notes de Gödel.

Dans son article de 1931, « Sur les propositions formellement indécidables », Gödel définit les fonctions primitives récursives, qu'il appelle simplement « fonctions récursives ». Il utilise, comme on le fait encore, deux schémas, récursion sur une variable et substitution [Gödel 1931, p. 158–159]. On sait dès 1931 que cette définition ne comprend pas toutes les fonctions calculables. Hilbert, en 1926, donnait déjà l'exemple, qu'il attribue à Ackermann, d'une fonction, définie par une double récurrence, et qui n'est pas « récursive » au sens de Gödel. [1] Le problème de trouver une définition plus large et, si possible, omni-englobante des fonctions calculables, est donc ouvert.

Or l'extension, et la portée, du premier théorème de 1931 dépend d'abord d'une telle définition. On le sait, Gödel code les formules du système, qui représente l'arithmétique élémentaire et dont on veut montrer l'incomplétude, par des numéraux. Une fois ce codage effectué, il faut, au moins, que la classe des axiomes et la relation de conséquence immédiate s'expriment par des relations arithmétiques, des relations qui puissent être définies à l'intérieur du système. On peut considérer que la notion même de système formel exige que celles-ci soient calculables, c'est-à-dire que l'on puisse déterminer par des moyens finis (selon une procédure réglée et en un nombre fini d'étapes) si tel numéral est le code d'un axiome et si, étant donné deux numéraux, l'un est le code d'une formule qui suit immédiatement de l'autre. Maintenant, dans la mesure où l'on ne dispose pas en 1931 d'une définition exhaustive de la

1 [Hilbert 1926], [Ackermann 1928], [Gödel 1934, p. 368].

notion de calculabilité, on ne peut pas absolument exclure qu'il existe des méthodes de calcul effectives qui ne puissent pas s'exprimer dans l'arithmétique élémentaire. Autrement dit, on ne peut pas exclure en 1931 que des systèmes formels, qui permettraient d'exprimer l'arithmétique élémentaire, échappent au théorème d'incomplétude. [2] Au fond, le codage de Gödel met en évidence que la notion même de système formel dépend d'une définition de la calculabilité. Et on ne pourra dire avec rigueur que le théorème d'incomplétude, le premier théorème de 1931, s'applique à tout système formel comprenant l'arithmétique élémentaire, qu'à la condition de disposer d'une définition de la calculabilité, définition qui, *via* le codage de Gödel, fixe la notion même de système formel.

Il faut également parler de la résistance de Gödel à faire jouer le second théorème de 1931 contre le programme de Hilbert. En effet, Gödel établit que la consistance du système ne peut pas être prouvée à l'intérieur du système. Toutes les méthodes finitistes connues s'expriment dans l'arithmétique élémentaire. Mais, contrairement à la plupart de ses contemporains, Gödel refuse de conclure que la consistance de l'arithmétique ne peut pas être prouvée dans l'arithmétique : [3]

(1) « *It is conceivable that there exist finitary proofs that cannot be expressed in the formalism [under consideration]* » [Gödel 1931, p. 195].

En 1933, dans sa conférence « *The present situation in the foundations of mathematics* », Gödel semble prendre une position plus tranchée. Il formule plusieurs contraintes, un système de contraintes qu'il appelle A et que devront vérifier tous raisonnements finitistes. Or

(2) « *[...] all the intuitionistic proofs complying with the requirements of the system A which have ever been constructed can easily be expressed [...] in the system of classical arithmetic, and there are reasons for believing that this will hold for any proof which one will ever be able to construct. So it seems that not even classical arithmetic can be proved noncontradictory by the methods of the system A [...]* » [Gödel 1933, p. 52].

---

2 Gödel [1931, p. 180–181] énonce des conditions suffisantes à l'application du théorème. Celles-ci ne sont pas nécessaires et laissent ouverte la question de savoir à quels systèmes (en dehors des *Principia mathematica*) le théorème d'incomplétude s'applique. Il y a une série d'objections à l'application du premier théorème de 1931 à d'autres systèmes que les *Principia mathematica*, objections qui ne seront clarifiées qu'avec la définition de la calculabilité par Church et Turing. Celle-ci fixe une notion stricte de système formel et permet alors de montrer que le théorème d'incomplétude s'applique à tout système formel en ce sens exprimant l'arithmétique élémentaire. Cf. [Sieg 1994], [Webb 1990] sur les objections de Herbrand au théorème de Gödel, [Herbrand 1931], ainsi que la lettre de Church à J. Dawson [2003, p. 362–262].

3 Sur ce point en particulier, *cf.* les discussions à Vienne et à Göttingen, rapportées dans Mancosu [1999].

Cependant, dans le brouillon, Gödel ajoute immédiatement après cette conclusion une remarque plus prudente, que, il est vrai, il rayera ensuite :

(3) « *The reason why this statement cannot be made with absolute certainty is this : all functions of integer which can be calculated are allowable in the system A but it is impossible to describe all the procedures for the construction of such functions and therefore impossible to give a rigorous proof that all of them are expressible in classical arithmetic, although it can be plausible that nobody will ever be able to construct any such function* » (Brouillon pour Gödel [1933], papiers Gödel 7b, 26, nous soulignons). [4]

Cette note, finalement barrée, rappelle la fin de la lettre de Herbrand à Gödel, d'avril 1931 [Gödel, *Works*, t. V, p. 21]. Elle est, dans le texte de Gödel, difficile à interpréter, et il est vrai que Gödel discute, dans ce passage, du finitisme. [5] Cela dit, dans la phrase que nous soulignons, Gödel évoque les « fonctions entières qui peuvent être calculées » et pose qu'il est « impossible » d'en donner une définition omni-englobante. Cela interdit d'utiliser avec rigueur le deuxième théorème de 1931 contre le programme de Hilbert, comme, du reste, de donner sa pleine généralité au premier théorème : établir l'incomplétude de tout système formel comprenant l'arithmétique élémentaire. Il semble en outre que la conviction qu'il est impossible de définir la calculabilité — cette anti-thèse de Gödel si l'on peut dire — joue encore l'année suivante, en 1934, et, en réalité, jusqu'à la lecture de l'article de Turing.

Gödel passe le printemps 1934 à Princeton où il donne un cours sur le théorème d'incomplétude. Ce cours ne sera publié qu'en 1965, dans l'anthologie de M. Davis, *The undecidable* [Davis 1965]. Les notes prises par Kleene et Rosser, puis revues par Gödel, qui serviront de base à la publication, ont cependant circulé dès l'été 1934. C'est au cours de ce séjour, en février ou mars, que Church communique à Gödel sa définition des fonctions calculables, fondée sur le λ-calcul. Cette définition, que l'on appellera la thèse de Church, n'emporte pas l'adhésion de Gödel, qui la juge « tout à fait non convaincante (*thoroughly unconvincing*) » [Kleene 1987, p. 55]. Dans la dernière section du cours, Gödel donne lui-même une définition des fonctions récursives (générales récursives) à partir d'une suggestion de Herbrand. [6] Gödel, cependant, ne pose pas son énoncé comme une définition de la calculabilité. Il écrira en 1964 à M. Davis :

4 Nous conserverons pour éviter toute ambiguïté l'anglicisme Papiers Gödel dans les références bibliographiques.

5 Sur ce point en particulier, cf. [Sieg 2005, p. 179–180].

6 Cf., en particulier, [Sieg 2005].

(4) « *As far as the second half of section 3 is concerned, it is not true that footnote 3 is a statement of Church's Thesis. The conjecture stated there only refers to the* equivalence *of 'finite (computation) procedure' and « recursive procedure'. However, I was, at the time, not at all convinced that my concept of recursion comprises all possible recursions ; and in fact the equivalence between my definition and Kleene's in Math. Ann 112 [Kleene 1936] is not quite trivial* » [Papiers Gödel, 1b, 138 ; lettre reprise dans [Davis 1965, p. 40]].

La note 3, dans le texte de Gödel, se rapporte à la phrase suivante :

> « *Recursive [primitive recursive] functions have the important property that, for each given set of values of the arguments, the value of the function can be computed by a finite procedure* ».

Gödel note alors :

(5) « *The converse seems to be true if, beside recursions according to scheme (2), recursions of other forms (e.g. with respect to two variables simultaneously) are admitted. This cannot be proved, since the notion of finite computation is not defined, but serves as a heuristic principle* ».

En 1964, pour la reprise dans l'anthologie de M. Davis, Gödel ajoute que « cet énoncé est maintenant hors de propos (*outdated*). Voir le Postscriptum » [Gödel 1934, p. 348].

La lettre à Davis (citation (4)) montre donc que ce principe énoncé dans la note (citation (5)), ou le paragraphe 9 du cours, sur les « fonctions générales récursives », ne devait pas constituer dans l'esprit de Gödel une définition de la calculabilité. Dans la note (citation (5)), Gödel semble alors indiquer de façon informelle que les fonctions calculables sont définies par certaines récursions, sans affirmer que sa définition de la « récursivité générale » comprend toutes les formes de récursions possibles, ni même que cette définition est équivalente à celle de Church, comme Kleene le prouvera en 1935. La construction de la citation (5) est ambiguë, puisque l'on ne sait pas si « sert comme un principe heuristique » concerne la notion de calcul fini ou la proposition énoncée dans la phrase qui précède. Disons donc vaguement que, comme en 1933, Gödel considère la notion de calculabilité comme non définie et objet seulement d'un principe heuristique.

Dans cette perspective, cependant, le brouillon de la lettre à Davis contient un paragraphe surprenant. Gödel écrit :

(6) « *In conclusion of my lectures, I gave expression to (put forth, articulated) the conjecture (not rendered in the above notes) that general recursiveness (as defined in 9) would prove to be an adequate definition of the concept of 'finite procedure' and consequently would make a completely general formulation of the previous results [the incompleteness*

theorems] possible. The definition of the concept of general recursiveness in 9 which was obtained only in the course of these lectures had the aim of making completely general versions of the previous results possible » [Papiers Gödel, 4b, 33, 020634]. [7]

Gödel affirme avoir formulé une thèse similaire à celle de Church, ce qu'il nie dans la lettre qu'il enverra effectivement à M. Davis. Et il est difficile d'accepter cette hypothèse d'une « thèse » de Gödel, puisque, comme on l'a vu, Gödel en 1933 et encore en 1934, ne voit dans la notion de calculabilité que l'objet d'un principe heuristique, et non celui d'une définition. Quel sens alors donner à cette remarque dans le brouillon de la lettre à M. Davis ? C'est ou bien que, rétrospectivement, Gödel n'est pas certain de la portée qu'il attribuait en 1934 à sa définition des « fonctions générales récursives », ou bien que le logicien ne veut pas laisser à Church seul le mérite d'une définition de la calculabilité. [8]

Une autre lettre à M. Davis, un peu plus tardive, revient du reste sur ce point. Gödel affirme à nouveau avoir énoncé une thèse équivalente à celle de Church, en se référant non plus à son cours de 1934, mais à un article de 1936.

(7) « *Moreover it occurred to me that, in case you mention explicitly that in these lectures I did not state an equivalent of Church's Thesis, you should also mention that I did so (for computation procedures) in the last paragraph of my paper on the length of proof* » [Papiers Gödel, 1b, 38].

Cet article [Gödel 1936] reprend de celui de 1931 la notion de fonction calculable dans un système formel S. Une fonction $\phi(x)$ est calculable dans S si pour tout numéral m il existe un numéral n tel que $\phi(m) = n$ est prouvable dans S. À partir de l'arithmétique élémentaire (au premier ordre, S_1), Gödel considère une hiérarchie de systèmes S_i à l'ordre i. Il remarque qu'une fonction calculable dans l'un de ces systèmes est déjà calculable dans S_1. Ainsi, cette « notion de 'calculable' est en un certain sens absolue » [Gödel 1936, p. 399]. Gödel fera une remarque analogue à propos de la notion de « récursivité générale (ou calculabilité de Turing) :

7 Ce brouillon ne se trouve pas dans le dossier de la correspondance avec M. Davis, mais dans celui de la correspondance avec l'éditeur de [Davis 1965].

8 Gödel indique aussi avoir obtenu sa définition durant son séjour à Princeton. Les Archives comprennent différents brouillons pour différentes parties du cours mais (à ma connaissance) aucun ne concerne la définition de la récursivité générale si ce n'est un long texte (l'item 040117) qui présente l'ensemble du cours et pourrait du reste avoir été utilisé comme notes en classe. Dans ce texte, Gödel définit les fonctions générales récursives en des termes proches (mais non identiques) aux notes qui seront ensuite distribuées.

« Par une sorte de miracle », « ces concepts ne dépendent pas du formalisme choisi » [Gödel 1946, p. 150].

Cependant, il n'est pas clair que la note de 1936, qui remarque le caractère « absolu » de la notion de « fonction calculable dans un système formel », c'est-à-dire le fait que cette notion ne s'élargit pas à mesure que l'on s'élève dans la hiérarchie des systèmes S_i, puisse avoir constituée dans l'esprit de Gödel à l'époque une thèse analogue à celle de Church. La notion de « fonction calculable dans un système formel (comprenant l'arithmétique élémentaire) » s'avèrera sans doute équivalente à la calculabilité au sens de Church, de Turing ou de Gödel [1934]. Cependant, il faut prendre en compte que, jusqu'à la lecture de Turing, Gödel n'a sans doute jamais été convaincu que sa définition, ou celle de Church, ait inclus la totalité des fonctions calculables. Plus précisément, dans la citation (3), Gödel envisage l'existence de fonctions calculables qui ne s'exprimeraient pas dans l'arithmétique élémentaire. Or, dans ce cas, l'absoluité de la notion de « fonction calculable dans un système formel », le fait que cette notion ne s'élargisse pas lorsque l'on monte dans la hiérarchie des systèmes, ne signifie pas qu'elle comprenne la totalité des fonctions calculables. [9]

Il faut donc prendre avec précaution ces passages où Gödel affirme avoir énoncé une thèse comparable à celle de Church. L'ambiguïté peut venir de ce que, pour Gödel, la thèse de Church n'est pas convaincante par elle-même. Et, dans ce cas, le travail de Church a surtout consisté à donner une définition équivalente à celle de Turing et que celle-ci justifie alors. Or, une telle définition, Gödel en a également donné une, et à peu près au même moment. Personne avant Turing ne semble, dans l'esprit de Gödel, avoir pu établir l'adéquation des définitions proposées.

Ainsi, dans un texte non daté, mais vraisemblablement de la fin des années trente, Gödel ajoute après avoir présenté la définition des fonctions récursives de 1934, qu'il attribue du reste seulement à Herbrand :

(8-bis) « *That this* [cette définition] *really is the correct definition of computability was established beyond any doubt by Turing* » [Gödel 193?, p. 168].

Wang écrit également, avec la permission de Gödel, qui a souvent lui-même rédigé et, en tout cas, relu et corrigé ces passages :

---

9 Sieg [1994 ; 2006] met également en question cette « thèse » de Gödel dans l'article de 1936, en soulignant que la notion de « fonction calculable dans un système formel » dépend de celle de système formel, que Gödel ne considèrera pas comme clairement fixée avant les travaux de Turing.

(8) « *Hence we have an excellent example here of a concept which did not appear sharp to us but has become so as a result of a careful reflection. The resulting definition of the concept of mechanical procedure by the sharp concept of 'performable' by a Turing machine is both correct and unique. [...] Gödel emphasizes that there is at least one highly interesting concept which is made precise by the unqualified notion of a Turing machine. Namely a formal system is nothing but a mechanical procedure for producing theorems. The concept of formal system requires that reasoning be completely replaced by 'mechanical operations' on formulas in just the sense made clear by Turing machines. [...] In fact, the concept of formal systems was not clear at all in 1931. Otherwise, Gödel would have then proved his incompleteness theorem in a more general form* » [Wang 1974, p. 84, nous soulignons].

Ces deux dernières phrases pourraient également décrire la situation en 1934. Contrairement à ce qu'il indique dans le brouillon de la lettre à M. Davis (citation (6)), Gödel n'applique pas sa définition des fonctions « générales récursives » à la notion de système formel et à la généralisation du théorème de 1931. Cette définition n'apparaît que dans le dernier paragraphe du cours de 1934, comme un appendice. Les conditions pour l'application du théorème d'incomplétude (dans la section 6 du cours) sont équivalentes à celle de 1931. C'est seulement Turing qui, pour Gödel, donne une définition convaincante, adéquate au sens de la notion. Gödel reconnaît dans des notes ajoutées en 1963 à l'article de 1931 et en 1964 au cours de 1934 :

(9) « *In consequence of later advances, in particular of the fact that, due to A. M. Turing's work, a precise and unquestionably adequate definition of the general concept of formal system can now be given, the existence of undecidable arithmetical propositions [...] can now be proved rigorously for* every *consistent formal system containing a certain amount of finitary number theory* » [Gödel 1934, Postscriptum 1964, p. 369].[10]

Et, comme le souligne W. Sieg [2006], c'est une lettre à Nagel qui est sur ce point la plus claire :

(10) « *It was only by Turing's work that it became completely clear, that my proof is applicable to every formal system containing arithmetic* » [Gödel à Nagel, 14 mars 1957, in [Gödel, *Works*, t. V, p. 147], Gödel souligne].

2. AVEC TURING : LES PROCÉDURES MÉCANIQUES

Le paragraphe de Wang (citation (8)) illustre plusieurs points, décisifs pour l'interprétation gödelienne de la thèse de Turing : (i) la notion en question sous le terme de « calculabilité » est celle de procédure mécanique, ou de système formel ; (ii) cette notion joue déjà un rôle en logique

10 Cf. également [Gödel 1931, note de 1963, p. 195].

avant Turing, sous la forme d'un concept qui n'est perçu que de façon confuse ; (iii) c'est bien Turing qui résout le problème de la calculabilité, parce que sa définition est adéquate à la notion en question.

Nous développerons simplement ces trois points. La notion à laquelle se rapporte la thèse de Turing, ou que les machines de Turing permettent de définir et à laquelle elles donnent une forme mathématiquement utilisable, est celle de « procédure mécanique ». Cette notion est au fondement de celle de système formel. Il y a en effet de nombreux passages où Gödel décrit la déduction de théorèmes au sein d'un système formel comme une procédure mécanique.

(11) « *'finite procedure' should mean 'mechanical procedure'. This meaning, however, is required by the concept of formal system, whose essence it is that reasoning is completely replaced by mechanical operations on formulas* » [Gödel 1934, Postscriptum 1964, p. 369].

Ou encore

(12) « *Formal systems in the proper sense of the term, whose characteristic property is that reasoning in them, in principle, can be completely replaced by mechanical devices [...]* » [Gödel 1931, note de 1963, p. 195].

Le point crucial est que Gödel lui-même énonce explicitement cette clause, que le raisonnement formel est mécanique, avant l'article de Turing. Le terme « mécanique », ou la référence à une machine, à propos du raisonnement logique, apparaît chez Frege et dans l'article de von Neumann de 1927. De même, Gödel dans son cours de 1934 note :

(13) « *a formal system consists only of symbols and mechanical rules relating to them* » [Gödel 1934, p. 349].

Mais, nulle part avant l'article de Turing, cette référence au mécanique n'est aussi claire que dans la conférence de 1933 et son brouillon. Gödel définit alors l'idée de formalisme de la façon suivante :

(14) « *The outstanding feature of the rules of inference being that they are purely formal, i.e. refer only to the outward structure of the formulas, not to their meaning, so that they could be applied by someone who knew nothing about mathematics, or by a machine* » [Gödel 1933, p. 45].

Ce passage est développé dans le brouillon. Gödel ajoute :

(15) « *[The rules of inference] could be applied by someone who knew nothing about the meaning of the symbols. One could easily devise a machine which would give you as many correct consequences of the axioms as you like, the only trouble would be that it would give*

consequences at random and therefore not the results one is interested in » [Papiers Gödel, 7b, 26, 040109, brouillon pour [Gödel 1933], nous soulignons]. [11]

La même phrase apparaît dans le texte d'une conférence à la *Philosophical Society of New York University*. Gödel définit donc la notion de formalisme par l'idée que les symboles y sont considérés abstraction faite de leur sens, et cela revient à dire qu'ils sont utilisés « mécaniquement ». En fait, avant Turing mais confusément, Gödel imagine une machine capable de déduire les (ou peut-être seulement des) théorèmes qui suivent des axiomes. Cependant, de la même façon qu'il voit dans la notion de calculabilité un principe heuristique, qui ne fait pas l'objet d'une définition, Gödel ne se rend pas compte que l'on peut décrire de façon précise ces machines à déduire. Et c'est ce que fera Turing. On peut considérer le passage cité du livre de Wang (citation (8)), pour autant qu'il a été lu et approuvé par Gödel, comme autobiographique. Le travail de Turing a été de préciser et de rendre mathématiquement utile une notion à laquelle Gödel lui-même renvoyait mais qu'il n'apercevait que confusément.

C'est pourquoi également, après l'article de Turing, Gödel, dans ses conférences « populaires », ne modifie pratiquement pas l'explication donnée pour la notion de système formel. Tantôt, le logicien ne fait que spécifier la machine, à laquelle il se référait déjà en 1933, comme une machine de Turing. Ainsi, la conférence « Gibbs » de 1951 évoque :

(16) « *a well-defined system of axioms and rules [...]. This requirement for the rules and axioms is equivalent to the requirement that it should be possible to build a finite machine, in the precise sense of a 'Turing machine', which will write down all the consequences of the axioms one after the other* » [Gödel 1931, p. 308].

Un passage du cours de logique donné à l'université de Notre-Dame en 1939 ne mentionne du reste pas même Turing. C'est que, au fond, Gödel ne fait que préciser cette machine dont il avait déjà eu l'idée six ans auparavant :

(17) « *It is the chief purpose of the axiomatisation of logic to avoid reference to the meaning of formulas i.e. we want to set up a calculus which can be handled purely mechanically (i.e. a calculus which makes thinking superfluous and which can replace thinking for certain questions). In other words, we want to put into effect as far as possible Leibniz' program of a calculus ratiocinator [...] This program has been partly carried out by the axiomatic system for logic. [...] It would actually be possible to construct a machine which would do the following thing : the supposed machine is to have a crank and whenever you turn the crank once the machine would write down a tautology of the calculus of predicates. And*

11 Également, 7b, 30, item 040123, conférence à la *Philosophical Society of New York University*, 18 avril 1934.

it would write down every existing tautologies of the calculus of predicates if you turn the crank sufficiently often. So the machine would really replace thinking completely as far as deriving formulas of the calculus of predicates is concerned » [Papiers Gödel, 8a, 59, item 040210]. [12]

On comprend donc mieux pourquoi, à la différence de la thèse de Church, celle de Turing, avec la référence à la machine, a de quoi convaincre Gödel. C'est qu'elle exprime sous une forme rigoureuse, une idée confuse dans la logique antérieure et dans les textes de Gödel eux-mêmes. Elle est adéquate au contenu de la notion en question, comme du reste la formulation de Post, au moyen de son « travailleur ». Elle est « intensionnellement adéquate » :

(18) « *The precise definition of the concept of formal system is achieved by means of the general concept of mechanical (or computational procedure or algorithm) which, no doubt, is adequately explained by the concept of a Turing machine. Extensionally equivalent concepts of computability had been defined before Turing's work and their adequacy had been highly probable (see : Church, 1936). Turing's definition was given simultaneously by E. Post in the* Journal of symbolic logic. *But only Turing has really analysed the content of computation and has definitively established the adequacy of the definitions. [...] Only Turing's and Post's definitions are also intensionally adequate and thereby settle the question definitively* » [Papiers Gödel, 8c, 106, item 040332]. [13]

Revenons en arrière. Dès la conférence de 1933, et à l'instar d'une certaine tradition logique, Gödel caractérise la notion de système formel par ceci que les symboles y sont considérés abstraction faite de leur sens, et qualifie alors le raisonnement de « mécanique ». C'est, pour Gödel, cette idée de procédure mécanique que vient fixer la thèse de Turing. J'ai déjà mentionné cette note où Gödel oppose les procédures non mécaniques (dont on discutera dans la section suivante) et les procédures mécaniques pour, manifestement, définir les unes et les autres :

---

12 Cours de logique à Notre Dame, 1939. Cité en partie dans [Sieg 2006].

13 Brouillon pour le Postcriptum à [Gödel 1934] Gödel souligne et se réfère aux articles de Church, « *An unsolvable problem of elementary number theory* » reproduit dans [Davis 1965], et de Post, « *Finite combinatory processes. Formulation I* », reproduit dans [Davis 1965]).
Dans un autre texte, un brouillon pour la révision de 1972, Gödel insiste aussi sur la forme mathématique que prend avec Turing l'idée de procédure mécanique : « *Turing has given a much more elaborate definition of the [...] concept of mechanically computable number theoretic function, and it was exactly this greater detail which apparently made this concept precisely intelligible and thereby accessible to mathematical treatment* » [Papiers Gödel, 9b, 141, 040450].

(19) « *Präzise Formulierung 'non mechanical procedure' : procedure into which some meaning essentially enters (as an element which cannot be eliminated and which is successful for all problems of a class for which no Turing procedure exists).*

[A] mechanical procedure of [typing ?] formulas composed of finite number of letters is a procedure which, given any desired amount of paper for typing but nothing else outside the machine, can be carried out by a machine with a finite number of parts each of which, as far as they are relevant for the work of the machine, has only a finite number of different possible states or positions » [Papiers Gödel 8c, 106, 040332].

Nous discuterons plus loin de la première partie de cette note. Nous nous attachons pour l'instant à la deuxième partie sur les procédures mécaniques. Gödel y décrit une machine finie que, puisque le nom de Turing est mentionné dans la première partie dela note, on peut considérer comme une machine de Turing. Le passage revient alors à ceci : une procédure mécanique est une procédure de Turing. Or, si cette phrase ne doit pas être une tautologie (une simple identité, a=a), il faut prendre « procédure mécanique » au sens où Gödel l'entendait avant Turing. Les citations (11), (12), (13), (14), (15), (17) semblent poser, comme une définition du terme « mécanique » :

Procédure mécanique $\underset{\text{DEF}}{=}$ *Procédure qui ne repose que sur la forme des symboles et fait abstraction de leur sens.*

La citation (19) établit alors :

(Thèse de Turing vue par Gödel)
Procédure mécanique = *Procédure réalisable sur une machine de Turing*

Bref, une procédure qui ne repose que sur la forme des symboles dans l'abstraction de leur sens est représentable par une machine de Turing. En outre, comme le montre la citation (18), avec la référence à une « adéquation intensionnelle », cette égalité ne concerne pas l'extension des concepts mais les concepts eux-mêmes : le concept de machine de Turing (ou réalisable par une machine de Turing) est identique à celui de procédure mécanique. C'est, pour Gödel, le même concept d'abord vu de façon confuse puis de façon précise.

La citation (19) a le mérite d'énoncer clairement une thèse, identifiant deux termes, procédures mécaniques et machines de Turing. Elle soulève plusieurs difficultés. D'abord, cette thèse n'est pas celle que défend Turing dans son article de 1937. Prenons-le à la lettre. Turing [Turing 1937, p. 116] entend définir les « procédures finies » ou « calculables par des moyens finis », et non les procédures « mécaniques ». La citation (19) présente donc seulement la thèse de Turing vue par Gödel. Au regard de la thèse explicite que défend Turing, « mécaniques » a été substitué à « finies ». Cette première remarque pose la question de l'usage par Gödel lui-même du terme « fini » ou « procédure finie ».

Il faut dès maintenant noter un changement important, sur lequel nous insisterons encore dans la dernière partie de cet article. L'année 1964 marque un tournant. Jusque là, Gödel qualifie également de « finies » ces « procédures mécaniques ». On le voit notamment dans les citations (4), (5), (6). Il brouille alors le sens de la thèse qu'il tire de l'article de Turing. La thèse de Turing, pour Gödel, a toujours porté sur les « procédures mécaniques », c'est-à-dire les procédures qui ne font pas référence au sens des symboles. Cependant, ces « procédures mécaniques », Gödel, jusqu'en 1964, les dit également « finies ». C'est seulement à partir de 1964 (et avec un certain flottement pour l'année 1964) que Gödel distingue l'usage des deux termes « mécanique » et « fini », ce qui rend alors claire la thèse qu'il lit chez Turing : les procédures « mécaniques » sont définies par les machines de Turing, alors qu'il existe des procédures « finies » non représentables par des machines de Turing. Nous reviendrons sur ce deuxième point.

Si notre interprétation des textes est correcte, Gödel lit dans l'article de Turing la thèse que les procédures « mécaniques », qui ne font pas référence au sens des symboles, sont représentables par des machines de Turing, certains dispositifs possédant des caractéristiques que fixe Turing. Comment, pour Gödel, cette thèse est-elle justifiée ? Comment montrer que toute procédure que peut réaliser un calculateur humain sans référence au sens des symboles, peut également être produite sur une machine de Turing ? Pour reprendre en français une distinction souvent formulée en anglais, il y a le « calculator » (*computor*) humain et « le calculateur » (*computer*) de Turing. La thèse de Turing identifie les procédures finies du « calculator » avec celles du « calculateur ». Cette identification, Turing vise à la justifier par une série de trois arguments. Gödel déplace la thèse de Turing, en la faisant porter non plus sur les procédures « finies » mais « mécaniques ». Cependant, il lui resterait à justifier cette nouvelle thèse (d'autant plus que, comme on le verra, Gödel critique l'un des arguments que Turing utilise pour établir sa propre thèse). Or il n'y a nulle part de textes où Gödel argumente de façon détaillée l'identification des procédures du « calculator », qui fait abstraction du sens, et celles du calculateur de Turing. Va-t-il de soi que toutes les procédures que l'on peut effectuer en manipulant des symboles sans considération de leur sens peuvent être réalisées par une machine de Turing ? Si cela ne va pas de soi, pourquoi Gödel ne s'attache-t-il jamais à discuter de cette question ?

En particulier, l'article de 1958, repris en 1972, « *On an extension of finitary mathematics...* », traduit l'arithmétique classique dans un certain système dont les objets sont des fonctions calculables (une hiérarchie de fonctions calculables). Puisque cette notion doit alors donner un fondement à l'arithmétique, on pourrait s'attendre à ce que Gödel en explicite le sens intuitif. Or, au contraire, Gödel pose simplement :

(20) « *The phrase 'well-defined mathematical procedure' is to be accepted as having a clear meaning without any further explanation* » [Gödel 1972a, p. 275]. [14]

Le logicien commente en note, dans la version de 1958 :

(21) « *As is well known, A. M. Turing, using the notion of a computing machine, gave a definition of the notion of computable function of the first order. But, had this notion not already been intelligible, the question whether Turing's definition is adequate would be meaningless* » [Gödel 1958, p. 245].

Et la version de 1972 précise :

(22) « *[...] However, if the term 'mechanically computable' had not had a clear, although unanalysed, meaning before, the question as to whether Turing's definition is adequate would be meaningless, while it undoubtedly has an affirmative answer* » [Gödel 1972a, p. 275].

On voit donc à nouveau que, dans l'esprit de Gödel, Turing a seulement transformé une notion claire mais confuse en une notion claire et distincte. La notion de procédure mécanique préexiste au travail de Turing, qui la fixe et lui donne un contenu mathématique. Et c'est à son adéquation à cette notion préexistante que la portée de la définition de Turing doit être mesurée. Mais cela ne se comprend que dans la mesure où la notion en question ne préexiste pas seulement comme une idée confuse dans la logique avant Turing mais comme un concept en soi, une idée platonicienne et que, selon le mot de Gödel [Gödel 1951, p. 311–312], l'ego n'a pas créée à partir de rien. C'est ce qu'un autre texte dans le livre de Wang montre clairement :

(23) « *'If we begin with a vague intuitive concept, how can we find a sharp concept to correspond to it faithfully ?' The answer Gödel gives is that the sharp concept is there all along,*

---

14 Une première version, [Gödel 1941], définissait ces fonctions par certains schémas, ce qui obligeait ensuite à prouver que les fonctions ainsi définies vérifiaient les axiomes du système. Cette preuve utilise une induction sur les segment $[0, \epsilon_0]$, de sorte que ce fondement de l'arithmétique repose sur une induction transfinie, sans qu'il y ait gain par rapport à la démonstration de consistance que donnait Gentzen en 1935. Dans les versions de 1958 et de 1972, Gödel considère plutôt ces fonctions calculables comme une notion primitive, « immédiatement intelligible », ce qui le dispense de recourir à l'induction précédente.

only we did not perceive it clearly at first. This is similar to our perception of an animal first far away and then nearby. We had not perceived the sharp concept of mechanical procedure sharply before Turing, who brought us to the right perspective. And then we do perceive the sharp concept. There are more similarities than differences between sense perceptions and the perception of concepts » [Wang 1974, p. 84-85].

De façon explicite dans ce texte, Gödel considère le concept de « procédure mécanique », ou de machine de Turing, comme un objet en soi qu'il compare aux objets perceptifs et décrit donc dans les mêmes termes que le concept d'ensemble. Ces concepts nous sont donc d'abord donnés dans une intuition avant d'être fixés dans certains énoncés. La thèse de Turing, dans l'interprétation de Gödel, fixe le concept de « procédure mécanique » comme les axiomes de la théorie des ensembles fixent le concept d'ensemble. La justification alors de la thèse de Turing semble tenir à l'intuition et à l'évidence. C'est pourquoi, semble-t-il, Gödel ne la discute pas. L'évidence la montre « indubitablement » adéquate (citation (22)). La thèse de Turing est un énoncé analogue à ces axiomes qui « s'imposent à nous comme étant vrais » [Gödel 1964, p. 268]. Cet appel à l'intuition constitue un argument assez faible mais c'est, à notre connaissance, la seule justification de la thèse que Gödel tire de Turing.

3. APRÈS TURING : LES PROCÉDURES NON MÉCANIQUES

On l'a vu, Gödel interprète la thèse de Turing comme une définition de la notion de procédure mécanique et, par conséquent, de celle de système formel. Mais il s'agit d'un déplacement par rapport à la portée que Turing veut donner à sa thèse. Turing [Turing 1937, p. 116, déjà cité] parle de calcul « par des moyens finis » (« *by finite means* »). Comme le souligne J. C. Webb, Gödel indique explicitement, dans le Postscriptum ajouté en 1964 au cours de 1934, qu'il n'accepte la thèse de Turing que dans la mesure où

(24) « *'finite procedure' is understood to mean 'mechanical procedures'* » [Gödel 1934, p. 370].[15]

Cet énoncé, il faut toutefois le situer dans son contexte, pour comprendre ce que signifie l'expression de « procédure mécanique ». On voit alors bien que Gödel opère un déplacement par rapport à la thèse de Turing : des procédures finies aux procédures mécaniques. Gödel défend une autre thèse, qui ne recouvre pas la thèse de Turing. En particulier, dans ce Postscriptum de 1964, Gödel peut donc évoquer l'existence de

15 Cf. la discussion de Webb [Webb 1990, p. 296 et suivantes].

« procédures finies non mécaniques », qui ne tombent pas dans le cadre de la thèse de Turing telle qu'il l'interprète, sans remettre en question l'adéquation de la thèse de Turing comme définition des procédures mécaniques. Il y a là — soutient Gödel — deux questions « indépendantes » :

(25) « *Note that the question of whether there exist finite non-mechanical procedures, not equivalent with any algorithm, has nothing whatsoever to do with the adequacy of the definition of 'formal system' and of 'mechanical procedures'* » [Gödel 1934, p. 370].

Ainsi, dans l'esprit de Gödel, Turing a correctement défini un certain genre de procédures, les procédures mécaniques (par exemple, les procédures impliquées dans la dérivation des théorèmes d'un système formel), mais il existe d'autres procédures, réglées, univoques, qui ne sont pas mécaniques et n'entrent pas dans la définition de Turing. C'est de ce problème, de l'existence de procédures non mécaniques finies, qu'il faut maintenant discuter.

Gödel, dans le Postscriptum de 1964, soulève deux questions à la fois. Il y a la question de savoir si l'on peut envisager des procédures non mécaniques. Et il y a la question de savoir dans quelle mesure ces procédures peuvent être dites « finies ». Or Gödel a toujours défendu l'existence de procédures non mécaniques, mais leur qualification comme finies ne date que de 1964. Du reste, le brouillon du Postscriptum énonçait seulement à la place de la citation (25)

(26) « *It is possible that there exist non-mechanical procedures non replaceable by Turing machines* » [Papiers Gödel, 8c, 106, 040332].

Le caractère fini de ces procédures n'y est pas mentionné. De la même manière, l'article de 1958 attribue à Turing la définition des « fonctions calculables » (*computable functions*), alors que la version révisée en 1972 ne lui attribue plus que la définition de « fonctions mécaniquement calculables » (*mechanically computable functions*) [Gödel 1972a, p. 275]. Et, dans de nombreux brouillons, Gödel évoque des « calculs par des procédures de pensée » (*computations by thought procedures*) et soutient :

(27) « *There are no sufficient reasons for expecting these two concepts [« computation by thought procedures » et « mechanical computation » dans les mots de Gödel] to have the same extension* » [Papiers Gödel, 9b, 141, 040450].

Ainsi, c'est seulement après 1964 (et non sans réserve [16]) que Gödel qualifie de « finies » ou de « calculs » (*computations*) des procédures qui

---

16 Gödel ajoute au passage précédent : « *Also it may be questioned that computation would be a suitable name for such procedures* » [Papiers Gödel, 9b, 141, 040450].

débordent l'analyse de Turing. Cependant, Gödel a toujours cru à la possibilité de procédures réglées non mécaniques. La célèbre note à l'article 1931 suggère déjà :

(28) « *The true reason for the incompleteness inherent in all formal systems of mathematics is that the formation of ever higher types can be continued into the transfinite (see Hilbert [Hilbert 1926]), while in any formal system at most denumerably many of them are available* ». [17]

Or, puisque Gödel, du moins en 1933, décrit le raisonnement à l'intérieur d'un système formel comme une procédure « mécanique », la possibilité de dépasser (par la formation de types transfinis et, apparemment, au-delà du dénombrable) le formalisme semble alors impliquer la reconnaissance de raisonnements qui ne pourront plus être dits « mécaniques ». Le terme « mécanique » serait alors pris en un sens pré-turingien. De façon explicite et après l'article de Turing, les « Remarques » de Gödel à Princeton en 1946 évoquent une procédure pour l'extension d'un système formel par l'adjonction de nouveaux axiomes, « qui pourraient être décrits et rassemblés de façon non constructive (*could be described and collected together in some non-constructive way*) et pour lesquels « vaudrait un résultat de complétude » [Gödel 1946, p. 151]. [18] Si cette procédure n'est pas décrite avec précision, elle supposerait manifestement des règles permettant d'obtenir à chaque étape un nouvel axiome. Et, si le système ainsi constitué doit pouvoir être complet, chaque proposition formulable pouvant y être ou démontrée ou réfutée (et non les deux à la fois), il faut qu'elle ne se réduise pas à une procédure mécanique, représentable par une machine de Turing. Il faut donc admettre l'existence de procédures réglées, bien définies et, pourtant, non mécaniques.

Ainsi, la nouveauté du Postscriptum de 1964 n'est pas dans la mention de procédures non mécaniques mais dans leur qualification comme « finies ». Cela donne un point d'appui pour analyser la position de Gödel, telle qu'elle est décrite dans la citation (25). En effet, soutenir l'existence de procédures finies non mécaniques semble engager à une critique de l'article de Turing, qui, pris à la lettre, entend définir par l'idée de machine l'ensemble des procédures finies : les calculs « par des moyens finis ».

Gödel développe en effet un argument contre Turing. Une première version en a été retrouvée dans les épreuves corrigées par Gödel de son article pour *Dialectica*, de 1972. Une seconde version a été publiée par Wang,

17 Gödel [1931, p. 181] se réfère à l'article de Hilbert, « Sur l'infini » [Hilbert 1926].

18 Sieg [2006] discute également des remarques de Gödel en les confrontant à l'intervention préalable de Tarski publiée par Sinaceur [2000].

dans son livre de 1974, *From Mathematics to Philosophy*. Nous nous attacherons surtout à la deuxième version qui est plus claire :

(29) « *Turing, in* Proc. Lond. Math. Soc. *42 (1936) p. 250, gives an argument which is supposed to show that mental procedures cannot carry any further than mechanical procedures. However, this argument is inconclusive, because it depends on the supposition that a finite mind is capable of only a finite number of distinguishable states. What Turing disregards completely is the fact that* mind*, in its use, is not static but constantly developing. This is seen, e.g., from the infinite series of ever stronger axioms of infinity in set theory, each of which expresses a new idea or insight. A similar process takes place with regard to the primitive terms. E.g., the iterative concept of set became clear only in the past few decades. Several more primitive ideas now appear on the horizon, e.g., the selfreflexive concept of proper class. Therefore, although at each stage of the mind's development the number of its possible states is finite, there is no reason why this number should not converge to infinity in the course of its development. Now there may exist systematic methods of accelerating, specializing, and uniquely determining this development, e.g. by asking the right questions on the basis of a mechanical procedure. But it must be admitted that the precise definition of a procedure of this kind would require a substantial deepening of our understanding of the basic operations of the mind* » [Gödel, *Works*, t. V, p. 576. Gödel souligne]. [19]

Une première difficulté concerne la date de cet argument. J. Dawson [Dawson 1997, p. 232], s'appuyant sur le journal de Morgenstern, qui rapporte une conversation avec Gödel à ce sujet, propose décembre 1969. Nous conjecturons pourtant que cet argument est antérieur. Nous avons pour cela plusieurs éléments. Il y a des remarques qui sous-entendent cet argument et ne prennent sens que sur cette base dans les brouillons pour le Postscriptum de 1964 (par exemple, citation (31) ci-dessous). Il y a également un papier daté « septembre-décembre 1963 » qui isole la citation de Turing (citation (30) ci-dessous), dont part l'argument de Gödel, et l'entoure d'une multitude de flèches, ce qui montre bien que, dans l'esprit de Gödel, c'est déjà là le point faible de la démonstration qu'il prête à Turing. [20] C'est donc vraisemblablement en préparant le Postscriptum pour la réédition de ses cours de 1934 que Gödel a mis au point cet argument, qui, dans son esprit, établit contre Turing la possibilité de ces « procédures finies non mécaniques ».

Gödel se réfère, dans l'article de Turing, de 1937, « Sur les nombres calculables » au paragraphe 9 où Turing développe trois arguments visant

19 La première version est publiée dans [Gödel 1972b].

20 Il s'agit de l'item 060164, boîte 11b, dossier 15, intitulé par Gödel « *Aufsatz im 'Entschluss', Sep-Dec 1963* » La seconde moitié de la même feuille comporte une série de commentaires en Gablesberger, que C. Dawson a transcrits, mais qui concernent la théologie comme science exacte, sans rapport avec Turing ou le problème de la calculabilité.

à établir que les machines définies précédemment, les « machines de Turing » permettent en effet de représenter toutes les fonctions calculables. Ce sont ces arguments qui justifient la thèse que les fonctions calculables sont les fonctions calculables par une machine de Turing. Dans son premier, et principal, argument, Turing analyse le calculateur humain (le calculator) pour montrer que chacune de ses opérations peut être reproduite par une machine de Turing. Mais, pour maintenir l'analogie entre le calculateur humain et les machines à calculer, Turing doit encore poser que :

(30) « *the number of states of mind which need be taken into account is finite. [...] If we admitted an infinity of states of mind, some of them will be 'arbitrarily close' and will be confused* » [Turing 1937, p. 136].

Les états possibles d'une machine de Turing sont, par définition, en nombre fini. La possibilité de considérer le calculateur humain comme une telle machine exige d'admettre que le nombre des états par lesquels le calculateur humain peut passer au cours d'un « calcul », d'une procédure donnée est également fini. Cela revient à limiter la mémoire du calculateur humain. Les états internes jouent en effet le rôle d'une mémoire. [21]

Or c'est d'abord ce point, cette limitation du nombre des états possibles, de la mémoire, du calculateur que refuse Gödel (en particulier citation (28)). On peut concevoir, soutient Gödel, que, à chaque étape de la procédure, le nombre des états par lesquels le calculateur est déjà passé reste fini mais que la procédure dans son ensemble mette en jeu une infinité d'états. Nous pouvons donc parler d'une infinité potentielle de l'esprit dont la mémoire (les états possibles) est toujours finie mais peut s'enrichir au-delà de toute borne. Le premier point, pour Gödel, est de montrer que cette infinité potentielle est concevable : elle n'est pas un non-sens. Néanmoins, la possibilité d'attribuer à un esprit « fini » cette « infinité » potentielle dépend du cadre métaphysique dans lequel on se place. La critique de Turing exige un détour par la métaphysique :

(31) « *A briefer way of describing the situation is this : it is perfectly possible that we are capable of infinitely many well distinguishable mental states although at each moment only a finite number can have been actualised. In fact, [that] contradicts the 'finitude' of the human mind as little as would eternal life. The latter also presupposes the possibility of*

21 Notamment [Kleene 1988, p. 22] et [Sieg 1994, p. 93]. Parce qu'il est dans un état, ou dans un autre, le calculateur, dans les mêmes circonstances, c'est-à-dire devant le même symbole sur le ruban, accomplira telle action, ou telle autre. En cela, l'action du calculateur peut dépendre du passé, des actions et des circonstances antérieures. La référence à ces états internes permet donc d'introduire une sorte de mémoire dans le calcul.

infinitely many well distinguishable experiences in one finite human being. It is only a materialist prejudice which excludes this because of the finiteness of our head. Moreover in order to prove that the procedure actually can be carried out a much more profound understanding of the essence and the working of the human mind » [... Le texte s'interrompt ici. Papiers Gödel, 8c, 106, 040332].

L'exemple de la vie éternelle vient seulement montrer qu'il est possible en principe, qu'il est sensé, de donner à un 'esprit fini' une infinité potentielle. Cela, continue Gödel, ne devient absurde que dans le cadre d'un matérialisme qui limiterait la capacité de l'esprit en référence au cerveau. Ce point est développé dans la suite du texte que Gödel a donné à Wang, pour son livre de 1974.

(32) « *In our discussions, Gödel added the following. Turing's argument becomes valid under two additional assumptions, which today are generally accepted, namely : 1. There is no mind separate from matter. 2. The brain functions basically like a digital computer (2. may be replaced by : 2′ The physical laws, in their observable consequences, have a finite limit of precision.) However, while Gödel thinks that 2 is very likely and 2′ practically certain, he believes that 1. is a prejudice of our time, which will be disproved scientifically (perhaps by the fact that there aren't enough nerve cells to perform the observable operations of the mind)* » [Wang 1974, p. 326]. [22]

La capacité, la mémoire (le nombre des états possibles) de l'esprit humain est-elle bornée ? Cette remarque de Turing, que le nombre des états possibles du calculateur humain est fini, dépend selon Gödel d'hypothèses « matérialistes ». C'est, du reste, manifeste dans l'article de 1937, où Turing fait immédiatement suivre sa remarque sur les « états d'esprit » du calculateur par une discussion concernant leurs « contre-parties physiques ». De façon générale, ce « matérialisme », qui, pour Gödel, vient parasiter cette question de la calculabilité, commence par corréler l'esprit au cerveau (à chaque état de l'esprit doit correspondre un état du cerveau), et montre

22 Ce texte est de la main de Gödel (Papiers Gödel, 8c, 117, 040393). Gödel ajoute que « le mécanisme en biologie est un préjudice de notre temps qui sera réfuté ». On montrera, pense Gödel, que la constitution d'un organisme comme le corps humain purement selon les lois physiques est aussi peu probable que « la séparation de l'atmosphère en ses composants » [Wang 1974, p. 326]. Voir également, papiers Gödel, boîte 20, où se trouve une autre version du texte de Wang, corrigée de la main de Gödel. Ce rejet du mécanisme en biologie est une conviction ancienne de Gödel, dont on peut se demander quel rôle elle joue ensuite dans la critique de la thèse de Turing, à propos des procédures mathématiques. Ainsi, Gödel se demande autour de 1942 : « *Ist der Unterschied des Lebendigen vom Leblosen vielleicht, daß sich seine Wirkungsgesetze nicht in 'mechanischen' Regeln fassen, d.h. nicht 'formalisieren' lassen ? [...] die dem Vitalismus entgegengesetzte Anschauung behauptet also einfach : alles Lebendige ist tot* » [*Cahier philosophique* XI, p. 44, transcription C. Dawson].

ensuite que le cerveau ne peut prendre qu'un nombre fini d'états différents. Gödel accepte ce deuxième point, sur le cerveau, mais refuse la thèse d'un « parallélisme » esprit/cerveau, qu'il dénonce comme une illusion. [23] Il évoque, dans la citation (32), une réfutation empirique, qui montrerait que les états possibles de ce système fini qu'est le cerveau, sont moins nombreux que les opérations élémentaires que l'on peut distinguer dans l'esprit. Or le fait de détacher ainsi l'esprit du cerveau permettrait de donner à celui-là une infinité potentielle. Cette infinité potentielle de l'esprit humain, avec son soubassement métaphysique, anti-matérialiste, apparaît également dans la note suivante :

(33) « *G. seriously questions (1) [that there is no mind separate from matter]. For all we know mind or* [Wang écrit : « *spirits* », Gödel remplace le terme par : « *the ego* »] *may be distinct from the brain and, in the course of an infinite time, be capable of an infinite number of distinguishable states. The brain may be essentially a digital computer* [Gödel ajoute « *temporally* » dans le texte de Wang] *connected with a finite mind, capable of unlimited* [Gödel ajoute « *systematic non mechanical* » dans le texte de Wang] *development. It is not denied that the mind is finite and capable of only finitely many states at each stage of its development. This observation points up to a possible world (which may be the real world) in which there would for finite minds exist mental procedures not equivalent to any Turing machine* » [Papiers Gödel, Boîte 20, dossier Wang].

Ainsi, Gödel accepte que le cerveau est une machine de Turing de sorte que, dans le cas même où sa vie durerait un temps infini, le cerveau ne serait capable que d'un nombre fini d'états internes. En revanche, Gödel semble soutenir que l'esprit est indépendant (ou, pour une part, indépendant) du cerveau de sorte que, si, à chaque étape, l'esprit reste fini, c'est-à-dire n'est encore passé que par un nombre fini d'états mentaux, il est susceptible d'une infinité d'états, au cours d'un développement analogue à une vie éternelle. L'énoncé de Turing (cité plus haut (30)), que le calculateur humain comme la machine ne dispose que d'un nombre fini d'états possibles, ou d'une mémoire finie, ne se justifie que sous l'hypothèse, ou bien d'une sorte d'identité ou bien d'une sorte de parallélisme entre l'esprit et le cerveau. L'argument de Gödel est d'abord négatif. Il montre que l'énoncé de Turing dépend d'un cadre métaphysique, un cadre « matérialiste », qui n'est pas justifié par Turing et qui, pour Gödel, ne va pas de soi. Mais Gödel vise ensuite bien à établir, dans un autre cadre métaphysique,

23 Également [Wang 1974, p. 190 et p. 198].

que l'esprit est capable d'une infinité d'états et peut par conséquent réaliser des procédures, qui ne pourraient pas être reproduites par une machine de Turing. La discussion avec Turing est liée aux convictions métaphysiques, voire religieuses, de Gödel. J. C. Webb [Webb 1990, p. 299] indique que l'argument fait appel à un « esprit infini ». Il s'agit plus exactement d'un esprit potentiellement infini, un esprit qui, tout en restant toujours fini, se développe au-delà de toute borne.

On peut se demander dans quelle mesure la critique de Gödel adressée à Turing est justifiée. En effet, comme l'a montré W. Sieg [Sieg 2006, p. 98], [Sieg 1994, p. 15 et suiv.], Turing dans ce passage de l'article de 1937 ne vise pas à restreindre les possibilités de l'esprit en général, pour montrer que les processus de l'esprit sont toujours des procédures mécaniques, susceptibles d'être représentées par une machine de Turing. Turing cherche uniquement à montrer que l'esprit, au cours de ce qu'on appelle un calcul, est une telle machine, sans vouloir nier qu'il puisse exister dans l'esprit d'autres processus, d'autres facultés qui ne soient pas des calculs. Ainsi, l'article sur les logiques ordinales distingue dans le raisonnement mathématique deux facultés : « l'ingénuité », d'une part, qui consiste dans un savoir-faire logique, à savoir appliquer les règles logiques le mieux possible, et que rend en principe superflue l'existence (au moins virtuelle) de ces machines capables de déduire les théorèmes d'une logique ; « l'intuition », d'autre part, qui ne se résume pas dans un calcul, l'application de règles, mais « consiste à faire des jugements spontanés, qui ne sont pas le résultat d'une suite consciente de raisonnements » [Turing 1939, p. 208]. Cette faculté d'intuition peut alors se représenter par un « oracle », un dispositif qui est accouplé à la machine. Cette faculté d'intuition, cet oracle, dont on peut seulement « dire qu'il n'est pas une machine » [Turing 1939, p. 167], semble alors représenter une composante de l'esprit qui échappe à l'analyse que proposait Turing en 1937. Celle-ci ne vise donc pas à élucider le fonctionnement de l'esprit en général, mais seulement le fonctionnement de l'esprit dans l'une de ses activités, le calcul. La critique de Gödel (dans la première phrase de la citation (29)) passe à côté de la lettre du texte de Turing. Cela dit, la thèse de Gödel est que l'esprit, parce que il est capable d'un développement indéfini, est susceptible de réaliser des procédures qui dépassent les machines de Turing. C'est de cette thèse, qu'elle contredise ou non le texte de Turing, qu'il faut discuter.

Une machine, définie de façon analogue aux machines de Turing mais susceptible de prendre un nombre infini d'états internes et disposant d'une liste infinie d'instructions relatives à ces états, peut « calculer » n'importe quelle fonction entière, n'importe quel nombre réel, ou écrire n'importe quelle suite de formules et, par exemple, toutes les formules vraies de l'arithmétique élémentaire dans un modèle donné. Cependant, une telle machine, personne (sinon un dieu) ne pourrait la programmer ou en comprendre le programme, puisque celui-ci comporte une infinité d'instructions. C'est pourquoi du reste, dans l'article de 1937, Turing qui ne considère que des calculs « finis », définis par une liste finie d'instructions, passe très vite sur cette question du nombre des états possibles du calculateur humain.[24] Cela dit, on peut bien sûr imaginer que l'esprit humain est une machine infinie avec un nombre infini d'états possibles et qui suit un programme infini, une liste infinie d'instructions relatives à ces états. Cet esprit ne serait pas très différent de l'automate spirituel que décrit Leibniz.[25] Et il pourrait, purement selon les lois qui régissent son fonctionnement, écrire une suite de formules, infinie, qu'aucune machine de Turing ne pourrait écrire. Cependant, dans ces passages que nous étudions, Gödel ne dit pas que l'esprit est pris dans une telle procédure, qui resterait comme inconsciente et qui, par exemple, le conduirait comme malgré lui à écrire les unes à la suite des autres toutes les formules vraies de l'arithmétique élémentaire ou de la théorie des ensembles. Lorsque Gödel parle d'un développement de l'esprit (dans la citation (29)), c'est plutôt, et comme dans les « Remarques » de 1946 déjà évoquées, que l'esprit, d'abord comparable à une machine de Turing, déduisant des formules dans un système donné, est toujours capable d'introduire de nouveaux axiomes et, par conséquent, d'étendre les systèmes dans lesquels il travaille suivant une procédure que ne peut pas reproduire une machine de Turing. Admettons que l'argument (qu'il soit celui de Turing ou non), qui dirait qu'un esprit fini ne peut pas faire autre chose que des calculs mécaniques, est réfuté. Admettons, pour suivre Gödel, qu'un esprit fini est susceptible au cours d'un temps infini de passer par une infinité d'états distincts. Bref, l'esprit en principe peut conduire des procédures qui ne sont pas réalisables sur une machine de Turing. Mais il reste à savoir comment, selon quelles règles, selon quelle méthode, partant d'un système donné dans lequel le calculateur humain est analogue à une machine de Turing, il est possible d'élargir ce système, avec

24 Turing [Turing 1937, p. 137] remarque simplement (dans le cas où l'on ne dispose que d'une liste d'instructions finie) : « *This restriction is not one which seriously affects computation* ».

25 Du moins, l'automate spirituel comporte bien une infinité d'états possibles, et ses lois ne s'expriment pas sous la forme d'une liste finie d'instructions, [notamment [Leibniz 1969], § 403 et 1765, II, XII, § 17]. Sur l'écart entre la notion leibnizienne d'automate spirituel et la machine de Turing, cf. [Cassou-Noguès 2002].

de nouveaux axiomes selon une procédure qui n'est pas représentable par une machine de Turing. Gödel exige que la procédure qui guide ce développement soit déterministe. Cela est sans doute sous-entendu dans le terme « systématique » des citations (29) et (33). Mais, de façon explicite :

(34) « *Note also that the term well-defined is supposed to imply that in any application of a procedure the result as well as the intermediate stages depend only on the initial elements and on no other circumstance* » [Papiers Gödel, 9, 147, 040486].

Les règles qui déterminent ces procédures « bien définies » mais non mécaniques doivent être univoques, prescrivant sans ambiguïté une suite d'actions (que ce soit pour le « calcul » d'une fonction entière ou l'extension d'un système formel). La question maintenant est de savoir de quelle nature doivent être ces règles. Comment énoncer des règles univoques mais que ne pourrait suivre aucune machine de Turing ? Gödel, de façon très générale, renvoie au *sens*. Ces règles, des procédures non mécaniques, devront faire appel au « sens », alors que les règles mécaniques ne portent que sur les symboles, abstraction faite de leur sens. C'est ce que montre par exemple la citation (19). De même, le Postscriptum de 1964 indique que ces procédures finies non mécaniques « impliquent l'usage de termes abstraits sur la base de leur sens » [Gödel 1934, p. 370].

Ce sens, invoqué comme moteur des procédures non mécaniques, dépend, dans la perspective de Gödel, d'une intuition des concepts. Comme cela apparaît déjà dans les « Remarques » de 1946, Gödel est convaincu qu'une intuition suffisamment claire des concepts fondamentaux des mathématiques fournirait d'elle-même une procédure précise pour formuler de nouveaux axiomes, étendre indéfiniment nos systèmes formels jusqu'à obtenir une propriété de complétude et, par conséquent, selon une procédure qu'aucune machine ne peut suivre. De la même façon, la première version de l'argument donné contre Turing [Gödel 1972b] renvoie de façon assez générale à des « concepts abstraits ». [26] La deuxième version (citation (29)) ouvre une autre perspective. C'est que, pour obtenir une telle procédure, il faudrait remplacer le concept d'ensemble à la base de nos mathématiques, par un autre plus général, comme le serait celui de classe propre ou, son corrélat, le concept de concept. Le concept de concept a cette propriété, selon Gödel, de s'appliquer à lui-même : le concept de concept est un concept. Si l'on considère l'extension de ce concept, il faudrait dire qu'elle appartient à elle-même. Le concept de concept semble

26 [Gödel 1972b, p. 306] : « *we understand abstract terms more and more precisely as we go on using them, and [...] more and more abstract terms enter the sphere of our understanding* ».

plus général que celui d'ensemble. On peut dire que chaque ensemble correspond à un concept, mais il y a des concepts dont l'extension n'est pas un ensemble. On pourrait parler de classe, en admettant alors qu'une classe puisse appartenir à elle-même. Gödel envisage que le concept de concept, ou celui de classe en ce sens, puisse remplacer celui d'ensemble, comme base de nos mathématiques. Les deux citations suivantes, des brouillons pour la deuxième moitié de la citation (29), développent cette idée : la définition d'une procédure réglée non mécanique dépend de l'intuition d'un concept donnant lieu indéfiniment à de nouveaux axiomes, comme le concept d'ensemble mais plus large que celui-ci.

(35) « *Several more primitive ideas now appear on the horizon destined for the same development [as the concept of set]. Two instances are the self-reflexive concept of proper class and the most general concept of concept (and the concept of true analogue). It would be furthermore an achievement of unprecedented importance since it would imply an acceleration so that developments of whole centuries would be compressed into a few years in analogy to the development of modern technology which has achieved more in a 100 years than was achieved in the preceding 6000 years* » [Papiers Gödel, 8c, 117, 040395].

(36) « *It is moreover conceivable that someday one may hit upon an abstract concept much more clearly discernible than the general concept of set so that the corresponding procedure of developing on its basis much more abstract concepts and their evident mutual relationships would from the beginning be unambiguous and proceed without much learning by use. All this may sound fantastic and unfounded. However, note that mere logical possibilities are sufficient to disprove the conclusiveness of an argument in logic or mathematics. Nor is it inconceivable, although hard to imagine at present, that we should develop a sufficiently clear concept of our mind to see that such a procedure can be continued indefinitely with well-determined results. Note that something like this actually seems to happen in the procedure of forming higher and higher axioms of infinity* » [Papiers Gödel, 9b, 147, 040488]. [27]

Gödel ne précise pas ce qu'il entend par « analogue véritable », et c'est à ma connaissance la seule occurrence de cette expression dans ses notes. L'important est que la possibilité d'une extension indéfinie de nos systèmes d'axiomes dans une procédure univoque et non mécanique, non reproductible par une machine de Turing, semble dépendre de la substitution, ou de l'adjonction, au concept d'ensemble d'un concept plus général, comme le concept de concept ou de classe propre. La citation (35) indique également que cette perspective reste, dans l'esprit de Gödel, lointaine : son actualisation correspondrait à une transformation complète de nos mathématiques.

---

27 On voit dans ces deux citations que Gödel n'a manifestement pas en tête un principe de réflexion analogue à ceux de Turing (1938) et surtout Feferman [1991].

Maintenant, la saisie précise du sens d'un concept, comme elle est exigée pour cette extension non mécanique de nos systèmes formels, repose sur une réflexion sur nos actes, notre pensée, dans l'intuition de ce concept. C'est ce qui ressort en particulier de la référence à la phénoménologie de Husserl dans la conférence de 1961. Gödel y explique que la production de nouveaux axiomes suppose une clarification du sens des concepts et que celle-ci, comme le veut la phénoménologie dans la lecture qu'en fait le logicien, exige une analyse réflexive de nos actes dans l'usage des concepts. Et Gödel conclut que, précisément, c'est ce processus, qui suppose par conséquent un retour réflexif sur la pensée même, « qu'une machine ne peut pas imiter » [Gödel 1961/?, p. 384-385].

Cette référence à la phénoménologie comme méthode pour la clarification du sens éclaire un point que la comparaison des deux versions de l'argument élaboré contre Turing, comme d'autres notes de Gödel, semble rendre ambigu. Gödel lie la découverte d'une procédure non mécanique à une meilleure compréhension tantôt des concepts mathématiques, tantôt de l'esprit humain, de son essence et de son fonctionnement. C'est, pour Gödel, que l'une ne va pas sans l'autre. Et, en dernier ressort, c'est dans une réflexion phénoménologique que Gödel cherche cette analyse de l'esprit qui doit donner à la fois une clarification des concepts mathématiques et la formulation d'une procédure non mécanique. Ainsi, à la fin de la note mentionnée plus haut (citation (27)) sur ces « procédures de pensée » qui fourniraient des fonctions « calculables » mais « non mécaniques », Gödel remarque :

(37) « *This insight is based on a psychological (phenomenological) reflection, whose fruitfulness for the foundations of mathematics is thereby clearly demonstrated* » [Papiers Gödel, 9b, 141, 040405].

Ces procédures de pensée, bien définies mais non mécaniques, dépendent donc d'une réflexion phénoménologique et, plus largement, de cette faculté qu'a la pensée, selon Gödel, de pouvoir se réfléchir et se connaître elle-même. C'est un thème qui apparaît également dans une lettre à Tillich. [28]

Ces remarques de Gödel, autour de la référence au sens, ne résolvent cependant pas les difficultés auxquelles est confronté le logicien dans sa critique de Turing. En fait, il y a deux problèmes exactement symétriques. On l'a vu, Gödel soutient que l'esprit, tout en restant fini, est capable d'un développement indéfini, de sorte que, si on le compare à une machine de Turing, on peut dire que le nombre des états par lesquels il passe

28 Cf. [Atten 2006].

augmente avec le temps et que, dans l'ensemble de son développement, le nombre de ces états est infini. Une « machine », analogue aux machines de Turing mais possédant une infinité d'états internes et une liste infinie d'instructions, peut écrire n'importe quelle suite de formules ou calculer n'importe quelle fonction entière. Si l'esprit, dans le cours de son développement est comparable à une telle machine, il est en principe capable de procédures qui ne sont pas mécaniques au sens de Turing. Mais la question reste de savoir comment formuler une procédure bien déterminée, que nous puissions réellement appliquer et qui soit pourtant non mécanique. Par exemple, partant d'un système donné, un système formel de théorie des ensembles, comment, c'est-à-dire selon quelles règles, peut-on l'étendre de façon à obtenir une série d'axiomes, bien définie mais qu'aucune machine ne puisse écrire et pour laquelle vaudrait un résultat de complétude ? La thèse qu'une telle procédure doit faire intervenir le sens des concepts est un leitmotiv, dans les notes de Gödel, mais elle n'apporte pas de réponse. La référence au sens dans la définition de la procédure ne donne aucune assurance sur le caractère non mécanique de celle-ci. Ce n'est pas, évidemment, parce que l'on enchaîne des axiomes (ou des phrases en général) en s'appuyant sur leur sens qu'une machine ne peut pas écrire les mêmes axiomes (les mêmes phrases) en procédant en quelque sorte à l'aveugle, selon un programme, ou des règles qui ne prennent en compte que les symboles.

Mais, inversement, s'il est entendu que l'esprit, au cours de son développement, est comparable à une « machine » susceptible de passer par une infinité d'états et que, dans ce cas, l'esprit est en principe capable de procédures qui ne sont pas représentables par une machine de Turing, pourquoi la définition d'une telle procédure devrait-elle forcément faire référence au sens ? Partant d'un système donné, demandons à nouveau comment, c'est-à-dire encore une fois selon quelles règles, ajouter de nouveaux axiomes (indéfiniment) pour obtenir un système complet. Imaginons que ces règles, qui appartiennent de toute façon à un métalangage, soient formulées comme des phrases dans le langage naturel. Ces phrases, en français disons, donnent certaines instructions permettant, à chaque étape, pour chaque système obtenu, de formuler encore de nouveaux axiomes. Pourquoi ces règles, ces phrases en français, devraient-elles s'appuyer sur le sens des symboles du système et ne peuvent-elles consister simplement en une série d'instructions ne prenant en considération que les symboles eux-mêmes ?

Il faut revenir à l'interprétation par Gödel de la thèse de Turing. Des règles, comme celles qui définissent un système formel, qui ne portent que sur les symboles, sans considération de leur sens, définissent une procédure formelle, mécanique au sens où Gödel l'entendait avant Turing. Et c'est, d'après Gödel, la thèse de Turing : une telle procédure est représentable par une machine de Turing. La citation (19) montre que Gödel accepte cette thèse — une procédure mécanique en ce sens peut être représentée par une machine de Turing —, alors même qu'il critique l'un des arguments de l'article de Turing. La justification de cette thèse semble, dans les textes de Gödel, ne dépendre que d'une intuition du concept de procédure « mécanique » ou faisant abstraction du sens des symboles. Quoi qu'il en soit, la conséquence de cette thèse est qu'une procédure qui ne serait pas représentable par une machine de Turing doit s'appuyer sur le sens. La citation (28) semble montrer que Gödel a toujours été convaincu de l'existence de procédures non mécaniques, de procédures définies par des règles dans lesquelles le sens est irréductible, c'est-à-dire que l'on ne peut pas remplacer par des règles qui ne porteraient que sur les symboles abstraction faite de leur sens. Cela est du moins clair dans les « Remarques » de 1946, qui évoquent l'existence de procédures pour l'extension des systèmes formels, bien définies mais non mécaniques et dépendant, plutôt que de règles portant sur les symboles, de l'intuition d'un concept. L' « argument » que Gödel met au point en 1964, n'ajoute presque rien de ce point de vue. Il est négatif, visant à réfuter un énoncé que Gödel prête à Turing, sans doute à tort, à savoir que la pensée mathématique ne peut s'exprimer que dans des procédures mécaniques. Cet énoncé, montre Gödel, s'appuie sur un cadre métaphysique, un certain matérialisme, qui n'est pas justifié et que l'on n'est pas tenu d'accepter. Gödel, après cet argument, considère l'esprit comme une sorte de machine potentiellement infinie. Dans la mesure où cette machine, à chaque étape de son développement, n'a utilisé qu'un nombre fini d'états internes (ou, si l'on veut, n'a jamais qu'une mémoire finie), Gödel peut qualifier les procédures non mécaniques de « finies ». Mais cet argument ne résout en rien le problème auquel Gödel est déjà confronté en 1946 : comment fixer une procédure, bien définie, avec des règles que le mathématicien puisse suivre mais qu'aucune machine de Turing ne puisse appliquer ? Non seulement l'argument donné contre Turing ne répond pas à ce problème, mais il tendrait plutôt à mettre en question la thèse même que Gödel tire de l'article de Turing et l'identification entre procédure mécanique et machine de Turing. Du moins, Gödel attaque l'argumentation par laquelle Turing justifiait sa propre thèse et sa définition des procédures « finies », sans chercher à lui substituer une autre argumentation. C'est apparemment que, dans l'esprit de Gödel, cette tâche est inutile. Si l'on en croit les citations (20), (21), (22), (23), il faudrait considérer la thèse de Turing telle que la formule Gödel (une procédure « mécanique », qui fait abstraction du sens des symboles, est représentable par une machine de Turing) et sa conséquence immédiate (une procédure non représentable par une machine de Turing s'appuie sur le sens des symboles) comme des propositions évidentes qui ne dépendent que de l'intuition du concept de procédure mécanique, concept qui a une existence en soi (notamment, citation (23)). Aucune de ces deux propositions ne semble être autrement justifiée. Mais Gödel, sans doute, n'a jamais prétendu être en mesure de décrire une telle procédure, qui exigerait une transformation profonde de notre culture mathématique :

(38) « *It should be noted that the realization of such procedures [non mechanical] would require /a development of the faculties of the human mind well beyond the stage that has been achieved (or potentially achieved) in our culture (the science of) today /a substantial improvement in our understanding of the abstract ideas either beforehand or continuously during the application of the procedure.*

A clear knowledge of the feasibility and successfulness of such procedures would even presuppose a development of our culture (the science of) today /a very substantial improvement (progress, advance) in our understanding of the abstract ideas of mathematics and of the process of understanding itself.

Such a procedure is certainly not within reach before a very substantial advance in our understanding of the abstract ideas of mathematics has taken place » [Papiers Gödel 8c, 106, 040332].

Remerciements

Cet article utilise les textes inédits de Gödel, conservés à la bibliothèque de l'université de Princeton et que nous avons pu consulter grâce à une bourse de la Society of the Friends of Princeton University Library, puis dans le cadre du projet Preuve de l'UMR STL. Une première version de ce texte a été présentée en janvier 2006 aux Archives Poincaré à Nancy. Nous sommes redevables à deux rapporteurs anonymes et à W. Sieg, dont les nombreuses remarques nous ont permis d'améliorer ce texte sur des points importants.

BIBLIOGRAPHIE

ACKERMANN (Wilhelm)

[1928] Zum Hilbertschen Aufbau der reellen Zahlen, *Math. Ann.*, 99(1) (1928), p. 118–133.

Atten (Mark (van))

[2006] Two draft letters from Gödel on self-knowledge of reason, *Philos. Math.*, 14(2) (2006), p. 255–261.

Cassou-Noguès (Pierre)

[2002] Deux figures de l'automate spirituel : Leibniz et Turing, dans Fédi (Laurent), éd., *La migration des concepts*, Paris : L'Harmattan, 2002, p. 51–68.

[2005] Gödel and 'the objective existence' of mathematical objects, *Hist. Philos. Logic*, 26(3) (2005), p. 211–228.

[2007] *Les démons de Gödel*, Paris : Seuil, 2007.

Dawson (John W. Jr)

[1997] *Logical Dilemmas : The Life and Work of Kurt Gödel*, Wellesley, MA : A K Peters Ltd., 1997.

[2003] Introductory note to the Gödel-Church correspondence, 2003 ; in [Gödel, *Works*, t. IV, p. 361–366].

Davis (Martin)

[1965] *The Undecidable. Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*, Ed. by Martin Davis, Hewlett, N.Y. : Raven Press, 1965.

[1982] Why Gödel didn't have Church's thesis, *Inform. and Control*, 54(1-2) (1982), p. 3–24.

Feferman (Solomon)

[1991] Reflecting on incompleteness, *J. Symbolic Logic*, 56(1) (1991), p. 1–49.

Gandy (Rudy)

[1980] Church's Thesis and principles for mechanism, dans Barwise & others, éd., *The Kleene Symposium*, Amsterdam : North-Holland, 1980.

Gödel (Kurt)

[*Works*] *Collected Works*, ed. by Solomon Feferman, John W. Dawson *et al.*, Oxford : Oxford University Press, 1986–2003.

[*Archives*] Archives Gödel de l'*Institute for Advanced Studies*, conservées à la bibiothèque de l'université de Princeton (cote 0282), 1986–2003.

[1931] Über formal unentscheidbare Sätze der *Principia mathematica* und verwandter Systeme I, 1931 ; in [Gödel, *Works*, t. I, p. 144–195].

[1933] The present situation in the foundations of mathematics, *Philos. Math.*, 16(1) (1933), p. 100–112 ; in [Gödel, *Works*, t. III, p. 45–53].

[1934] On undecidable propositions of formal mathematical systems, 1934 ; in [Gödel, *Works*, t. I, p. 346–372].

[1936] Über die Länge von Beweisen, *Philos. Math.*, 16(1) (1936), p. 100–112 ; in [Gödel, *Works*, t. I, p. 396–398].

[193 ?] Undecidable Diophantine propositions, 193 ? ; in [Gödel, *Works*, t. III, p. 164–175].

[1941] In what sense is intuitionistic logic constructive ?, *Philos. Math.*, 16(1) (1941), p. 100–112 ; in [Gödel, *Works*, t. III, p. 189–201].

[1946] Remarks before the Princeton bicentennial conference on problems in mathematics, 1946 ; in [Gödel, *Works*, t. II, p. 149–152].

[1951] Some basic theorems on the foundations of mathematics and their philosophical implications, *Philos. Math.*, 16(1) (1951), p. 100–112 ; in [Gödel, *Works*, t. III, p. 304–323].

[1958] Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes, *Dialectica*, 12 (1958), p. 280–287 ; in [Gödel, *Works*, t. II, p. 240–252].

[1961/ ?] The modern development of the foundations of mathematics in the light of philosophy, *Philos. Math.*, 16(1) (1961/ ?), p. 100–112 ; in [Gödel, *Works*, t. III, p. 374–387].

[1964] What is Cantor's Continuum Problem ?, 1964 ; in [Gödel, *Works*, t. II, p. 254–270].

[1972a] On an extension of finitary mathematic which has not yet been used, 1972 ; in [Gödel, *Works*, t. II, p. 271-280].

[1972b] Some remarks on the undecidability results, 1972 ; in [Gödel, *Works*, t. II, p. 305–306].

Herbrand (Jacques)

[1931] Sur la non-contradiction de l'arithmétique, *Journal für die reine und angewandte Mathematik*, 166 (1931), p. 1–8.

Hilbert (David)

[1926] Über das Unendliche, *Math. Ann.*, 95(1) (1926), p. 161–190.

Kleene (Stephen)

[1936] General recursive functions of natural numbers, *Math. Ann.*, 112(1) (1936), p. 727–742.

[1981] Origins of recursive function theory, *Ann. Hist. Comput.*, 3(1) (1981), p. 52–67.

[1987] Gödel's impression on students in logic in the 1930s, dans Weingartner (Paul) & Schmetterer (Leopold), éd., *Gödel Remembered, History of Logic*, 4, Gödel Symposium, Salzburg, July 10–12, 1983, Naples : Bibliopolis, 1987, p. 49–64.

[1988] Turing's analysis of computability, and major applications of it, dans Herken (Rolf), éd., *The Universal Turing Machine : A Half-Century Survey*, New York : The Clarendon Press, Oxford University Press, 1988, p. 17–54.

Leibniz (Gottfried W.)
[1710] *Essais de théodicée* (1710), éd. de Jacques Brunschvicg, Paris : Garnier-Flammarion, 1969.
[1765] *Nouveaux essais sur l'entendement humain* (1765), éd. de Jacques Brunschvicg, Paris : Garnier-Flammarion, 1966.

Mancosu (Paolo)
[1999] Between Vienna and Berlin : the immediate reception of Gödel's incompleteness theorems, *Hist. Philos. Logic*, 20(1) (1999), p. 33–45.

Sieg (Wilfried)
[1994] Mechanical procedures and mathematical experience, dans George (A.), éd., *Mathematics and mind*, Logic Comput. Philos., Oxford : Oxford Univ. Press, 1994, p. 71–117.
[1997] Step by recursive step : Church's analysis of effective calculability, *Bull. Symbolic Logic*, 3(2) (1997), p. 154–180.
[2005] Only two letters : the correspondence between Herbrand and Gödel, *Bull. Symbolic Logic*, 11(2) (2005), p. 172–184.
[2006] Gödel on computability, *Philos. Math.*, 14(2) (2006), p. 189–207.

Sinaceur (Hourya)
[2000] Address at the Princeton University Bicentennial Conference on Problems in Mathematics, by Alfred Tarski, *Bull. Symbolic Logic*, 6 (2000), p. 1–44.

Turing (Alan)
[1937] On computable numbers, with an application to the Entscheidungsproblem, 1937 ; in [Davis 1965, p. 115–153].
[1939] Systems of logic based on ordinals, 1939 ; in [Davis 1965, p. 154–222].

Wang (Hao)
[1974] *From Mathematics to Philosophy*, Londres : Routledge & Paul, 1974.

Webb (Judd C.)
[1980] *Mechanism, Mentalism and Metamathematics*, Dordrecht : Reidel, 1980.
[1990] Introductory note to [Gödel 1972b], remarks 3, 1990 ; in [Gödel, *Works*, t. II, p. 292–304].