

CENTRE DE RECHERCHES SEMIOLOGIQUES

TRAVAUX DE LOGIQUE

KURT GÖDEL

Actes du Colloque,
Neuchâtel, 13 et 14 juin 1991

Denis Miéville (éditeur)

CdRS



Université de Neuchâtel

Centre de Recherches Sémiologiques
Travaux de logique
N° 7 - Juin 1992

KURT GÖDEL
Actes du Colloque,
Neuchâtel 13 -14 juin 1991
Denis Miéville (éditeur)



Université de Neuchâtel

**Centre de Recherches Sémiologiques
Université de Neuchâtel
Espace Louis-Agassiz 1
CH-2000 Neuchâtel (Switzerland)**

©1992 by Centre de Recherches Sémiologiques. Tous droits réservés

Sommaire

Denis MIEVILLE (Université de Neuchâtel)	
Gödel: un passé, un présent, un avenir!	1
Hourya SINACEUR (Université de Paris I)	
Sur le rôle de l'analyse des concepts selon Gödel et son rapport avec la théorie des modèles.....	11
John CORCORAN (State University of New York at Buffalo)	
Unprovability and undefinability	37
Richard VESLEY (State University of New York at Buffalo)	
Boolos' nonconstructive proof of Gödel's theorem.....	67
Jacques Paul DUBUCS (Université de Paris I)	
Arguments gödéliens contre la psychologie computationnelle.....	73
Henri LAUENER (Université de Berne)	
Remarques sur l'interprétation du théorème d'indécidabilité.....	91
Carlo CELLUCCI (Università degli studi di Roma «La Sapienza»)	
Gödel's incompleteness theorem and the philosophy of open systems	103
Evandro AGAZZI (Université de Fribourg)	
Some philosophical implications of Gödel's theorem.....	129
Jean-Paul DELAHAYE (Université de Lille)	
Importance ou insignifiance mathématique des indécidables de Gödel.....	161

GÖDEL: UN PASSE, UN PRESENT, UN AVENIR

Denis Miéville

[La] limitation fondamentale [que Gödel a mise en évidence] est de ne pouvoir parler adéquatement de la transcendance dans le langage de l'effectuable, c'est-à-dire, du fini.

(Ladrière)

Gödel, un logicien hors du commun

Le samedi 14 janvier 1978, K. Gödel décède de «malnutrition et inanition» à l'hôpital de Princeton (Pennsylvanie, USA). Il s'éteint à l'âge de 71 ans. Natif de Moravie, formé à l'Université de Vienne en physique et mathématique, Gödel va marquer de son empreinte l'histoire de la logique. Sa mort absurde contraste fortement avec la liste des honneurs qui l'ont couronné. Bénéficiaire du premier prix Einstein (1950), il obtint le titre de docteur en Littérature *honoris causa* de l'Université de Yale, puis de docteur en Sciences *honoris causa* de l'Université de Harvard. Membre permanent de l'Institute for Advanced Studies de Princeton depuis 1938, il reçoit du président Ford la Médaille Nationale de la Science en 1975. Cet être surprenant à plus d'un titre aimait à ce qu'on l'appelle «le découvreur de la vérité la plus importante de son siècle». D'aucuns même n'ont pas craint de le comparer à Aristote.

Pour mieux comprendre son influence, pour mieux saisir la subtilité de son génie, il n'est pas inutile de préciser le climat in-

telle que celui-ci prévalait à l'aube de ce siècle dans la communauté des logiciens et de certains mathématiciens.

Pendant plus de deux mille ans, la logique reste intimement liée au discours. Que l'on pense à Aristote et ses syllogismes, que l'on songe à l'enseignement de la logique dans les écoles médiévales: il s'agissait alors, comme l'a écrit Kotarbinski, d'un programme de spéculation verbale. Puis la logique trouve enfin son identité formelle avec Gottlob Frege (1848-1925). Avec lui, c'est la logique mathématique qui prend naissance. On attribue même une année de naissance à cette logique: celle de la parution de l'ouvrage de Frege, la *Begriffsschrift*, 1879. Le projet de Frege est double. D'une part, mathématicien, il veut expliciter ce qui, dans le discours des démonstrations mathématiques, est exprimé dans une langue naturelle et règle les preuves. Il s'agit donc d'aller en deçà des structures mathématiques et d'extraire la substance logique même qui gère le raisonnement mathématique. Il réalise son projet en construisant un système hypothético-déductif pour la logique, système qui permet de fonder les raisonnements de la démonstration mathématique. D'autre part, Frege est persuadé que la logique est le fondement même de l'arithmétique, que les propositions arithmétiques peuvent être réduites à celles de la logique, que l'arithmétique, donc, est dépendante de la logique.

La réalisation de Frege est caractéristique d'un esprit fécond qui a su offrir les assises d'une science nouvelle, libérant la logique de la tutelle des discours tenus dans une langue naturelle, réalisant ainsi une partie du pari de Leibniz. De plus, Frege a su échapper à l'omniprésence du modèle mathématique pour mieux en saisir les fondements.

Frege possède un lecteur attentif de ses oeuvres, il s'agit de Bertrand Russell (1872-1970). En 1902, ce dernier lui fait part de son admiration, mais également lui communique une propriété pour le moins gênante que son système possède: celui-ci est contradictoire. Cette découverte ébranle fortement l'édifice mathématique. En effet, la contradiction semble régner dans les bases mêmes de cette science.

L'antinomie que révèle Russell oblige les mathématiciens et les logiciens à prendre conscience de la fragilité des règles de raison-

nement, et davantage encore, que l'arithmétique, la «reine des sciences», repose sur des fondements fragiles. Le problème de sa non-contradiction se pose alors avec une grande acuité. En effet, il s'agit de s'assurer avant de développer une théorie mathématique, que ses principes fondateurs sont exempts de toute contradiction et que les règles de déduction ne peuvent en produire une. C'est à ce problème difficile et important que va s'atteler David Hilbert (1862-1943). La solution proposée est la marque d'un esprit subtil.

Hilbert tire une leçon de l'histoire des sciences, celle qui met en évidence la faiblesse, la non-certitude de nos sens, de nos jugements lorsqu'ils tentent de cerner un contenu conceptuel en un processus rationnel. Il est persuadé que la mathématique doit être libre de tout préjugé et que, pour cela, il faut en aborder les fondements d'une manière uniquement formelle.

Hilbert s'engage dans une réflexion sur la théorie de la démonstration en privilégiant la dimension syntaxique, théorie dans laquelle il n'est pas question de considérer une organisation mathématique chargée de sens. Une de ses idées est d'avoir formulé la distinction précise entre le langage utilisé pour une théorie mathématique et le langage qui sert à en parler. Il sépare ainsi nettement mathématique et métamathématique. Celle-ci se présente comme un système axiomatique qui s'organise sur la base de matériaux symboliques et qui contient des règles formelles de formation et de transformation, règles purement opératoires. Une telle structure doit être contrôlable. Il faut donc exiger des conditions d'effectivité, et n'admettre que des procédures finitaires. Dans cette perspective, le problème de la non-contradiction se pose en d'autres termes. Il s'agit de démontrer qu'en appliquant les règles formelles aux propositions de base, il est impossible d'obtenir à la fois une formule du système et sa négation. On veut davantage encore. Si l'on admet que la mathématique contient tous les modes de raisonnement rationnel, elle contient donc ceux qui s'appliquent à la métamathématique. La métamathématique posséderait donc, de manière réfléchissante, les moyens de montrer sa propre non-contradiction. Ce rêve s'éteint lorsque Kurt Gödel démontre en 1931 l'impossibilité de prouver la non-contradiction de toute théorie contenant l'arithmétique si les démons-

trations utilisées consistent en ces procédés finis, voulus par Hilbert. Dans cette perspective, la mathématique est essentiellement incomplète, et davantage encore, elle ne saurait être complétée: la méthode axiomatique possède certaines limites internes qui excluent la possibilité d'axiomatiser dans sa totalité l'arithmétique, et donc la mathématique classique. Il n'y a pas de système axiomatique de type hilbertien capable d'envisager la totalité des problèmes de la pensée mathématique, ainsi que la possibilité d'en proposer toutes les réponses.

Mais les théorèmes critiques de Gödel ne marquent cependant pas le crépuscule d'un temps et d'une manière de vivre la pensée mathématique.

1991: Gödel, quelque soixante ans plus tard

En 1979, l'oeuvre de Gödel est tombée, si je puis m'exprimer ainsi, dans le domaine public. Cette année-là, Hofstadter publie son ouvrage à succès: *Gödel, Escher, Bach*. L'auteur exploite de manière talentueuse et fascinante, une certaine idée de la circularité qui apparaît dans la démarche démonstrative que Gödel a mise en évidence. Cette association généreuse qui a imprégné de poésie les théorèmes de limitation, a conduit à certaines conséquences navrantes. Gödel est apparu à l'image d'un magicien, d'un représentant de l'ésotérisme logique, «qui a réussi à maîtriser le dire sur les choses dans les choses elles-mêmes», a-t-on pu lire. Une certaine gödelite s'est alors propagée, pour emprunter le terme à Girard (1989). L'engouement que cet ouvrage a suscité s'explique certes par le talent de son auteur, mais également par le fait que les présentations des démonstrations des théorèmes de limitation qui se veulent accessibles, invitent, par leur sobriété, l'imagination à se mouvoir dans le mystérieux, l'incompris. On n'y révèle jamais les mécanismes démonstratifs, on ne fait que les suggérer: pour le théorème d'incomplétude, Dubarle consacre deux pages, Nagel et Newman, six pages. Il faut admettre que Gödel lui-même n'a pas été très généreux, puisque sa démonstration occupe l'espace de quelque vingt-cinq pages! Il n'en demeure pas moins que la responsabilité de l'auteur de *Gödel*,

Escher, Bach est grande dans ce qui a pu être dit des résultats gödeliens. Et même s'il a réussi à faire rêver sur le thème logique - en soi il s'agit d'un exploit -, il a ouvert la porte à des suggestions épistémologiques douteuses. La réflexion s'est effacée devant des slogans. En voici quelques uns:

- Il est impossible de démontrer la non-contradiction de l'arithmétique (on oublie un peu vite Gentzen, et surtout le cadre dans lequel les théorèmes de Gödel prennent sens).
- On mentionne la ruine de l'école de Hilbert, et sa fameuse collère lorsqu'il prend connaissance du théorème d'incomplétude.
- Le Figaro parle même d'une bombe mathématique, il n'en fut rien!
- Albert Ducrocq affirme, comme si cela était évident, que le théorème de Gödel introduit une notion de relativité mentale.
- On apprend qu'il y a davantage de vérités que de preuves, on le chante de différentes manières: il y a au plan de l'évidence des propositions vraies qu'au plan de la certitude nul ne parviendra à saisir.
- On parle également de l'inadéquation de la formalisation pour capter l'intuition.
- On apprend que Gödel a prouvé que nous ne sommes pas des machines.
- On nous dit que le théorème de Gödel se réduit simplement à la compréhension du théorème du point fixe.

Il est indispensable d'aborder ces aphorismes avec un esprit critique, mais il est surtout nécessaire de les confronter au test de la réalité, de celle que constitue la démonstration des théorèmes d'incomplétude de Gödel. C'est la raison pour laquelle Marie-Jeanne Borel et Henri Volken (Université de Lausanne), et Denis Miéville (Université de Neuchâtel) ont organisé l'an passé un IIIème cycle de formation post-grade qui voulait offrir un parcours complet des démonstrations relatives aux théorèmes de limitation. C'était aussi une manière de mettre en évidence une certaine beauté que possèdent ces démonstrations, une beauté que leur confère le génie humain qui les a inventé. En effet, la ma-

nière de s'inspirer du paradoxe, tout en se donnant les moyens de l'éviter, est déjà surprenante, mais que dire de la démarche qui inscrit la complétude pour la logique des prédicats, et qui utilise des procédés de construction inattendus, qui d'expansions en extensions, réalisent conceptuellement des systèmes catégoriques, qui découpent dans la syntaxe des modèles internes? Mais, il y avait davantage dans notre volonté de présenter l'oeuvre de Gödel que son autorité marquée dans l'histoire de la pensée.

Un Gödel revisité

Comment Gödel a-t-il réagi après avoir démontré son théorème d'incomplétude? Gödel a-t-il vraiment dit quelque chose sur ce qu'il a fait qui montrerait qu'il n'a pas compris la portée de ce qu'il a fait, comme l'a prétendu Wittgenstein? Et qu'elle était la philosophie de Gödel? Était-ce par goût du formalisme qu'il s'est plongé dans ces démonstrations? Est-il formaliste? Certes non! On a écrit que le point de vue de Gödel s'oppose aussi bien au formalisme qu'à l'intuitionnisme et à l'empirisme. On pourrait penser que, homme de Vienne, Gödel ait quelque affection pour le positivisme logique et la philosophie analytique. Il tenait dit-on en piètre estime cette forme de pensée. Dans quel champ conceptuel se situait-il donc? Le théorème de Gödel a-t-il eu des conséquences sur la philosophie de l'esprit? L'intérêt que l'on porte aujourd'hui aux choses de l'intelligence artificielle offre-t-il une nouvelle lecture des théorèmes critiques de Gödel?

Les résultats critiques de Gödel constituent l'un des thèmes centraux d'une réflexion épistémologique sans cesse renouvelée. Il est reconnu, et de manière incontestable, que Gödel a jeté sur l'ensemble des mathématiques une lumière nouvelle, qu'il a fourni à sa discipline un corpus de concepts, des méthodes et des résultats dont elle tire à ce jour une bonne part de sa substance. Il a très largement contribué à former, à côté des mathématiciens appliqués, et des mathématiciens purs, une troisième race de mathématiciens - pour reprendre l'expression de Demazure -, ceux qui sont imprégnés de logique et d'informatique. Il a contribué à

constituer la logique en art d'inventer des mathématiques, comme l'a écrit Mme Sinaceur dans son ouvrage *Corps et Modèles* (1991: 380). Mais comment dire et comment lire Gödel? Pour clore notre cycle de formation postgrade, il nous a semblé indispensable de réunir un groupe de savants et d'experts pour saisir davantage l'importance du message gödelien, et ceci à la lumière d'une réflexion qui a su prendre distance avec l'événement lui-même. C'est une ouverture sur une aventure intellectuelle nouvelle que nous voulions offrir, de celle qu'anticipait déjà le Révérend Père Dubarle à la fin des années cinquante, lui qui écrivait cette très belle page:

«Les formalismes logico-mathématiques que nous sommes capables de mettre sur pied sont du genre des herbiers. Ils nous font voir distinctement un état de la pensée mathématique. Ils ne nous montrent pas directement ces virtualités antérieures aux états mêmes, ces virtualités préscientifiques que la pensée mathématique vivante et en instance d'évolution emporte avec elle comme son plus précieux capital d'avenir. Leur fonction est précisément *de ne pas* les représenter, de les faire perdre au clair concept de la science afin de mieux faire sentir, comme par un choc en retour, l'originalité et le prix de ce qui est absent de ce clair concept. Que ce quelque chose existe comme une source indéfinie de renouvellement, ils l'attestent d'une façon indirecte. Les théorèmes critiques de Gödel, Church, Löwenheim-Skolem sont des théorèmes qui forcent à comprendre la non-autarchie de ces formalismes, l'impossibilité qu'ils ont de se clore de façon suffisante sur eux-mêmes. Les formalismes manifestent alors, pour ainsi dire, l'adhérence du résultat qu'ils sont à un indéfini qui n'est *pour eux* que néant et obscurité, mais d'où montera leur renouvellement essentiel. *Pour l'esprit du mathématicien* qui vit sa pensée, l'indéfini en question n'est en effet nullement néant et obscurité, mais faculté indéfinie et encore inexplorée des clartés ultérieures. Le bienfait des formalismes à cet égard est d'apprendre où il faut désormais regarder pour entrer dans le radicalement nouveau, dans le véritable imprévu qui seul fait les grandes conquêtes.

La logique de la pensée mathématique paraît ainsi s'être conquise à la fois en vertu d'une espérance naïve de la rationalité mathématique et comme en vue de démontrer à l'esprit le succès bien relatif de cette es-

pérance, c'est-à-dire au fond sa catastrophe. Mais les résultats de cette conquête ne sont pourtant pas aussi négatifs qu'il peut le paraître au terme de cette démonstration (...). Un certain départ se fait. Ce qui passe des mathématiques et de leur logique dans les formalismes c'en est précisément le passé, si riche soit-il encore d'avenues à parcourir et incomplètement parcourues. Mais la nette conscience de ce passé, une fois bien vu qu'il ne saurait être tout, permet d'aborder le tout avec des yeux renouvelés, formés à l'exploration de l'avenir, aussi bien ouverts à la création mathématique la plus essentielle que préparés à la logique, inédite encore, de ces créations qui viendront». (1957: 84-85)

Dans cette perspective, Madame Hourya Sinaceur (Université de Paris I) et Messieurs Evandro Agazzi (Université de Fribourg), Carlo Celluci (Universita degli studi di Roma «la Sapienza»), John Corcoran (State University of New York, at Buffalo), Jean-Paul Delahaye (Université de Lille), Jacques Paul Dubucs (Université de Paris I), Henri Lauener (Université de Berne), Richard Vesley (State University of New York, at Buffalo), ont présenté un Gödel revisité, pensé aujourd'hui, d'une rare qualité. A travers ces pages, j'invite le lecteur à suivre à son tour cette visite que les conférenciers à ce colloque nous ont offerte, et qui en ont fait un événement.

*Séminaire de logique
Faculté des lettres
Espace Louis-Agassiz 1
CH - 2000 Neuchâtel*

Références bibliographiques

- DUBARLE, R. (1957). *Initiation à la logique*. Paris: Vrin.
- FREGE, G. (1879). *Begriffsschrift*. Halle: Nebert.
- GIRARD, J.-Y. (1989). In: E. Nagel *et al.* (eds) 1989.
- HOFSTADTER, D. (1979). *Gödel, Escher, Bach*. Hassocks: The Harvester Press.
- KOTARBINSKI, T. (1964). *Leçons sur l'histoire de la logique*. Paris: PUF.
- NAGEL, E., NEWMAN, J.R., GÖDEL, K., GIRARD, J.-Y. (1989). *Le théorème de Gödel*. Paris: Seuil.
- SINACEUR, H. (1991). *Corps et Modèles*. Paris: Vrin.

DU ROLE DE L'ANALYSE DES CONCEPTS SELON GÖDEL ET DE SON RAPPORT A LA THEORIE DES MODELES

Hourya Sinaceur

Il y a au moins trois raisons de s'intéresser au thème philosophique de l'analyse des concepts dans l'œuvre de Kurt Gödel. 1°) Il s'agit d'un thème récurrent quoiqu'il ne soit pas amplement développé; de 1944 à 1974 il apparaît dans trois articles différents, une note sur l'analyse non standard et deux lettres adressées à Hao Wang (Gödel 1944, 1947, 1958, 1974 et Wang 1974: 8-11). 2°) Ce thème a exercé une influence assez profonde sur la philosophie et les réflexions de certains logiciens contemporains, et non des moindres, Georg Kreisel par exemple (et nombre de ses élèves) ou Solomon Feferman. 3°) Enfin et surtout, l'étude de ce thème permet d'apercevoir l'amorce, dans la pensée de Gödel, d'un tournant décisif de la conception moderne de la logique, qui cesse, dans ces années 1930, de se présenter comme un fondement et préfère constituer, pour les mathématiques, une méthodologie ou même un «art d'inventer».

Après un examen chronologique des textes d'après lesquels nous tâcherons de décrire ce que Gödel entendait par «analyse des concepts», nous montrerons que, de façon parallèle et relativement ou totalement indépendante, Alfred Tarski a proposé une «reconstruction mathématique» en vue de transformer l'analyse conceptuelle en une «science déductive». C'est ainsi que naît la théorie des modèles dont le développement a grandement contribué à faire de la logique une branche mathématique comparable aux branches traditionnelles et capable d'interaction avec elles.

1. Cinq caractères de l'analyse des concepts selon Gödel

1. Dans la première page de son article sur la logique de Russell (1944), Gödel oppose «l'analyse des preuves» aux «théorèmes» en remarquant que Frege n'a tiré de sa soigneuse analyse de preuves que les propriétés les plus élémentaires de la suite des entiers, tandis que Peano a obtenu une grande collection de théorèmes mathématiques dans le nouveau langage symbolique, mais sans preuves. Il fallut, écrit-il, les *Principia Mathematica* pour dériver effectivement de larges parties des mathématiques d'un très petit nombre de concepts et d'axiomes logiques. Son but à lui, Gödel, n'est pas d'entrer dans le détail du formalisme ou du contenu mathématique des *Principia* mais d'analyser les concepts et axiomes sous-jacents à la logique mathématique, je dis bien à la logique mathématique en général, à laquelle les *Principia* ont apporté des contributions essentielles mais partielles. Donc cet article sur la logique des *Principia* est en réalité un article sur la logique mathématique en général.

D'ailleurs, Gödel commence par rappeler dans les premières lignes la dualité de cette discipline, d'une part secteur des mathématiques traitant de classes, de relations, de combinaisons de symboles au lieu de traiter de nombres comme l'arithmétique, de fonctions comme l'analyse ou de figures comme la géométrie, et d'autre part, science «fondamentale» s'il en est, «science des principes sous-jacents à toutes les autres sciences». C'est de la logique en ce deuxième sens que s'est occupé Russell après Frege, et, ajoute Gödel, après Leibniz qui nous a laissé le paradigme d'une *characteristica universalis*, c'est-à-dire d'une langue symbolique universelle. Et c'est de la logique en ce deuxième sens que s'occupe Gödel dans cet article (et dans d'autres réflexions sur l'histoire de la logique ou dans certains commentaires de ses propres résultats). Première constatation: la logique en ce deuxième sens est étroitement liée à ce qu'il appelle «l'analyse des concepts».

2. Deuxième point: il est significatif que l'analyse des concepts soit unie par lui à l'idée de *characteristica universalis*, évoquée en commençant et de nouveau dans la conclusion de cet ar-

ticle de 1944. Evocation «des plus frappantes», mais aussi des plus «énigmatiques» selon Ch. Parsons qui a écrit la note introductive dans le volume II des *Collected Works*. C'est ce caractère énigmatique que nous allons tâcher d'élucider ici, espérant jeter quelque lumière sur un tournant décisif de la logique mathématique.

Gödel (1944) est une critique de la «no class theory» de Russell. Il est donc naturel que presque tout l'article tourne autour de la question de savoir ce que sont les concepts¹ mathématiques et logiques, s'il ont une réalité objective ou ne sont que des façons de parler, et si on doit les accepter ou en faire l'économie dans une théorie logique des mathématiques. L'expression «analyse des concepts», ou simplement «analyse», ou encore «élucidation», «clarification» a exactement 8 occurrences. Ce n'est pas beaucoup mais c'est déjà suffisant pour un auteur aussi concis et économe que Gödel, à constituer un «motto» de l'article. Comme exemples tirés des *Principia*, il évoque l'analyse des descriptions définies et l'analyse des paradoxes (qui montre que nos intuitions immédiates des notions de vérité, de concept, de classe, etc., sont contradictoires). L'analyse des concepts est donc l'héritière directe des travaux menés dans la perspective de la recherche de fondements logiques pour les mathématiques. Comme le souligne Gödel au début de son article, celle-ci ne produit pas de théorèmes, mais clarifie la base sur laquelle ils reposent. La théorie simple des types, convenablement interprétée, et la théorie axiomatique des ensembles, ont fourni une telle base. Pourtant les «concepts primitifs» de l'une ou l'autre demeurent insuffisamment élucidés. Et Gödel, à la fin de l'article, soupçonne que c'est cela «qui est responsable du fait que la logique mathématique est demeurée jusqu'alors si en deçà des espérances de ceux, Peano entre autres, qui en attendaient (conformément aux prétentions de Leibniz) une facilitation des mathématiques théoriques identique à celle que le système décimal des nombres a permise pour les calculs numériques». Le soupçon, la supputation de Gödel répond ainsi à une exigence dont elle explique la non-satisfaction, du moins jusqu'alors. Cette exigence c'est que

1 Sur la distinction entre «concept» et «ensemble» cf. Wang 1987, p. 306-307, 309.

la logique mathématique doit «faciliter» les mathématiques, permettre de «résoudre de façon systématique des problèmes par une simple analyse des concepts impliqués dans ces problèmes» (Gödel 1944, C.W., II, p. 140).

Exigence éminemment leibnizienne, comme le souligne Gödel, qui ne prend pas la *characteristica universalis* pour une utopie dépassée, et rappelle que Leibniz la tenait responsable de toutes ses découvertes mathématiques. J'ai mené antérieurement une sorte d'enquête systématique quoique non exhaustive à travers l'œuvre logique, mathématique et philosophique de Leibniz sur le thème de la fécondité mathématique de la caractéristique et de l'analyse logique des concepts, ou, pour reprendre une façon leibnizienne de s'exprimer, sur la logique considérée comme un art d'inventer des mathématiques (Sinaceur 1988b, 1989a, 1989b, 1990a). Cette enquête était largement motivée, dans une lecture rétrospective de Leibniz, par diverses indications laconiques de Gödel, dont je viens de citer celles qui sont chronologiquement les premières dans ses écrits publiés. Il restait à prendre ces indications pour objet même de l'étude au lieu d'y puiser seulement le motif d'une relecture de Leibniz. C'est ce que je fais ici. Nous allons voir que contrairement à ce qu'écrit Parsons dans son introduction (C.W., II: 117-118), la tonalité leibnizienne des écrits de Gödel ne s'est pas estompée avec le temps. Au contraire, elle a donné lieu, dans le style lapidaire et incisif propre à Gödel, à un véritable leitmotiv, dont témoignent d'ailleurs des personnalités plus ou moins proches de Gödel, Hao Wang (1987)² et Simon Kochen (1976: 315) entre autres. On sait aussi que le *Nachlass* (Dawson 1991: 2, 4) contient des notes abondantes sur les écrits de Leibniz.

3. Revenons au texte de 1944. Après avoir opposé, au début, analyse des concepts et production ou dérivation de théorèmes, Gödel les entrelace dans ses remarques finales. L'ambition de la logique symbolique: résoudre systématiquement des problèmes mathématiques en analysant les concepts qui y occurrent, n'est pas une vaine ambition. L'importance de l'œuvre mathématique

2 Index, entrée 'Leibniz' et plus particulièrement p. 74, 165, 210, 226, 298 («Gödel semble suggérer que lorsque la théorie des concepts [la logique] sera développée adéquatement, elle pourra aider les mathématiques»).

d'un Leibniz en témoigne assez pour convaincre même un Poincaré. Mais il est raisonnable de commencer par une analyse des concepts primitifs³ et des axiomes où ils sont mis en œuvre. Il n'est pas absolument clair dans ces dernières lignes si Gödel conserve l'idée d'un fondement (au sens d'une réduction) logique des mathématiques — comme il l'exprime au début en disant en disant que les *Principia* ont permis de dériver une grande partie des mathématiques de quelques concepts et axiomes logiques — ou s'il pense seulement à l'utilité mathématique d'une analyse logique de ce qui constitue la base des théories mathématiques: concepts primitifs et axiomes. Autrement dit — et c'est notre troisième point — on a là sans doute le premier glissement chez Gödel de l'idée classique de fondement vers l'idée, moderne, de fondation, qui veut que la recherche fondationnelle est 1°) interne aux mathématiques et 2°) devrait déboucher sur des résultats proprement mathématiques. On trouve un indice de cette attitude dès 1929. A la fin de l'introduction à son mémoire sur la complétude du calcul des prédicats du premier ordre, Gödel souligne l'indépendance du problème qu'il a résolu par rapport à la controverse sur les fondements des mathématiques.

Même si on n'avait jamais douté de la valeur [Geltung] du contenu de la mathématique 'naïve', écrit Gödel, le problème de la complétude (contrairement à celui de la consistance) aurait pu être judicieusement [sinnvoll] posé à l'intérieur de cette mathématique naïve, parce qu'une restriction des moyens de preuve ne semble pas plus pressante ici que pour n'importe quel autre problème mathématique» (C.W., I: 64; c'est Gödel qui souligne).

C'est prendre le contre-pied de Hilbert (1922: 161; Sinaceur 1991a: 308-310), pour qui il s'agissait d'un problème extérieur aux mathématiques, et amorcer un retournement qui deviendra, chez Tarski, la base d'un nouveau point de vue (Sinaceur 1991a: 317-318). C'est aussi prendre le contre-pied de Frege (et de Russell) pour qui l'élucidation des concepts mathématiques

3 Wang 1987 (p. 309) note que Gödel n'a jamais donné une liste complète de ce qu'il considère comme «concepts primitifs» de la logique.

fondamentaux sert seulement à nous «éclairer»⁴ sur nos objets mathématiques les plus familiers. S'il est vrai qu'il s'agit d'une activité engendrée *de l'intérieur des mathématiques*, alors il serait naturel qu'elle rejaillisse sur elles. Pour Gödel clarifier n'a pas seulement un objectif théorique: élucider les concepts qui font difficulté devrait conduire à la production *systématique* de théorèmes.

Que veut dire «systématique»? Si l'on revient à la phrase de Gödel qui comporte cet adjectif («résoudre de façon systématique des problèmes par une simple analyse des concepts impliqués dans ces problèmes»), on verra que «systématique» veut d'abord dire *a priori*, puisqu'il s'agit, en attaquant un problème, d'en analyser les concepts et les termes, *avant même* de se mettre en quête des éléments d'une solution. En somme, la simple analyse (logique) de l'énoncé indique la voie d'une solution. N'est-ce pas là seulement l'extension grâce aux techniques logiques d'une exigence méthodologique triviale dans la pratique mathématique? Mais «systématique» ne le dit-on pas aussi d'une solution *uniforme* pour un ensemble de problèmes de même type? Certes. Il faut alors supposer que l'analyse des concepts met au jour des parentés profondes, l'identité structurale, de problèmes distincts. L'on est bien en droit de retenir ce second sens, Gödel ayant clairement invoqué le patronage de Leibniz qui voyait dans la caractéristique le moyen de traiter aussi bien *intrinsèquement* que *systématiquement* des problèmes apparemment divers. Leibniz cherchait à réduire la diversité des matières ou des expressions à l'unité des formes symboliques. Pourtant, ce second sens ne va pas tout à fait de soi, s'agissant de Gödel dont les résultats (de 1931) nous enseignent qu'il n'est pas de langue (caractéristique) universelle et qu'il n'est même pas toujours possible, dans un langage (formel) donné, de résoudre systématiquement la question de la décision pour *toutes* les propositions de ce langage. Dans son commentaire de Gödel 1944, Hao Wang (1987: 311) a du reste avoué ne pas bien comprendre ce que Gödel entendait suggérer. Il me semble pourtant qu'il y a une façon naturelle de comprendre cet emploi de «systématique». Gödel n'était pas

4 *Die Grundlagen der Arithmetik, Einleitung*, trad. fr., Paris: Seuil, p. 116.

moins subtil que précis; il en résultait chez lui des positions non tranchées quoique très bien définies. Il en appelle à Leibniz, mais il ne s'identifie pas à Leibniz. Celui-ci avait foi dans la recherche de méthodes dérivées de l'établissement d'une caractéristique universelle. Les résultats de Gödel ne ruinent pas cet optimisme mais forcent à préciser les limites dans lesquelles on peut le *maintenir*, c'est-à-dire à la fois le restreindre et le garder. L'espoir peut être maintenu de trouver des méthodes générales *relativement à une* catégorie de problèmes formulés dans *un* langage donné. Attitude nuancée, comparable à celle de son commentaire au deuxième résultat d'incomplétude du mémoire de 1931. Rappelons que Gödel fait observer que son résultat négatif n'interdit pas de chercher une démonstration finitaire de consistance de l'arithmétique élémentaire, mais seulement d'abandonner l'espoir de pouvoir faire cette démonstration avec des moyens internes au système considéré. Ici, il suggère que la logique peut bien jouer, dans des limites à préciser, le rôle d'*ars inveniendi* que lui assignait Leibniz. Cette interprétation est confortée par les remarques de Gödel à propos de l'*Analyse non standard* d'Abraham Robinson⁵: celle-ci offre l'avantage de «simplifier de manière substantielle les preuves, non seulement pour des théorèmes élémentaires mais aussi pour des résultats profonds»; or, la simplification des preuves «facilite la découverte» écrit-il (1974, C.W., II, p. 311) d'une façon qui reprend exactement l'expression et les préoccupations de son article de 1944. Autrement dit, comme le confirme par ailleurs Simon Kochen (1976: 315), la théorie des modèles à la manière de Robinson (et, ajoutons-nous, de Tarski), paraît à Gödel offrir une réalisation de l'ambition leibnizienne de constituer la «vraie logique» en *ars inveniendi*. Nous illustrerons plus précisément ce jugement de Gödel dans la deuxième partie de cette étude.

4. Le thème de l'analyse des concepts, des procédures et des résultats est peut-être plus prégnant dans Gödel (1947), dont la section 3 porte le titre: «Reformulation du problème [du continu]

5 Wang 1987, qui pense qu'il «n'est pas facile de voir comment la logique mathématique telle qu'elle est pratiquée aujourd'hui pourrait être étendue de manière à fournir de puissantes méthodes (ou seulement des guides efficaces) pour de nouvelles découvertes» (p. 311), ne fait précisément pas état de ces remarques de Gödel.

sur la base d'une analyse des fondations⁶ de la théorie des ensembles et des résultats obtenus dans les différentes perspectives». Gödel y affirme la nécessité, pour une solution complète du problème du continu, d'une «analyse plus profonde (qu'il n'est habituel de faire en mathématiques) des significations [meanings] des termes ... "ensemble", "correspondance un à un"⁷, etc.». Avec l'entrée en scène des termes "sens" et "signification", l'analyse des concepts devient une analyse des significations — c'est notre quatrième point.

L'axiomatique de la théorie des ensembles a transformé en un problème de pure combinatoire de signes la question de savoir si telle proposition dérive des axiomes retenus, mais ces signes ont une signification. «Même l'intuitionniste le plus radical doit reconnaître qu'ils ont une signification [meaningful]» écrit Gödel (1947, *C.W.*, II: 181). Cette phrase n'a pas un sens absolument évident. Hao Wang (1987: 297), par exemple, voit une ambiguïté dans l'usage de l'adjectif "meaningful". Pour moi la difficulté est ailleurs. La présence d'une ou plusieurs significations inscrites dans le symbolisme d'un système axiomatique — à ne pas confondre avec système formel, comme le fait remarquer Wang ailleurs dans son livre (par exemple, p. 247). — est en effet une assumption indispensable pour qui veut entreprendre une analyse des significations et en tirer des résultats. Gödel se démarque donc (implicitement) ici du formalisme pour lequel, formalisées, les mathématiques seraient des combinaisons finies et discrètes de signes matériels.

«Comment peut-on penser à *exprimer* la métamathématique dans des systèmes mathématiques eux-mêmes, écrira-t-il en 1967, si on considère que ceux-ci consistent en symboles sans signification qui

6 Gödel entend par là les différentes axiomatiques de la théorie des ensembles et non les différentes attitudes dictées par la conception (philosophique) qu'on se fait de la nature des objets mathématiques (réalisme, intuitionnisme, semi-intuitionnisme, etc.)

7 On peut bien dire que Tarski s'est, de son côté, livré à une analyse de la relation de correspondance bijective dans sa théorie des algèbres cardinales et ordinales. Commencée dès 1926, elle aboutit, entre autres, à la publication d'un ouvrage en collaboration avec B. Jónsson en 1949 (*Cardinal algebras*, Oxford University Press).

n'acquièrent quelque substitut de signification que par l'intermédiaire de la métamathématique?»⁸

Ce commentaire *a posteriori* fait voir clairement une chose: traduire un langage dans un autre, comme Gödel l'a fait en codant par des entiers ses énoncés métamathématiques sur l'arithmétique, *présuppose* que ce langage ait une signification. D'une façon analogue, Tarski associe le fait que les énoncés ont une signification et le fait qu'on puisse *interpréter* un langage dans un autre, c'est-à-dire définir les concepts primitifs d'une théorie dans une autre. Une grande partie de l'activité mathématique aussi bien que métamathématique consistant à user de la méthode des interprétations (par exemple, pour des preuves d'indécidabilité), il n'est pas question de croire qu'on manipule des signes vides de sens⁹. Pousser ainsi sur le devant de la scène mathématico-logique les catégories associées de signification et d'interprétation est même une des idées-forces qui ont donné naissance à la théorie des modèles.

Gödel se démarque aussi (explicitement), dans son texte de 1947, de «l'intuitionnisme radical», mais seulement dans la mesure où pour celui-ci les significations ne sont pas objectives mais subjectives. En dépit de la tournure raccourcie de sa phrase, Gödel ne reproche pas aux intuitionnistes de vouloir manipuler des signes sans signification mais de rapporter celle-ci à l'activité du mathématicien. Gödel (1958, C.W., II: 244) ne doute pas que «les objets de la logique intuitionniste sont des énoncés et des preuves pourvus de signification», ce qu'il conteste c'est que

8 Lettre à Hao Wang du 7 décembre 1967, dans Wang 1974, p. 8-9 (les soulignements sont de Gödel).

9 Tarski 1936/1960 explique cela très simplement et très clairement: «Si, dans la construction d'une théorie, on se comporte comme quelqu'un qui n'aurait pas compris la signification des termes de cette discipline, ce n'est pas du tout la même chose que de dénier à ces termes toute signification. C'est un fait que nous développons quelquefois [donc pas toujours!] une théorie déductive sans attribuer un sens défini à ses termes primitifs, traitant alors ces derniers comme des variables; dans ce cas nous disons que nous traitons la théorie comme un SYSTÈME FORMEL. Mais cette situation... ne se produit que lorsqu'il est possible de fournir plusieurs interprétations du système axiomatique de la théorie, c'est-à-dire s'il y a plusieurs moyens qui s'offrent d'attribuer des significations concrètes aux termes qui figurent dans la théorie, mais que nous n'avons pas l'intention d'accorder d'avance la préférence à aucune de ces significations particulières» (p. 113-114).

cette signification dépende de notre esprit ou de nos constructions. Ou que, comme il l'a écrit explicitement, la métamathématique soit considérée comme une «théorie de l'activité humaine consistant à manipuler des symboles» plutôt que comme une «science décrivant des situations mathématiques objectives» (Wang 1974: 10). Bref ce qui est opposé à «l'intuitionnisme le plus radical» c'est le fameux réalisme gödelien — sur lequel je ne m'étendrai pas ici.

Un passage du début de Gödel (1958, *C.W.*, II: 240, auquel je viens d'emprunter l'explicitation sur la logique intuitionniste) permet de comprendre plus clairement cette insistance sur la présence de significations dans les symboles. Si l'on comprend par «mathématique finitaire» la mathématique de l'évidence *intuitive*, alors il faut reconnaître, observe Gödel, qu'une preuve de consistance de l'arithmétique a besoin de certains concepts *abstrait*¹⁰. Les concepts abstraits renvoient à des «constructions mentales» [Denkgebilde]: démonstrations, énoncés pourvus de signification [sinnvolle Aussagen], etc. Dans les preuves relatives à ces constructions mentales interviennent des «vues» [Einsichten, insights] suggérées¹¹ non par la combinaison matérielle de signes représentant les concepts abstraits mais par la signification [Sinn] de ces signes. On retrouve la triade concept-signé-sens, dont les éléments ne semblent pas pouvoir, pour Gödel, être dissociés. Renversant le point de vue de Hilbert pour qui *seul* un processus fini a du contenu, on pourrait dire que *même* un formalisme finitaire, une pure combinatoire de signes, a un contenu [Inhalt]; la réflexion sur ce contenu conduit à des concepts abstraits (Gödel 1958, *C.W.*, II, p. 242, note 4). C'est parce que le contenu est inéliminable que l'analyse est à la fois possible et fructueuse. Les commentaires de Gödel sur les axiomes forts de l'infini laissent penser que pour lui — c'est-à-dire d'un point de vue réaliste — le contenu est même en quelque sorte inépuisable. Les nouveaux axiomes «ne font que déployer le contenu du concept d'ensemble» écrit-il (1964, *C.W.*, II: 261) en effet dans un style dont il convient de remarquer la tonalité très

10 Soulignements de Gödel.

11 Dans la version anglaise, Gödel emploie l'expression, plus forte, de «dérivé de» (Gödel 1972, *C.W.*, II, p. 273).

leibnizienne, même si pour expliquer notre connaissance intuitive de ce contenu Gödel recourt, dans le supplément à la seconde édition de son texte de 1947, à l'idée kantienne de synthèse rationnelle du divers.

5. «Analyser» ne se réduit d'ailleurs pas à «clarifier», qui n'a qu'une valeur contemplative. Analyser a une valeur opératoire, comme l'indique nettement la note 5 de Gödel (1972C.W., II: 275). Gödel y remarque que l'expression «fonction mécaniquement calculable» avait dès avant les travaux de Turing un sens *clair* mais *inanalysé*. Observons que l'analyse est, dans ce cas, plutôt ou autant le *produit* que l'origine du passage d'une notion intuitive (plus ou moins vague) à une définition mathématique précise. Quoiqu'il en soit, analyser c'est *préciser mathématiquement* un concept - tel est notre cinquième point. Mais tous les concepts mathématiquement précis ne constituent pas une «analyse» de la notion intuitive correspondante; par exemple, «la notion de démontrabilité formelle n'est pas une *analyse*¹² du concept de vérité mathématique» (Wang 1974: 10).

Revenons au problème du continu. Il se peut qu'il y ait de nouveaux axiomes de la théorie des ensembles que nous reconnaitrions comme impliqués par les concepts sous-jacents à la logique et aux mathématiques, si nous avions de ces derniers une compréhension plus profonde (Gödel 1947, C.W., II: 182). Commençons donc par les analyser. Nous reconnaitrions qu'il existe au moins deux interprétations du terme «ensemble», c'est-à-dire deux classes d'objets satisfaisant les axiomes de la théorie des ensembles. L'une est constituée de multitudes arbitraires; l'autre de multitudes qui sont des extensions de propriétés définissables. Que, dans cette deuxième interprétation (où Gödel suppose que nous savons ce qu'est une propriété définissable), les ensembles soient identifiés à des propriétés définissables n'est ni énoncé explicitement ni contenu implicitement dans les axiomes de la théorie des ensembles. Cela laisse prévoir, conclut Gödel, qu'un nouvel axiome susceptible de décider l'hypothèse du continu, devrait «impliquer quelque chose sur la définissabilité des ensembles» (*Ibid.*, p. 184). Gödel a lui-même *vérifié*

12 Souligné par Gödel.

cette supposition, toute *théorique* quand elle découle comme ici d'une analyse du concept d'ensemble, en montrant que l'hypothèse généralisée du continu est dérivable de l'axiome: «tout ensemble est définissable en termes de notions primitives de la théorie des ensembles, c'est-à-dire, en un mot, est constructible», ce que nous écrivons aujourd'hui par la formule $V = L$ où L dénote l'univers des constructibles et V celui des ensembles. Gödel nous montre donc sur un exemple tiré de ses propres travaux techniques, logiques au premier sens distingué dans Gödel 1944, comment l'analyse des concepts — la logique au second sens — permet de *prévoir*¹³ un résultat. Les ensembles constructibles sont, «d'un point de vue purement mathématique» et «indépendant de la question philosophique de savoir si les définitions non prédicatives sont admissibles», une application «féconde» de la théorie ramifiée des types de Russell¹⁴. Cependant, Gödel ne nous confirme pas si lui-même avait *effectivement* trouvé son résultat grâce à une analyse du concept d'ensemble dérivée de la théorie des types, ou si une réflexion *a posteriori* sur son résultat lui a suggéré cette analyse, dont le statut est *logiquement* antérieur au résultat lui-même, ou si, troisième éventualité, les deux processus d'analyse et d'invention furent plus ou moins concomitants. Quoi qu'il en soit, on peut dire qu'*en principe*, l'analyse du concept d'ensemble, adossée à la théorie des ordinaux combinée avec la notion logique de définissabilité, conduit à une demi-solution du problème du continu: la consistance avec les axiomes habituels de la théorie des ensembles; de même l'analyse des concepts et de leurs significations nous met sur la voie de nouveaux axiomes dont découlera probablement, pense Gödel, la fausseté de l'hypothèse du continu.

Résumons les acquis de notre lecture de Gödel: l'analyse des concepts est une activité logique mathématiquement féconde; elle

13 Il est plus intéressant, pensait Leibniz, de prévoir que de trouver un résultat. Sans en appeler à Leibniz, Kreisel a beaucoup commenté et pratiqué pour son propre compte le point de vue de Gödel; il a lui-même souvent insisté sur l'utilité *technique* de l'analyse des concepts (cf. Sinaceur 1991b).

14 Gödel 1944, *C.W.*, II, p. 136 : «The theory of orders proves more fruitful if considered from a purely mathematical standpoint, independently of the philosophical question whether impredicative definitions are admissible...».

consiste en une réflexion sur des contenus mathématiques et cette réflexion est utile dans la recherche systématique de résultats; elle peut réussir simultanément à déployer la signification de ces contenus et à la préciser mathématiquement.

2. L'analyse des concepts comme science déductive

Du point de vue technique, il y a de multiples points de contact entre les travaux de Gödel et la théorie des modèles, ne serait-ce d'abord que l'utilisation centrale dans les premiers de la notion logique de modèle, formellement définie par Alfred Tarski en 1936. Bien entendu, c'est le parallélisme entre les concepts de décidabilité et de définissabilité qui donne lieu à la convergence la plus fine entre les résultats de Gödel et ceux de Tarski. On peut montrer aussi, et c'est ce que je me propose de faire maintenant, une affinité également profonde, en dépit d'éventuelles divergences philosophiques (d'ailleurs non radicales), sur leur conception du rôle de la logique par rapport aux mathématiques.

1. *L'analyse des concepts comme méthodologie des sciences deductives.*

On ne trouve pas dans les travaux d'Alfred Tarski l'expression «analyse des concepts». Tarski était avare en commentaires épistémologiques sur sa façon de travailler. Sans doute aussi se voulait-il moins tributaire que Gödel d'un certain vocabulaire philosophique, leibnizien ou non. Pourtant, une grande partie de ses résultats reposent sur une analyse de concepts qui le conduit à introduire de manière explicite et *formelle* des distinctions ou des définitions inexistantes ou seulement implicites chez d'autres. Ainsi en est-il des distinctions entre les concepts de dérivabilité et de conséquence (1936c), entre ceux de définition explicite et de définition implicite (1931b, § 4), etc.; ainsi en est-il de sa définition formelle du concept de définissabilité (1935). Quant à sa fameuse analyse du concept de vérité, elle pourrait être, du point de vue de ses objectifs, mise en parallèle avec celle du concept d'ensemble par Gödel, bien qu'elle soit, au moins en volume de pages, beaucoup plus importante.

Nous voici convaincus sur la foi de ces quelques exemples d'une chose: Tarski a amplement pratiqué l'analyse des concepts. C'est, note John Barwise (1989: 396-398), un de ses traits communs avec Gödel et avec Kleene, les deux autres pères fondateurs de la logique contemporaine. Bien plus que cela, sous le nom de «Méthodologie des sciences déductives» ou «Métamathématique», il a fait de l'analyse des concepts une discipline scientifique autonome et capable d'interaction avec d'autres disciplines scientifiques. C'est Tarski qui, le premier, a fait le plus grand pas dans la réalisation du vœu leibnizien d'une «vraie logique» comme *ars inveniendi* et mis en place les conditions générales d'applicabilité de la logique non pas seulement à l'analyse des procédures mathématiques mais aussi à la *recherche mathématique*.

La source commune à Gödel et à Tarski, c'est non seulement les *Principia mathematica* de Russell et Whitehead mais encore, bien entendu la métamathématique de Hilbert. Tarski hérite de cette «mathématique nouvelle» (Hilbert 1922: 174), définit¹⁵ plus précisément et élargit le champ de ses concepts fondamentaux et, surtout, *lui donne la configuration d'une science déductive*: l'interaction avec les autres sciences déductives (les autres théories mathématiques) en découle alors comme une *conséquence*¹⁶. En somme, Tarski est remonté du but à la source: il a compris que pour faire jouer à la logique son plein rôle dans la recherche de résultats mathématiques, il fallait organiser mathématiquement non seulement la logique elle-même (ce qui était chose faite) mais encore son activité réflexive sur les théories mathématiques (les métathéories de ces théories). Grâce à lui, cette activité réflexive (analyse des concepts et des procédures mathématiques) se présente sous le visage d'une *méthodologie* plutôt que d'un fondement. Dans son *Introduction à la logique*, ouvrage destiné à un large public, Tarski définit justement la méthodologie des sciences déductives (c'est-à-dire des mathématiques) comme

15 Dans un métamétalangage, comme l'observe John Corcoran dans son introduction à la seconde édition de Tarski 1956, p. XIX.

16 Cf. Sinaceur 1991a, 4e partie, chap. 1, §§ 1 et 2, p. 306-325.

«l'analyse détaillée et l'évaluation critique des principes fondamentaux ... qui doivent s'appliquer dans la construction de la logique et des mathématiques» (1936/1960: 103 les termes et expressions soulignées le sont par nous).

Désignation somme toute banale si elle n'était assortie de l'exigence que cette méthodologie devienne

«une science générale des sciences déductives dans un sens analogue où l'arithmétique est la science des nombres et où la géométrie est la science des figures géométriques» (*ibid.*, p. 123).

Arrêtons-nous sur cette analogie établie par Tarski entre la métamathématique d'une part et l'arithmétique ou la géométrie d'autre part. Car c'est une clef de son œuvre et présente dès 1928 (Tarski 1930a, *Collec. papers*, I: 313). L'arithmétique, l'algèbre ni la géométrie n'embrassent la totalité des objets mathématiques; chacune possède un domaine bien délimité: nombres, équations ou figures et un langage spécifique: des concepts primitifs et des axiomes particuliers. Une fois ces domaines découpés dans l'univers mathématique, il n'est pas rare qu'on puisse passer de l'un à l'autre, les «interpréter» l'un dans l'autre de façon à établir une correspondance bijective entre énoncés sur les objets de l'un et énoncés sur les objets de l'autre. Ainsi Descartes nous a montré comment interpréter la géométrie dans l'algèbre. Mais nous pouvons réciproquement interpréter l'arithmétique ou l'algèbre dans la géométrie; c'est le procédé bien connu sous le nom de «méthode des graphiques». Et depuis l'apparition des géométries non euclidiennes, c'est devenu un procédé systématique d'en montrer la consistance par interprétation dans un modèle euclidien. Par ailleurs, Gödel a arithmétisé la métathéorie du système dans lequel il construit sa preuve d'incomplétude de l'arithmétique élémentaire. C'est un premier pont de passage entre métamathématique et mathématique, établi à l'occasion d'un problème dans lequel — rappelons-le — Gödel voit l'indice de l'intériorité, et non de l'extériorité, de l'une à l'autre. Rapprochons Gödel de Descartes et généralisons. Il suffit d'effacer la distinction gödelienne entre logique au sens 1 et logique au sens 2. Si nous considérons, en effet, la logique comme une portion

bien délimitée de l'univers mathématique, bien qu'elle soit en même temps une étude des procédures mathématiques, alors il y a des chances pour que le jeu de l'interprétation soit possible entre elle et n'importe laquelle des autres branches mathématiques. C'est sur ce point que s'exerce le génie propre de Tarski: non pas appliquer les mathématiques à la logique ni même seulement appliquer pour ainsi dire de l'extérieur la logique aux mathématiques, mais tout simplement considérer la logique, *à la fois au sens 1 et au sens 2*, comme l'une des nombreuses disciplines mathématiques. Pensée semi-triviale, puisqu'elle consiste à prendre la mesure d'une semi-évidence: la logique est une théorie déductive; or «toute théorie déductive est une discipline mathématique» (Tarski 1936/1960:106). La logique au sens 1? Non, pas seulement la logique au sens 2 aussi! Puisqu'il n'y a pas démarcation entre les deux¹⁷. Du moins suffira-t-il de présenter déductivement toute analyse logique d'un concept mathématique ou impliqué dans le raisonnement mathématique comme ceux de vérité et de définissabilité, c'est-à-dire toute recherche métamathématique. Tarski retrouve sans le savoir, ou tout au moins sans le dire, cette conviction bien leibnizienne: la clef du passage d'une théorie dans une autre c'est bien la formalisation, pourvu qu'on mette en évidence par ailleurs les règles de passage, c'est-à-dire d'interprétation. Cette disposition si fondamentale pour les mathématiques, il faut la prolonger à la métamathématique, qui sera formelle *et* significative. La théorie de la vérité, souligne Tarski, est une «discipline déductive... qui établit de façon axiomatique les propriétés du concept primitif de vérité»¹⁸. Barwise remarque que Tarski procède en mathématicien: il développe la notion de vérité non en général mais dans un domaine particulier, relativement à un langage donné. Il faut bien déterminer les conditions du problème pour pouvoir le résoudre. Et pour une solution précise, rien de tel qu'une formalisation soignée.

17 Tarski 1944, *Collec. papers*, II, p. 693 (Vaught observe que cette absence de démarcation découle de la définition tarskienne de la vérité, cf. Hodges 1986, p. 876. Mais on peut supposer aussi que la définition de la vérité implique l'absence de démarcation).

18 *Collec. papers*, I, 616-617.

2. *La reconstruction mathématique de la métamathématique.*

Tarski a constamment cherché à «reconstruire» mathématiquement la métamathématique, c'est-à-dire à en faire une branche des mathématiques ordinaires, ayant certes ses concepts spécifiques (déduction, vérité, conséquence, définissabilité, etc.) et un objet particulier: analyser les théories mathématiques et les concepts intuitifs auxquels elles s'adossent, mais organisée selon les mêmes principes formels et puisant sa fécondité dans un jeu analogue entre expression et intuition, théorie et modèles, syntaxe et sémantique. Depuis Boole la logique appliquait des méthodes mathématiques à l'analyse de la pensée; depuis Frege elle s'appliquait à l'analyse des concepts et des procédures mathématiques; pour cette analyse Hilbert a proposé des moyens formels (l'axiomatisation) et Gödel a montré comment elle pouvait être exprimée dans l'arithmétique tout en souhaitant qu'elle eût des retombées sur les mathématiques elles-mêmes. Tarski, enfin, a fait en sorte qu'il n'y ait plus lieu de distinguer, *même du point de vue du mathématicien ordinaire*, entre logique et mathématique. En «poussant l'analyse» (comme eût dit Leibniz) jusqu'à la formaliser et en tirer des méthodes susceptibles d'être appliquées avec fruit à un domaine mathématique spécifique: algèbre ou géométrie, il a créé un nouveau domaine de la logique mathématique: la théorie des modèles. Celle-ci nous a donné les premiers échantillons de ce qu'on peut appeler une «mathématique logique», dont le trait distinctif est l'entrelacement dont elle se constitue de méthodes logiques (analyse syntaxique et analyse sémantique) et de méthodes mathématiques (calcul, constructions géométriques, approche topologique).

J'ai analysé en détail cet entrelacement dans le cas de la preuve, par généralisation d'un théorème d'analyse numérique, de la décidabilité de l'algèbre et de la géométrie élémentaires. En d'autres termes, j'ai examiné la transformation du théorème de Sturm sur le nombre de racines réelles d'un polynôme en méthode d'élimination des quantificateurs (Sinaceur 1991a, 4^e partie, 1^e section, 304-371). Je me bornerai ici à rappeler combien Tarski tient à faire comprendre que son résultat d'élimination des quantificateurs (théorème 31 de Tarski 1951, note 12, *Collec. Papers*, III: 354) a un intérêt «même du point de vue purement

mathématique». Cette préoccupation, dont nous avons trouvé l'expression chez Gödel à propos des ensembles constructibles, est une véritable obsession chez Tarski (Cf. Sinaceur 1991a: 321). Il faut dire que le développement actuel de la théorie des modèles en relation avec d'autres disciplines mathématiques, la géométrie algébrique en particulier, lui a enfin rendu justice. Laissant donc de côté la méthode tarskienne d'élimination des quantificateurs (sur laquelle il est bien curieux que Gödel ne se soit jamais exprimé, à ma connaissance), je voudrais revenir sur la notion de définissabilité, utilisée mais non définie par Gödel, et montrer rapidement comment Tarski en a fait un concept mathématiquement opératoire.

3. La définissabilité des nombres réels

Dès 1931 Tarski a montré sur l'exemple des nombres réels et celui des ensembles projectifs (ensembles obtenus à partir d'un ensemble fermé par un nombre fini d'opérations de projection orthogonales ou de passage au complémentaire) l'intérêt pour le mathématicien de déterminer des classes d'ensembles définissables dans un langage logique du premier ordre. Soit dit en passant, il est caractéristique de la méthode de Tarski d'examiner le problème de la définissabilité dans le cadre restreint d'un langage déterminé, celui de l'arithmétique des nombres réels par exemple, *avant* d'envisager la notion générale d'ensemble élémentairement définissable (Tarski 1931a, *in fine*), et aussi avant d'attaquer le problème sous son aspect formel, généralisant et justifiant la méthode de Padoa comme instrument systématique de preuve de l'indépendance d'une constante non-logique par rapport à un sous-ensemble de l'ensemble des autres constantes non-logiques d'un langage du premier ordre (Tarski 1935). C'est le propre du créateur de la sémantique scientifique et de la théorie des modèles que le travail sur le terrain sémantique (par le biais de langages interprétés dans un domaine particulier ou dans le domaine général de la théorie des ensembles) précède le travail dans un langage formel ininterprété.

Venons-en à la définissabilité dans le modèle particulier du corps ordonné des nombres réels, c'est-à-dire aux deux mémoires de 1931. Tarski indique que la détermination de

l'ensemble des sous-ensembles définissables de nombres réels découle du lemme d'élimination des quantificateurs auquel j'ai fait allusion plus haut. Ce lemme étant la généralisation d'un théorème mathématique usuel (le théorème de Sturm), on doit pouvoir «reconstruire ... dans le cadre des mathématiques» la notion de sous-ensemble définissable de nombres réels. C'est ce que fait Tarski. Il établit qu'un ensemble de suites de nombres réels, c'est-à-dire un sous-ensemble du produit cartésien $\mathbb{R}^n$, est élémentairement définissable si et seulement s'il est combinaison booléenne finie de sous-ensembles mathématiquement définis par des équations ou des inégalités algébriques à coefficients réels. Si l'on met en valeur l'aspect *métamathématique* de cette définition, on dira qu'un sous-ensemble de $\mathbb{R}^n$ est élémentairement définissable si et seulement s'il est l'extension d'une formule sans quantificateur du langage de la théorie des corps ordonnés. Si l'on met en valeur son aspect *mathématique*, on verra qu'en particulier un sous-ensemble de $\mathbb{R}$ est élémentairement définissable si et seulement s'il est somme finie d'intervalles de la droite réelle bornés algébriquement, et un nombre réel est définissable si et seulement s'il est algébrique. Voilà bien qu'on retombe sur des notions mathématiques usuelles. Ainsi Tarski a donné l'interprétation mathématique du concept de définissabilité dans le modèle des nombres réels. Il a donc illustré concrètement les faits suivants:

1°) Prise dans un sens relatif (à un langage formel donné L) la notion de définissabilité est non seulement non contradictoire mais encore parfaitement précise. Il n'y a pas à craindre le spectre d'un paradoxe comme celui de Richard.

2°) Un peu de travail d'analyse à la fois syntaxique (de la structure des formules du langage L) et sémantique (de la signification de l'existence d'une solution commune à un système fini d'équations et d'inégalités algébriques à coefficients réels) suffit en effet à montrer que cette notion n'est pas «en dehors des limites propres aux mathématiques» (Tarski 1931a, *Collec. Papers*, I: 519), qu'elle ne «diffère en rien des autres notions mathématiques» et ne doit «éveiller ni crainte ni doute» (ibid., p. 520) chez les mathématiciens; que ceux-ci cessent donc de se

méfier de la métamathématique qui n'est que la face abstraite de la (sémantique) mathématique.

3°) Puisque nous retombons dans l'algèbre et la géométrie ordinaires, nous pourrions déduire des résultats dont nous n'aurions pas eu seulement idée si nous en étions restés au domaine purement métamathématique; c'est le croisement — et non pas seulement l'application dans un sens ou dans l'autre — des méthodes logiques et mathématiques qui est fructueux.

4°) En croisant algèbre, géométrie et logique nous définissons le concept de base d'une nouvelle discipline mathématique; le concept d'ensemble définissable de nombres réels est en effet analogue à ceux de ligne et de surface en géométrie, relativement tout au moins car ici nous établissons une définition précise à partir d'une intuition vague, là notre intuition est déjà travaillée par la précision logique (*ibid.*, p. 521). De fait cette nouvelle discipline mathématique, dont Tarski voyait distinctement les contours en 1931, existe pleinement depuis une dizaine d'années sous le nom de géométrie algébrique réelle; son concept de base, celui d'ensemble semi-algébrique, n'est autre que celui d'ensemble définissable de nombres réels de Tarski généralisé au cas d'un corps réel clos quelconque, ce à quoi les mathématiciens peu au fait de l'histoire de la logique n'ont pas pris garde. Bien plus, le résultat d'élimination des quantificateurs dans la théorie élémentaire du corps ordonné des nombres réels et des corps réels clos est systématiquement utilisé aujourd'hui pour simplifier les démonstrations; on démontre plus facilement la stabilité des ensembles semi-algébriques par projection en se servant des propriétés syntaxiques des formules qui les définissent qu'en s'en tenant aux propriétés géométriques des projections; on démontre de même leur stabilité par produit cartésien et par adhérence topologique. L'exposé du résultat d'élimination des quantificateurs de Tarski figure naturellement dans les traités de géométrie algébrique réelle actuellement disponibles. Il figure aussi et depuis plus longtemps, ce qui est plus significatif, dans un traité d'algèbre abstraite comme celui de Nathan Jacobson (1974, 1985), qui salue dans sa préface l'émergence d'une nouvelle algèbre: l'algèbre modèle-théorique. C'est assez montrer le chemin

que l'analyse des concepts métamathématiques s'est frayée, depuis une décennie ou deux, dans les mathématiques mêmes.

En conclusion, je voudrais faire observer que l'ordre d'exposition et mon intention de braquer les pleins feux sur Gödel pour un colloque consacré essentiellement à lui ont fait que j'ai parlé de Tarski comme si son travail avait pris la succession de celui de Gödel. Il n'en est rien; et j'ai averti dans mon introduction de la relative ou totale indépendance des travaux des deux logiciens. On sait, et lui-même a souvent insisté sur ce fait, que Tarski était en possession de certains de ses résultats dès les années 1926-28, où il tenait séminaire à Varsovie. Plus précisément, on a vu que l'idée d'organiser déductivement la métamathématique est déjà présente dans des travaux publiés en 1928; que deux mémoires publiés avant que Tarski n'ait eu le temps de prendre connaissance du résultat d'indécidabilité de Gödel portent sur la définissabilité sémantique; que son fameux mémoire sur le concept de vérité tire son origine d'une conférence du 21 mars 1931 à la Société des Lettres et des Sciences de Varsovie. D'un autre côté, on sait que Tarski et Gödel se sont rencontrés plusieurs fois, à Vienne en 1935, à Princeton en 1946 (à la conférence sur les problèmes dans les mathématiques): le texte de Gödel de 1946 commence d'ailleurs par évoquer la conférence de Tarski; plus tard encore en 1962 (photographie retrouvée dans le *Nachlass* et reproduite dans le volume II, p. 252, des *Collected Works*). Il est dommage qu'il soit trop tard pour procéder sur les rapports de Gödel et Tarski à quelque chose comme le questionnaire Grandjean qui nous a livré certains renseignements factuels bien précieux!

*Institut d'histoire et philosophie
des sciences et des techniques,
13, rue du Four
F 75006 Paris*

Références bibliographiques

- BARWISE, J., FEFERMAN, S. & SUPPES, P. (1989). Commemorative meeting for Alfred Tarski. Stanford University-November 7, 1983. In: *A Century of Mathematics in America*, Part III. Rhodes Island: American Mathematical Society, 393-403.
- BENIS-SINACEUR (ou SINACEUR), H. (1988a). La logique comme *ars inveniendi*. In: A. Robinet (éd.), *Doctrines et concepts. Cinquante ans de philosophie de langue française 1937-1987*. Paris: Vrin, 319-334.
- BENIS-SINACEUR (ou SINACEUR), H. (1988b). *Ars inveniendi* et théorie des modèles. *Dialogue*, XXVII, 591-613 (numéro paru en 1989).
- BENIS-SINACEUR (ou SINACEUR), H. (1989a). *Ars inveniendi* aujourd'hui. In: *Les Etudes philosophiques*, Avril-Juin, 199-214.
- BENIS-SINACEUR (ou SINACEUR), H. (1989b). *Ars inveniendi* aujourd'hui. In: *Leibniz, Tradition und Aktualität*, V. Internationaler Leibniz-Kongress, Vorträge II. Teil. Hannover: Gottfried-Wilhelm-Leibniz-Gesellschaft, 59-70.
- BENIS-SINACEUR (ou SINACEUR), H. (1990a). Caractéristique de G.W. Leibniz et théorie des modèles selon A. Robinson. Prépublications de l'Equipe de Logique Mathématique, CNRS-Université Paris VII, Séminaire de Structures algébriques ordonnées 1988-1989, n° 2.
- BENIS-SINACEUR (ou SINACEUR), H. (1990b). Article «Définissabilité». In: *Encyclopédie philosophique universelle. Vol. II: Les notions philosophiques*, t. 1, 564-566.
- BENIS-SINACEUR (ou SINACEUR), H. (1991a). *Corps et modèles. Essai sur l'histoire de l'algèbre réelle*. Paris: Vrin.
- BENIS-SINACEUR (ou SINACEUR), H. (1991b). Logique: mathématique ordinaire ou épistémologie effective? *Hommage à Jean-Toussaint Desanti*. TER, 331-346.
- DAWSON, J.W. Jr. (1991). *The Nachlass of Kurt Gödel: An Overview 6/91*, à paraître dans le volume III de Gödel 1986 (copie aimablement communiquée par l'auteur).

- GÖDEL, K. (1944). Russell's mathematical logic. In : P.A. Schilpp (ed.), *The Philosophy of Bertrand Russell*. Library of living philosophers, vol. 5, Evanston, Northwestern University, 123-153. (C.W., II, 119-141).
- GÖDEL, K. (1947). What is Cantor's continuum problem? *American Mathematical Monthly* 54, 515-525. (C.W., II, 176-187).
- GÖDEL, K. (1958). Über eine noch nicht benützte Erweiterung des finiten Standpunkt. *Dialectica* 12, 280-287. (C.W., II, 240-251).
- GÖDEL, K. (1964). What is Cantor's continuum problem? Version revue et augmentée de Gödel 1947, C.W., II, 254-270.
- GÖDEL, K. (1972). On an extension of finitary mathematics. Version anglaise augmentée de Gödel 1958, C.W., II, 271-280.
- GÖDEL, K. (1974). Remark on non-standard analysis. C.W., II, 311.
- GÖDEL, K. (1986). *Collected Works, I et II*. S. Feferman et al. (eds), New York-Oxford: Oxford University Press.
- HILBERT, D. (1922). Neubegründung der Mathematik. Erste Mitteilung. *Abh. math. Sem. Hamb.* 1, 157-177. (*Gesam. Abh. III*, 157-177). (Reprint: New York, Chelsea, 1965).
- HODGES, W. (ed.) (1986). Survey /Expository papers by R.L. Vaught, B. Jónsson, F. McNulty, J.D. Monk, L.W. Szczerba, S. Givant, *The Journal of Symbolic Logic* 51, n° 4, 865-941.
- HODGES, W. (ed.) (1988). Survey /Expository papers by A. Levy, L. van den Dries, J. Doner & W. Hodges, W.J. Blok & DonPigozzi, J. Etchemendy, P. Suppes, *The Journal of Symbolic Logic* 53, n° 1, 1-91.
- JACOBSON, N. (1974-1980). *Basic algebra I, II*. San-Francisco: W.H. Freeman & C°. Second edition, vol. I, New York: W.H. Freeman & C°, 1985.
- KOCHEN, S., KÖRNER, S., ROQUETTE, P., YOUNG, A.D. (1976). The pure mathematician. On Abraham Robinson's work in mathematical logic. *Bull. London math. soc.* 8, 307-323.

- TARSKI, A. (1928). Remarques sur les notions fondamentales de la méthodologie des mathématiques. *Annales de la Société polonaise de mathématique* 7, 270-272. (*Collec. Papers IV*, 552-554).
- TARSKI, A. (1930a). Über einige fundamentale Begriffe der Metamathematik. *C. R. Soc. sc. et lettres de Varsovie* XXIII, Cl. 3; dans Tarski 1972-1974, t. 1, III; (*Collec. Papers I*, 311-320).
- TARSKI, A. (1930b). Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften I. *Monatsh. Math. Phys.* 37, 361-404; (dans Tarski 1972-1974, t.1, V; *Collec. Papers I*, 341-390).
- TARSKI, A. (1931a). Sur les ensembles définissables de nombres réels I, *Fund. Math.* 17, 210-239; (dans Tarski 1972-1974, t. 1, VI; *Collec. Papers I*, 517-548).
- TARSKI, A. (1931b). Les opérations logiques et les ensembles projectifs, *Fund. Math.* 17, 240-248; dans Tarski 1972-1974, t.1, VII; *Collec. Papers I*, 549-559.
- TARSKI, A. (1935). Einige methodologische Untersuchungen über die Definierbarkeit der Begriffe. *Erkenntnis* 5, 80-100; (dans Tarski 1972-1974, t. 2, X. *Collec. Papers II*, 637-659).
- TARSKI, A. (1936/1960). Introduction à la logique. Paris: Gauthier-Villars/Louvain: E. Nauwelarts. (2e éd. augmentée 1969. 3e éd. 1971).
- TARSKI, A. (1936a). Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philos.* 1, 261-404; (dans Tarski 1972-1974, t. 1, VIII; *Collec. Papers II*, 51-198).
- TARSKI, A. (1936b). *Grundlagen der wissenschaftlichen Semantik. Actes cong. intern. philosophie scientifique* 3. Paris: Hermann, Actualités scientif. et indus. 390, 1-8; (dans Tarski 1972-1974, t. 2, XV; *Collec. Papers II*, 259-268).
- TARSKI, A. (1936c). *Über den Begriff der logischen Folgerungen. Actes cong. intern. philosophie scientifique* 7. Paris: Hermann, Actualités scientif. et indus. 394, 1-11; (dans Tarski 1972-1974, t. 2, XVI; *Collec. Papers II*, 269-282).

- TARSKI, A. (1944). The semantic conception of truth and the foundations of semantics. *Philos. and Phenomenal. Res.* 4, 341-376; (dans Tarski 1972-1974, t. 2, XXI; *Collec. Papers II*, 661-699).
- TARSKI, A. (1951). *A Decision Method for Elementary Algebra and Geometry*. Berkeley & Los Angeles: University of California Press, second revised ed. (*Collec. Papers III*, 297-368).
- TARSKI, A. (1956). *Logic, Semantics, Metamathematics. Papers from 1923 to 1938*. Translated by J.H. Woodger. Oxford: Clarendon Press. (Second edition edited and introduced by J. Corcoran, Indianapolis: Hackett Pub. C°, 1983).
- TARSKI, A. (1972-1974). *Logique, sémantique, métamathématique 1923-1944*. 1, 2, (en abrégé *L. S. M.* 1, 2). Paris: A. Colin.
- TARSKI, A. (1986). *Collected Papers, I, II, III, IV*. S.R. Givant, McKenzie R.N. Birkhäuser.
- WANG, H. (1974) *From Mathematics to Philosophy*. London: Routledge & Kegan Paul.
- WANG, H. (1987). *Reflections on Kurt Gödel*. Cambridge (Mass.)/London: The MIT Press.

UNPROVABILITY AND UNDEFINABILITY

John Corcoran

*Dedication: for Robert McNaughton
logician, teacher, friend*

1.0. There are many refinements, generalizations, and extensions of the Gödel unprovability results and of the Tarski undefinability results. This paper focuses on one core result of Gödel (along with the main lemma used in its proof) and on one core result of Tarski. Both results can be taken to concern *one* of the many formalizations of *one* of the many axiomatic formulations of the body of knowledge known as number theory (*number science* would be a more apt name). Each result has been understood as establishing a limitation on the classical axiomatic method of organizing scientific knowledge about the natural numbers. This paper will consider the question of the premises necessary to extend to *all* axiomatizations of number theory the core results established for *one* axiomatization. In order to clarify the significance, the concreteness, and the objective validity of these two results, it is necessary to remind ourselves of certain relevant mathematical, ontological, epistemological, and historical details.

Number theory, the science of whole numbers, can be conceived of in many ways; what exactly the various conceptions have in common is hard to say, it might be an important philosophical or mathematical question. Fortunately, for our purposes it does not seem to be needed. Here number theory is taken to be the science of natural numbers beginning with zero. But many choices could have been made; we could have started

with one, or included the negative integers, or even the rational numbers. Number theory originated independently (as far as is known) in several different parts of the world: China, India, and Greece are often mentioned. Number theory existed as a science long before it was organized axiomatically, regardless of whether we regard one of the Greek treatments in the third century B.C. as a genuine axiomatization or whether we regard one of the nineteenth century treatments (e.g., Dedekind's or Peano's) as the first axiomatization.

Long before any axiomatic organization took place there was an accumulation of *theorems*, propositions known to be true. No square number is twice a square number. No square number is three times a square number. For any given number n exceeding two, the sum of the first n consecutive odd numbers is the square of n . For any given number n exceeding two the sum of the first consecutive powers of two starting with one is one less than the next power of two. For any given number n , the sum of the first n consecutive even numbers beginning with two is the product of n with n plus one. The proof of Gödel's unprovability results shows an astounding mastery of number theory; people have trouble following Gödel's proof because they don't realize that they need to learn number theory not only to understand what the result is but also to follow the thinking that yields the result. Tarski was also a master of number theory as is evident from his later work.

From its origins before the fifth century the body of number theoretic theorems continued to grow. Alongside the body of theorems the Greek mathematicians were careful to note *hypotheses*, propositions that were not yet known to be true and not yet known to be false. A person's mastery of number theory is measured as much by the hypotheses that he knows of and can discuss as it is by the theorems that he knows and can prove. The Greeks knew that every number is exceeded by a prime, i.e., that no matter how far out the sequence of numbers one counts more prime numbers will be found. Several similar propositions were hypotheses that challenged the Greek mathematicians: every number is exceeded by a prime whose second successor is

prime, every number is exceeded by a perfect number. These two propositions are still open problems (Shanks 1962: 2, 30).

As far as we know rudimentary axiomatizations of parts of the body of then-known mathematics were attempted more than a century before Euclid who flourished around 300 B.C. But for centuries before the rudimentary axiomatizations, Greek mathematical activity embodied several practices that can be seen in retrospect to be essential to the axiomatic method. It should be no surprise that the components of the axiomatic method had been in use before being combined into the method.

One of the most important preaxiomatic activities was that of *deduction*, i.e., determining or establishing that one given proposition (the conclusion) is a logical consequence of a given set of propositions (the premise-set) or of a given proposition (the premise). Deduction was used in several ways. It was the main component of the *hypothetical method* of testing hypotheses by deducing from them consequences that are simpler and more amenable to various forms of direct verification or falsification. Hypothetical applications of deduction, i.e., the deducing of consequences from a premise-set that includes one or more propositions not known to be true, are mentioned by Plato in several places. At one point Plato seems to give Socrates credit for some sort of innovation connected with the hypothetical method. Two forms of deduction were recognized. Direct deduction unpacks the information contained in a set of premises step-by-step without adding anything for purposes of reasoning; this turns out to be the natural way to proceed in many cases where the premises and the conclusion are all affirmative. Indirect deduction shows that a given conclusion follows logically from a given premise-set by deducing a contradiction from the result of augmenting the premise-set by adding the contradictory opposite of the conclusion. The basic idea behind indirect deduction is that in order for a given conclusion to be implied by a given premise-set it is sufficient for the contradictory opposite of that conclusion to be contradicted by the premise set. Indirect deduction is often quite convenient when the conclusion is negative, e.g., to deduce «Zero does not divide itself» from «No number that divides itself is divided by the

number zero» or «No number that the number zero divides is a number that divides itself».

In addition to its use in the hypothetical method, deduction was also used in connection with generating paradoxes and in connection with demonstration (Corcoran 1989). Some paradoxes result from deducing a proposition believed to be false from premises all believed to be true.

Another preaxiomatic activity was *demonstration*, using propositions already known to be true as a basis on which to show that a given hypothesis actually is true. For example, once the reader knows that for every number n the sum of the first n consecutive odd numbers is the square of n , other similar theorems can be proved. It is easy to see that the sum of the first n even numbers beginning with two can be obtained by adding one to each of the first n consecutive odd numbers and then adding the results. This shows that the sum of the first n even numbers beginning with two is the sum of n with the square of n , which is the same as the product of n with its own successor. Legend attributes to Thales the introduction into mathematical practice of the activity of demonstration. Whether Thales is too early or too late, it is nevertheless clear that demonstration had become well-entrenched before the middle of the sixth century, over a century before the earliest axiomatization, and over two centuries before anyone systematically studied axiomatization, or demonstration, or even deduction. All three of these activities were first systematically investigated by Aristotle who constructed a general theory of all three, a theory whose most general aspects are widely accepted today, and form an essential context for the core results due to Gödel and Tarski. It is surprising at first to discover that the axiomatizations, demonstrations, and deductions constructed and discussed by Gödel and Tarski in connection with the core results are much more in keeping with Aristotle's theories as presented in the *Prior Analytics* and *Posterior Analytics* than are those found in Euclid's *Elements* or even in Hilbert's *Foundations of Geometry* (Corcoran 1974).

In a sense Aristotle's *truth-and-consequence conception of demonstration* is his most profound contribution to this field of

investigation, more because than in spite of the fact that it has so pervaded modern thinking as to become an inconspicuous part of the background, rather than a central part of the foreground. Aristotle's view was that a proposition is demonstrated, proved to be true, by showing that it is a *logical consequence* of propositions already known to be *true*. For Aristotle a demonstration begins with a set of propositions already known to be true and continues with a chain of reasoning that shows that the conclusion follows logically. Aristotle was perfectly clear that truth depends on *matter* as opposed to form; that a proposition is true or false in virtue of its subject matter. In fact Tarski traces his own correspondence theory of truth to Aristotle. At the same time Aristotle realized that logical consequence depends on *form* as opposed to matter. In order for a given proposition to be a logical consequence of a given set of propositions it is necessary and sufficient for the same to hold in every formally similar case (Corcoran 1974: 105). In so far as truth involves matter and consequence involves form, demonstration is a function of matter and form.

Aristotle also developed a conception of the method by which one given proposition is *shown* to be a logical consequence of a given set of propositions and he fully realized that this consequence aspect of demonstration is totally separable from the truth aspect. Knowledge of truth requires reference to matter. Knowledge of logical consequence requires reference to form. According to Aristotle, there are *immediate inferences* that can be made by humans. These correspond to what are today called logical rules, rules of inference, rules of deduction.

Inference of «Two is an even prime» from the two premises «Two is even» and «Two is a prime» is immediate as is inference of «Some prime is even» from the single premise «Two is an even prime». But inference of «Some prime is even» from the two premises «Two is even» and «Two is a prime» would have been considered by Aristotle to be non-immediate, i.e., mediated. Aristotle's *immediate-inference-chaining conception of deduction* is that in order to show that a given proposition is logically implied by a given set of propositions it is necessary and sufficient to construct a chain of immediate inferences, *either* a

direct deduction which begins with the given set and ends with the given proposition *or* an indirect deduction which begins with the given set augmented by the contradictory opposite of the given proposition and ending with a proposition which is the contradictory opposite of one of the preceding. The fine details are not important. What is of crucial importance is that a demonstration is conceived of as a deduction whose premises are known to be true and that a deduction is conceived of as a chaining of immediate inferences.

This clears the way for formalization of deduction and thus for formalization of proof and of the axiomatic method. Aristotle's proofs are deductions and the deductions are chainings of propositions linked by formal (as opposed to material) inferences. Thus once the premises have been established it is no longer necessary (indeed no longer possible) to be involved with the subject matter in order to complete a proof. In particular, a deduction in arithmetic contains no arithmetic constructions and no computations, and a deduction in geometry contains no geometrical ruler-and-compass constructions and no superpositions. It should be added that the Aristotelian prohibition of constructions, diagrams, intuitions, etc. from *deduction* (i.e., from the deducing of consequences from premise-sets) was never intended to apply to the process of *induction* by which the premises are established nondemonstratively. For more on establishing the «ultimate» premises see the 1980 article by Jaakko Hintikka «Aristotelian Induction», which is reviewed in *Mathematical Reviews* (1982) 82m:00016.

We can use the expression *formal deduction* to indicate a deduction which proceeds from the initial premise-set to the conclusion by a series of formal steps, i.e., by steps which have to do solely with the logical forms of the propositions involved and which in particular are in no essential way dependent on knowledge of the matter. In this sense, Aristotle's immediate-inference-chaining conception is a theory of formal deduction as is the theory of deduction which is the basis of the *underlying logic* of the axiomatization of number theory constructed by Gödel. It is in this connection that Gödel's axiomatization fulfills

the Aristotelian requirements to a much greater extent than does Euclid's axiomatization or indeed perhaps any axiomatization constructed before the modern period.

Aristotle's theory of the axiomatic organization of the sciences is presented in the *Posterior Analytics* which presupposes the theory of proof and the theory of deduction found in the *Prior Analytics*, but it goes beyond those two theories in several respects. In the first place, it recognizes the need to determine from among all of the concepts involved in the given science a small set of concepts (the *principle* concepts) in terms of which the other concepts may be defined. In addition, it recognizes the need to determine from among all of the theorems of the science a small set of propositions (the *principle* propositions) which, when augmented by suitable definitions, is sufficiently rich to serve as a premise-set from which all of the theorems of the science may be *recovered* by logical deduction. It went without saying that the set of principle concepts and the set of principle propositions were both *finite*; the science to be axiomatized was already in existence as a human construction (admittedly based on an objectively existent universe of discourse).

2.0. It would have been entirely possible for Aristotle to have produced an axiomatization of number theory not differing essentially from the (non-formalized) axiomatization of number theory that was to be formalized by Gödel...had Aristotle thought of it. The universe of discourse, as already indicated, is the set of natural numbers: 0, 1, 2, 3,.... The principle concepts are the concepts «zero», «successor» and «number». And the principle propositions, which are called the *Gödel Axioms*, the *Peano-Gödel Axioms* or even (anachronistically) the *Peano Axioms*, are as follows.

The Peano-Gödel Axiom Set

- A1 No number has zero for its successor.
- A2 Every two distinct numbers have respectively distinct successors.

- A3 Every property that belongs to zero and to the successor of every number it belongs to belongs also to every number.

The first axiom, sometimes called *the zero axiom* amounts to the proposition that given any number n , the successor of n is not zero. The second axiom, sometimes called *the successor axiom*, amounts to the proposition that the successor function is one-to-one: given a number n , for any number m , if m is not n , then the successor of m is not the successor of n . By the way, arithmetic tradition dictates that 'equals' is used in the sense of the *is* of identity, but since this tradition does not extend to geometry it is sometimes misleading, especially in foundational discussions. Tarski's 1946 *Introduction to Logic* includes an interesting elaboration of this point. The third axiom is the principle of mathematical induction which became an integral part of modern thought as a result of the work of Blaise Pascal.

Contrary to the impression one gathers from various sources, principle concepts are not *ipso facto* undefinable in any ordinary sense of the word and, of course, it would be absurd to think that by dint of use as an expression for a principle concept, a previously meaningful expression (e.g., 'zero', 'successor', 'number') somehow becomes meaningless or somehow loses its status as constant and becomes a variable. The literature of this subject is replete with passages that betray lapses in precision, in objectivity, and in common sense (Keyser 1922: 54-59; Russell 1918: 5-9). In preformalization terminology it is said that a concept, e.g., the concept of zero, is definable in a given axiomatization if and only if a proposition giving a defining (necessary and sufficient) condition for it is provable (deducible from the axioms). For example, in order for «zero» to be definable in the above preformalized axiomatization, which is called *PGA* (Peano-Gödel Axiomatization), it is sufficient for the following proposition to be provable as a theorem.

- DZ In order for a given number to be zero it is necessary and sufficient for the given number to not be the successor of any number.

The property of not being *the* successor of any number, or more briefly, the property of not being *a* successor, i.e., the property of being a non-successor, is a defining property of zero (Tarski 1944: 345). Proposition DZ amounts to the proposition that the property of being zero is *co-extensive* with the property of not being a successor. Thus DZ is logically equivalent to the proposition that the property of being zero and the property of being a successor are mutually exclusive and jointly exhaustive of the numbers. The proposition that they are mutually exclusive, i.e., «No number both is zero and is a successor (sc. of a number)» can be seen to be of course logically equivalent to the first axiom and is thus *a fortiori* provable as a theorem. The proposition that they are jointly exhaustive of the numbers, «Every number either is zero or is a successor», is not implied by the first axiom nor even by the first two axioms, as was known by Dedekind (1888), but it can be deduced from the third axiom alone. This joint exhaustiveness proposition can be construed as a universal proposition with a disjunctive predicate: «Every number has the property of being either zero or a successor». To see that this proposition follows from the principle of mathematical induction (axiom three) alone it is sufficient to follow a chain of individually trivial steps of reasoning. In the first place, it is clearly *tautologous* that zero has the property of either being zero or a successor because this follows *a fortiori* from the tautology that zero has the property of being zero, which in turn follows immediately from the tautologous identity that zero is zero.

It is likewise easy to see that it is tautologous that every successor of a number having this disjunctive property also has the property, i.e., that for any given number n , if n is either zero or a successor then the successor of n is either zero or a successor. Notice that the consequent of the imbedded conditional is itself tautologous. The two propositions just established form respectively the so-called basis step and induction step for an «application» of mathematical induction yielding the required conclusion--that every number is either zero or a successor (sc. of a number). Let us display the three theorems just obtained and label them.

- T1 No number is both zero and a successor.
- T2 Every number is either zero or a successor.
- T3 Every number is such that it is zero if and only if it is a non-successor.

The third theorem amounts to the proposition DZ; the two are equivalent definitions of zero. As we have seen, then, «zero» is definable in the Peano-Gödel Axiomatization in terms of the concepts of (immediate) successor and (natural) number. This shows, as indicated above, that being a principle concept of an axiomatization does not *ipso facto* exclude being a definable concept of the same axiomatization. The real interest in definability lies of course with non-principle concepts.

The concept of one is defined by the property of being the successor of zero and by the property of being the successor of a non-successor. The concept of two is defined by the property of being the successor of one, and so forth for as many such concepts as one may consider.

- DI One is the successor of zero.
- DII Two is the successor of one.

The concept of precedence which is often said to give the natural ordering of the numbers is now known to be definable in terms of the principle concepts of the Peano-Gödel Axiomatization as a result of stunningly brilliant insights independently achieved by Dedekind (1888) and by Frege (1879). Here is the background of the insight. Let p be a number and let q be a number. In order for p to precede q it is sufficient for the successor of p to be q itself or for the successor of the successor of p to be q itself, or and so on. Dedekind and Frege discovered how to give logically rigorous expression to «and so on», i.e., to the definition of precedence pointed out by the above background remark.

DP In order for one given number to precede a second given number it is necessary and sufficient for the second to have every property belonging to the successor of the first and to the successor of any number to which it belongs.

Once the definition of precedence and a few other standard concepts of number theory had been constructed, it began to seem plausible to suspect that, without exception, absolutely every concept then-employed by number-theorists going back through Gauss, Fermat, Pascal and so on to Diophantus and the other early arithmeticians could indeed be defined in terms of the principle concepts of the Peano-Gödel Axiomatization. As a result of his own experience constructing definitions together with his encyclopedic knowledge of number theory, Peano was able to verify that every concept already in use in number theory in fact admitted of a definition in terms of zero, successor, and number. No comparable result had ever been achieved in the whole history of mathematics; a comprehensive conceptual basis of a science had been isolated.

Does this mean that every numerical concept is definable in terms of zero, successor, and number? Certainly not! Because of the finite complexity characteristic of the properties constructible from the principle concepts it was clear that the class of «definable» properties is countably infinite, i.e., that the class of «definable» properties is the same size as the set of natural numbers. In contrast, as a result of the work of Cantor it was known that there are uncountably many numerical properties, i.e., that there are more numerical properties than there are definable properties. Thus many, perhaps we should say "most", properties are not definable in terms of any given finite set of concepts and in particular in terms of the concepts of zero, successor, and number.

3.0. These results, positive and negative, carry over to the formalization of the Peano-Gödel Axiomatization which is the real topic of this paper. Before the formalization is considered it remains to discuss provability and unprovability in the

unformalized or preformalized setting. Before turning to that however it is important to notice a difference between what it means to say that a given principle concept is definable in an axiomatization in terms of other principle concepts, on the one hand, and, on the other hand, to say that a given non-principle concept is definable in terms of the principle concepts. This important point, often overlooked, helps to avoid confusion. In order for a given principle concept to be definable in terms of other principle concepts in a given axiomatization it is necessary and sufficient for a proposition of a certain kind to be provable as a theorem in that axiomatization, more explicitly, for a proposition in the form of a definition of the given concept in terms of the other concepts to be provable from the axioms. Now in the case of the non-principle concept, proving such a proposition is out of the question because the propositions involving non-principle concepts in an essential way are not even in the running, so to speak. When it is said that a non-principle concept of number theory is definable in terms of «zero», «successor» and «number» what is meant is that a proposition in a suitable definitional form is a theorem of number theory itself, the preaxiomatic science, the science being axiomatized.

In accordance with the thinking involved in the above point there are two ways of treating the addition of definitions to an axiomatization. For many purposes it is technically convenient to regard the definitions being added not as «real» definitions of non-principle concepts but rather as «nominal» («stipulative» or «abbreviational») definitions of «new» words (regarded as antecedently meaningless). For example, on this approach in the above Peano-Gödel Axiomatization, DI would be understood as stipulating that the word 'one' is to be used as an abbreviation of the term 'the successor of zero'. In this way, the only numerical concepts ever occurring in theorems of this axiomatization are those of zero, successor, and number. This is the *definitions-as-abbreviations* approach.

The other approach takes the «new» words with their standard meanings. The sentences expressing the added definitions therefore express genuine propositions involving the «new» concepts and the principle concepts already in use. Thus, this

approach takes the added definition to be a new axiom expanding the set of principle concepts of the axiomatization. This is the *definitions-as-axioms* approach favored by Tarski and the California logicians (Suppes 1957: 153).

In this paper the abbreviational approach to definition is adopted except where explicit indications to the contrary are given. This means that we take the propositions (whether true or false) relevant to the Peano-Gödel Axiomatization to be those whose universe of discourse is the class of numbers (the extension of the concept of number) and whose non-logical (material, mathematical) concepts other than «number» are only «zero» and «(immediate) successor». The sentence 'The successor of one is two' thus does not express a proposition containing the concepts «one» and «two» but rather is an abbreviation of the sentence 'The successor of the successor of zero is the successor of the successor of zero' which expresses a tautology. The proposition normally expressed by the sentence 'The successor of one is two' is not implied by the Peano-Gödel Axiom-Set, any more than is «The predecessor of Johnson was Kennedy» (cf. Keyser 1922: 57).

4.0. Once the class of propositions (true and false) relevant to the axiomatization is settled we can turn to the question of provability (in the axiomatization): is every (relevant) proposition of number theory provable in this axiomatization? In other words, is every proposition involving no concepts other than «zero», «successor» and «number» deducible from the PG Axiom-set? In still other words is the axiomatization as a whole *comprehensively complete*? This question concerns the entire axiomatization, its material aspect as well as its formal aspect, the mathematics *per se* as well as the logic *per se*, the axiom-set and the underlying logic. The question breaks down into two parts: one about the axiom-set alone and one about the logic alone. The first is whether the PG Axiom-set is rich enough in information to contain as logical consequences absolutely every relevant true proposition. The second is whether the human potentiality to perform logical deduction is rich enough to be able to deduce

from a given premise-set each and every proposition that is a logical consequence of the premise-set. If the answer to the first question is affirmative then the axiom-set is *implication complete* in the sense that every relevant proposition is either implied or contradicted by the axiom-set itself. An implicationally complete set of propositions is so rich in information content that it implies every (relevant) proposition whose negation it does not imply and it implies the negation of every relevant proposition that is not one of its implications. If the answer to the second question is affirmative then the system of deductions of the underlying logic («our logic») is *deduction complete*.

In 1904 Oswald Veblen presented an unformalized axiomatization of Euclidean geometry using as principle concepts «point» and «between». He went on to show that his axiom-set is *categorical*, i.e., that any two reinterpretations of the non-logical constants (used to express the principle concepts) will be isomorphic to each other if they both satisfy the axiom-set. The only reason for mentioning categoricity here is that, because of the nature of logical form of propositions, every categorical axiom-set is implication complete. In Veblen's words, every relevant proposition is either implied by his axiom-set or is contradicted by his axiom-set. This means, as already indicated, that his axiom-set is as rich in information as could be desired--no true proposition could be irredundantly added (Corcoran 1980, 1981).

Veblen went on to raise the questions of whether the axiomatization as a whole (geometry plus logic) is comprehensively complete and he realized that his implication completeness result reduces the question of comprehensive completeness to the question of deduction completeness of the underlying unformalized logic. Veblen specially raised the question of whether for each proposition logically implied by his axiom-set there exists a finite chain of immediate inferences certifying the implication. More specifically, Veblen said that each relevant true proposition is implied by his axiom-set «even were it not deducible from the axioms by a finite number of syllogisms» (Veblen 1904, 346n).

The Veblen implication completeness result concerning geometry is relevant to the concerns of this paper even though this paper focuses exclusively on axiomatizations of arithmetic. The Veblen result illustrates the fact that the distinction between implication and deduction had been made long before the work of Gödel and Tarski. It also illustrates the fact that deep and fascinating results concerning an *unformalized* axiomatization had been achieved. Veblen's axioms were stated in natural language augmented with a few mathematical symbols; no formal language was used much less described and nothing was said about the rules for making deductions--except that each deduction was a finite chain of immediate inferences. Moreover, in view of the fact that Veblen gave a proof about an axiomatization *per se* it is necessary for him to have used premises that were *about* the axiomatization and that were known to be true and it was necessary for him to have shown by a deduction that his conclusion was a consequence of his premises. Discussion of Veblen's premises would detract from the main theme of this paper.

Perhaps the most important reason for mentioning the Veblen Implication Completeness Result for geometry is to provide a parallel to the fact that the analogous result (or at least all hard mathematical work needed for it) had already been achieved for arithmetic as an outcome of the work of Dedekind and Cantor. And this is even before Veblen. In particular, this early work, which must have been known to Gödel and which certainly was known to Tarski, makes it clear that *the Peano-Gödel Axiom-set is categorical and therefore implicationally complete*.

There is another comparison which is well worth a small digression. As we saw above, it was found that every arithmetical concept employed by number theorists turns out to be definable in terms of «zero», «succession» and «number», the principle concepts of the Peano-Gödel Axiomatization. Veblen reported an analogous result for the principle concepts of his own axiomatization of geometry. In particular, he claimed that he proved that line congruence (or line equality) is definable in terms of «point» and «between». However, in the 1920s Tarski who was somewhat of an expert on definition, found fallacies in

Veblen's proof and then went on to prove himself, with the help of Lindenbaum, that the concept of line congruence (as in «line AB is congruent to line CD») is not definable in terms of the Veblen principle concepts. Thus there is a striking disanalogy between the Peano-Gödel Axiomatization and the Veblen Axiomatization. See Corcoran 1986 for a few more details on this curious episode in the history of the theory of definition.

5.0. In a sense Gödel's core result was designed to illuminate the nature of the axiomatic or deductive method. More particularly, again in a sense, it was designed to illuminate the nature of the Peano-Gödel Axiomatization and, thereby, to the extent that the Peano-Gödel Axiomatization is representative of axiomatizations of number theory, it was designed to make a statement about the scope and limitations of the axiomatic method. The method that Gödel used to establish his result was the standard method of mathematical proof--the truth-and-consequence method of proof: beginning with premises known to be true show by means of chaining of immediate inferences that the conclusion is logically implied by the premise-set.

Gödel's first step was to replace the vague and informal axiomatization by a mathematically precise axiomatization about which premises could be established. This first step itself involved three constructions. First, to model or express or represent the class of relevant propositions Gödel constructed a mathematically precise formalized language L of formal sentences about which various theorems could be established and in particular to which number theory could be applied. Second, to model or represent the class of deductions of the underlying logic of the PG Axiomatization Gödel constructed a mathematically precise system D of deductions, extended discourses, likewise amenable to treatment as mathematical objects. Third, to model or represent the mathematics or matter or content of the PG Axiomatization, Gödel constructed a set G of sentences of the language L .

In order to specify the language L , i.e., the set of sentences of the language L , Gödel presented a *sentential grammar*, which

specified an alphabet (or vocabulary) of meaningless (but potentially symbolic) characters, a set of simple strings of characters as a basis and a set of rules of generation or transformation for constructing complex strings from simple strings. In order to specify the system D of deductions, Gödel presented a *discourse grammar* by specifying another set of rules for generating «complex deductions» from «simple deductions» which were simply concatenations of premises. For more details on sentential grammars and discourse grammars see my 1971 series of articles «Discourse Grammars and the Structure of Mathematical Reasoning». The ideas for the «logical structure» of the formal sentences and for the «logical structure» of the formal deductions was taken with modifications from *Principia Mathematica* which itself was taken from the underlying logics presupposed in mathematical practice including the work mentioned above by Veblen, Peano and Dedekind. Once a person sees the connection between the fully formalized underlying logic in the 1931 Gödel paper and the unformalized writings of Veblen, Dedekind and others it is amazing how easily the latter admits of being «translated into formalized form». Indeed, were it not the case that the Gödel Axiomatization had and has the appearance of being a formalization of *preexisting nonformalized mathematical practice*, no one would have had the motivation to attach to it the importance that it has attracted.

It is important to realize that the underlying logic that Gödel took in formalized form for the axiomatization of number theory that was to become so famous was a full unrestricted predicate logic in which the principle of mathematical induction can be expressed in full by means of one sentence, as is the case of course in normal unformalized mathematical discourse. It is also interesting to point out that before Gödel undertook the work that was to lead to his unprovability result involving «the» unrestricted predicate logic (i.e., higher order logic) he had proved the deduction completeness of «the» restricted predicate logic (i.e., lower predicate logic, first order logic). See Shapiro 1990.

The axiom-set, properly so-called, of the Gödel Axiomatization is the set G consisting of three sentences from the

formal language L which are obtained by «translating» or «formalizing» the Peano-Gödel Axiom Set. Thus G contains formal analogues of the zero axiom, the one-one axiom and the axiom of mathematical induction. This completes the specification of the Gödel Axiomatization which is denoted GA .

As the logic of the Gödel Axiomatization was taken from *Principia Mathematica* which at the time was the most modern codification of the formal aspect of traditional demonstrative practice in mathematics, the content of the Gödel Axiomatization was taken from the Peano work, which represented the most modern codification of the contentual aspect of the body of number-theoretic knowledge tracing back to ancient Asian and European cultures.

By the way the propositions that are called 'logical axioms' in an underlying logic are really not axioms in the mathematical sense but rather are zero-place rules of inference. A *two-place* rule of inference such as *modus ponens* or *barbara* lengthens a deduction on the basis of *two* lines already in the deduction. A *one-place* rule of inference such as *universal instantiation* or any of the *conversion* rules (e.g., going from «zero is not the successor of one» to «one is not the successor of zero») lengthen a deduction on the basis of *one* line already present. A zero-place rule such as *self-identity* (e.g., enter «zero is zero», «one is one», etc.) lengthens a deduction without preconditions. Overlooking this point has led to some tragic confusions as pointed out by Blumberg (1967, esp. 24-25) and others. In this paper 'axiom' is used in the sense of subject-matter axiom. Regardless of what they are called, zero-place rules or «logical axioms» are part of the underlying logic having to do with the formal aspect of demonstration whereas axioms properly so-called are part of the content of an axiomatization.

The Gödel Axiomatization, GA , of number theory consists of an underlying logic (with language L and deduction system D) together with the axiom-set G whose non-logical constants are the two characters ' O ' and ' s '. The first of these, ' O ', is an individual constant intended to denote the individual number zero and the second, ' s ', is a one-place function constant intended to denote the one-place successor function. Thus ' sO ' the

concatenation of 's' with 'O', is a string of length two which is intended to serve as a name of the number one: 'sO' formalizes 'the successor of zero'. In the language L there are individual variables intended to range over the intended universe of discourse which is the class of natural numbers. There are property variables intended to range over the class of numerical properties (or, in extension, the class of sets of natural numbers), and there are property-property variables that range over the class of property properties (such as the property of belonging to countably many numbers which itself belongs to the property of being odd, the property of being even, etc.). There is no limit to the *order* of variables. As could be deduced from what has already been said, it was clear at the time that this axiomatization was sufficient for formalization of all of the then-known number theory together with the so-called arithmetization of analysis including integral and differential calculus.

Gödel's Core Result is that the *Gödel Axiomatization* is not *comprehensively complete*, i.e., that the combination of the axiom-set G with the underlying logic (involving the language L and the deduction system D) is not comprehensive enough to include for every sentence of L either a proof of the sentence itself or a proof of the negation of the sentence. This means that when we look at the set of all formal deductions in the underlying logic there will be a formal sentence f whose deductions and «antiductions» all use premises other than the three axioms in G . By a deduction of f is meant a deduction whose conclusion is f and by an antideduction of f is meant a deduction whose conclusion is the negation of f . Thus the Gödel Unprovability Result is not to the effect that some true *proposition* is not provable (by humans) but rather that some formal sentence has neither a formal deduction nor a formal antideduction using exclusively premises in G . From the point of view of the axiomatization GA «deductions» are certain concatenations of formal sentences, viz. those concatenations in D , and proofs are deductions whose premises are all in G : D provides the formal aspect of demonstration, G provides the material aspect or content.

There are two ways of looking at the class of formal deductions in D : we can choose a certain set S of sentences and look at the class of deductions each of whose premise-sets are subsets of S and we can choose a sentence f and look at the class of deductions having f or the negation of f as conclusion. There is the premise-oriented approach and the conclusion oriented-approach.

The conclusion-oriented approach was taken above when we noticed that the Gödel Unprovability Result implies that there exists a sentence f in the language L whose deductions all involve premises other than simply those in G and whose negation has the same property of being deducible only by deductions using premises other than those in G . Such sentences may be said to be *nonconcludable in GA* ; Gödel called them 'undecidable' but this word has an entirely different use today as we will have occasion to notice below. To take the premise-oriented approach we can call a premise-set *inconclusive with respect to a deduction system* if there exists a sentence such that neither it nor its negation is the conclusion of a deduction in that system. From this point of view the Gödel Unprovability Result is simply that the axiom-set G is inconclusive with respect to the deduction system D .

6.0. As has been emphasized above the core result by Gödel, the Gödel Unprovability Result, is about a formalized axiomatization constructed by Gödel in order to study questions concerning axiomatization of number theory. Can we conclude from the Gödel Unprovability Result that the unformalized Peano-Gödel Axiomatization is not comprehensively complete? Can we conclude from the Gödel Unprovability Result that no comprehensively complete axiomatization of number theory is possible?

Before inquiring into the implications of the Gödel Unprovability Result it might be useful to inquire into its subject-matter to see whether we can locate more precisely the deficiency in the axiomatization (whose components include the formal language L , the deduction system D , and the axiom-set G). Could it be that the formalized language L contains beyond the

formal sentences that can be used to express ordinary number-theoretic propositions such as those whose truth-values were determined by classical number-theorists other formal sentences which are artifacts of the construction and which do not correspond to any genuine propositions? This question, and variants of it, have led to interesting developments even though a deeper acquaintance with number theory and logic lead conclusively to a negative answer: there is no trouble with the formal language L and, in particular, it is not the case that bloating of the language with bizarre artifactual sentences is the cause of the trouble with the axiomatization.

The next component to be considered is the axiom-set G : could it be that G is deficient in information, i.e., that there are relevant propositions in number theory not implied by the set of propositions expressed by G ? This question can likewise confidently and conclusively be answered in the negative as a result of work done years earlier by Dedekind and others. The fact that the Peano-Gödel Axiom-set is categorical and therefore *implication complete* carries over without difficulty to the formalized analogue. The formalized Gödel Axiom-set G is implication complete in the relevant sense, i.e., given a formal sentence f in L , either f is a logical consequence of G or the negation of f is a logical consequence of G . In view of the Gödel core result, this means that there is a sentence f in L which is logically implied by G but not deducible from G by means of a deduction in D . Thus some logical consequence of a set of sentences is not deducible from that set using deductions in D . In traditional terminology there are logically valid arguments which can not be seen to be valid by a deduction constructed in accordance with the rules of deduction that Gödel chose for his axiomatization. The deduction system of the Gödel Axiomatization is *deduction incomplete*.

Now the underlying logic of the Gödel Axiomatization contains as a sublogic a logic of first order, a logic of second order, a logic of third order, etc. Gödel had already shown the year earlier that the first-order sublogic of this unrestricted underlying logic is deduction complete. Thus, the deduction

incompleteness must be found at one of the higher levels. Which one?

The Main Lemma of the Gödel core result, i.e., the lemma that culminates the chaining of reasoning and that yields more or less immediately the Unprovability Result, identifies by name a sentence, call it g , which expresses a proposition that can be proved to be true on the basis of premises that can be known to be true by *reflecting* on the axiomatization itself but which can not be known to be true by means of a formal proof concluding with g . This sentence g is a first-order sentence. The axiom-set G , known to imply g , is second-order. Thus, deduction incompleteness of the underlying logic enters at the lowest place compatible with known results: *Gödel's Second-Order Logic is deduction incomplete*.

Moreover, since the axiom-set G , which implies g but from which g is not deducible using Gödel's deduction system is finite, there is a single sentence logically equivalent to G , viz. the conjunction of the first axiom with the conjunction of the second with the third. Thus the conditional of this conjunction with the sentence g is a tautology which is not deducible in D from the null set and thus not «logically provable». This means that D is not complete in the weak sense. The above argumentation is due to Henkin (1950: 81).

As mentioned above Gödel's proof of his Main Lemma depends on applying number theory itself to the investigation of strings of initially meaningless characters; strings that once interpreted express propositions and proofs in number theory. The key to this application of number theory is based on the fact that the strings in question admit of being enumerated $S_0, S_1, S_2, \dots$ and thus each string is assigned a natural number. Once this is done then it is possible to imagine the formal sentences to be about the strings enumerated and not just about the numbers themselves. Tarski and Gödel thought of this strategy independently (Tarski 1983: 278).

In the course of his proof Gödel proved that the property of being a number of a provable sentence admits of being defined in terms of zero, successor and number by a certain formula $PR(x)$ which has exactly the variable x free. In other words, Gödel

produced a formula $PR(x)$ and then showed that for every number n , n satisfies the formula $PR(x)$ under the intended interpretation if and only if n is the number of a string that occurs as the conclusion of a deduction whose premises are strings (sentences) in G , the Gödel Axiom-set. Tarski, as is well-known, showed how to define the string-theoretic property of being a sentence of a given formalized language expressing a true proposition and in particular (in effect) for L interpreted in the intended way. Then Tarski continued in a different direction to get what is called here *Tarski's Core Result*, viz. that the number-theoretic property of being a number of a string which expresses a true proposition is not definable on the basis of zero, successor and number, in other words no matter which formula $F(x)$ from language L is considered, under the intended interpretation there is a number n which satisfies $F(x)$ but which is the number of a «false sentence» or there is a number n which does not satisfy $F(x)$ but which is the number of a «true sentence». In short, Tarski showed that any formula $F(x)$ considered as a «definition» of «number of a sentence of L expressing a true proposition» is either too narrow or too broad.

Tarski himself realized sometime later that his core result is actually a lemma that yields the Gödel core result more or less immediately. One premise in the argumentation is that every sentence provable as a theorem in the Gödel Axiomatization expresses a true proposition and thus that every number satisfying $PR(x)$ has the property TR of being the number of a sentence expressing a truth. But there is no formula that expresses TR . Thus the set of numbers satisfying $PR(x)$ is not the extension of TR . Thus some number having TR does not satisfy $PR(x)$. Thus some sentence expressing a true proposition is not provable in GA .

Gödel's core result was originally obtained from a lemma that identifies a sentence that can not be known to express a true proposition of number theory by a formal proof in GA but which is known to be true by *reflecting* on the axiomatization itself. Likewise Tarski's core result identifies a numerical (or number-theoretic) property which can not be defined in terms of the principle concepts of Gödel's axiomatization of number theory (and

hence can not be defined in terms of the normally used concepts of number theory) but which can be defined in terms of concepts that arise naturally in *reflecting* on the axiomatization.

Gödel's proof is deduction of his core result from premises which are known to be true. If it should ever be determined that even one of these premises is not known to be true by the community of investigators then Gödel's deduction will no longer be regarded as a proof of its conclusion. Among these premises will be found the definition of truth discovered by Tarski a few years later and the basic laws of string theory also set forth later. Some of these points are discussed in the 1974 paper «String Theory» by Corcoran, Frank and Maloney.

7.0. In conclusion allow me to say a few words about what would be involved in generalizing Gödel's proof so that, instead of proving that the Gödel Axiomatization of number theory is not comprehensively complete, we would prove that no axiomatization of number theory is possible. In order to do this we will have to discover several things about axiomatizations which are either not known at all or else which have not been explained clearly enough to warrant wide acceptance. One overarching need is for specification of the string-theoretic (syntactic) property that a «system» must have in order to be suitable to serve as an axiomatization. In particular, if a formalized language L is in fact suitable for use in an axiomatization of a science, what string-theoretic condition does it satisfy in virtue of its being suitable for such use? In other words, what is the *axiomatic-language-suitability condition*? Again, if a formalized system of «derivations» is in fact suitable for use as the deduction system in the underlying logic of an axiomatization, what string-theoretic property does it have in virtue of this suitability? In other words, what is the *deduction-system-suitability condition*? Finally, if a set of formal sentences is suitable for use as an axiom-set, then what string-theoretic property does it have in virtue of its suitability? In other words, what is the *axiom-set suitability condition*?

Gödel seems to have changed his mind about these issues between the time he announced his core result in 1930 and the time he published it in 1931. Gödel's views differ from those of Tarski 1969 and from the still different views expressed in Church 1956 and Corcoran 1969 and 1973. The variety of views on just the axiom-set-suitability condition is remarkable and representative. It is clear from surveying the literature that before 1930 the idea that formal languages must consist in finite strings over a finite alphabet was virtually universal and there was likewise universal or near-universal agreement that an axiom-set must be finite. See for example the revealing remark in Presburger's 1930 paper which was based on work supervised by Tarski (Presburger 1930, English 229, German 95). The idea of infinitely many axioms seemed to many mathematicians and logicians to be contrary to the intuitive concept of axiomatization and it still seems so. Gödel 1930 seems to condone the idea of infinitely many axioms as long as they are given in such a way that for any given number it is possible to prove as a theorem about the axiomatization a proposition to the effect that that number is (or is not) a number of an axiom whenever it in fact is (or is not) such a number. Tarski 1969 explicitly accepts an axiom-set whose set of numbers is definable by means of the principle concepts of the axiomatization. Today the overwhelming majority of mathematical logicians would easily accept as an axiom-set a set of formal sentences which is decidable in the modern sense, i.e., a set for which there exists an algorithmic procedure for determining of an arbitrary string whether or not it is in the proposed axiom-set (Mendelson 1987: 28, 165). It is important to note that being decidable does not mean being known to be decidable; there are many decidable sets for which no algorithm has been constructed or for which no constructed algorithm is known to «work». The condition of definability (proposed by Tarski) and the condition of decidability (accepted by Mendelson and others) seems so remote from the traditional axiomatic method that further explanation may be demanded--and yet none had been offered, at least not by Tarski or Mendelson. This is one small indication that investigation into the foundations of the axiomatic method

continues to challenge philosophers, logicians and mathematicians.

I hope that I have succeeded in showing how Aristotle's truth-and-consequence conception of proof and his immediate-inference-chaining conception of deduction paved the way for investigation of the phenomenon of axiomatization which led to formalized axiomatizations and to the studies of the inherent-limitations of formalized axiomatizations. In addition, it was my intention to show that the same investigations by Gödel and Tarski that led to the inherent limitation results also produced insights into the nature of axiomatization which made possible realization of potential for increased rigor already implicit in the traditional conception of axiomatization. Perhaps most importantly, I wanted to illustrate that in establishing results about the axiomatic method Gödel and Tarski both applied the truth-and-consequence conception of proof which was part of the basis of the axiomatic method. And finally, I hope that I showed that just as the activity of identifying the premises in proofs of number-theoretic propositions leads to axiomatization and to insights into the foundations of number theory, the activity of identifying the premises of the Gödel proof led to discoveries *and* the activity being carried out today of identifying and articulating the premises of the proofs by Gödel, Tarski and others continues to produce advances in understanding.

The most rigorous mathematical analogue of the traditional notion of axiomatization used above is the notion of *formalism* developed by Tarski and Givant in 1987.

Acknowledgements: This expository paper was prepared with assistance of Michael Scanlan and Sriram Nambiar for the International Colloquium on Kurt Gödel held at the University of Neuchâtel in June of 1991. The Colloquium was designed with certain explicit educational goals. Accordingly, I have permitted myself rather more pedagogical license than would otherwise be appropriate in a paper of this sort. The high level of preparation evident from the student participation at the Colloquium indicated to me that certain simplifications and repetitions were not necessary. Nevertheless, I thought that the paper would be

accessible to a broader audience were the pedagogical measures allowed to stand. It is my hope that this paper will be found to be useful by Swiss logic students and, in this way, I will be able to repay somewhat the warm hospitality that they and their teachers have extended to me not just in June of 1991 but on every one of my visits. Special thanks to Professor Denis Miéville for organizing an event that proved to be both pleasant and scientifically rewarding.

*Department of Philosophy
State University of New York at Buffalo
Buffalo, New York 14260 U.S.A.*

Bibliographical references

- BLUMBERG, A. (1967). Modern logic. In: Edwards 1967 Vol. 5, 12-34.
- BUNGE, M. (1973). *Methodological Unity of Science*. Dordrecht: Reidel Pub.
- CHURCH, A. (1956). *Introduction to Mathematical Logic*. Princeton: Princeton University Press.
- CORCORAN, J. (1969). Three logical theories. *Philosophy of Science*, 36, 153-77.
- CORCORAN, J. (1971). Discourse grammar and the structure of mathematical reasoning. In three parts *Journal of Structural Learning* 3 n° 1, 55-74; n° 2: 1-16; n° 3, 1-24 (reprinted in Scanduva 1976).
- CORCORAN, J. (1973). Gaps between logical theory and mathematical practice. In: Bunge 1973, 23-30.
- CORCORAN, J., FRANK, W. & MALONEY, M. (1974). String theory. *The Journal of Symbolic Logic*, 39, 625-37.
- CORCORAN, J. (1974). Aristotle's natural deductive system. In: Corcoran (ed.) 1974.

- CORCORAN, J. (ed.) (1974). *Ancient Logic and its Modern Interpretations*. Dordrecht: Reidel Pub.
- CORCORAN, J. (1980). Categoricity. *History and Philosophy of Logic*, 1, 187-207.
- CORCORAN, J. (1981). A note on categoricity and completeness. *History and Philosophy of Logic*, 2, 113-119.
- CORCORAN, J. (1982). Review of Hinkikka 1980. *Mathematical Reviews* 82m:00016.
- CORCORAN, J. (1986). Undefinability tests and the Erlanger Program (abstract) *International Congress of Mathematicians*. Berkeley, Ca.
- CORCORAN, J. (1989). Argumentations and logic. *Argumentation*, 3, 17-43.
- DEDEKIND, R. (1888). The nature and meaning of numbers. In: Dedekind 1901.
- DEDEKIND, R. (1901). *Essays on the Theory of Numbers*. Lasalle, Ill.: Open Court (New York: Dover 1963).
- EDWARDS, P. (ed.) (1967). *Encyclopedia of Philosophy*. New York /London: Macmillan/Collier-Macmillan.
- FREGE, G. (1879). *Conceptual Notation and Related Articles*. Oxford 1972 (translated by Terrell Bynum).
- GÖDEL, K. (1930). Some metamathematical results on completeness and consistency. (abstract). In Gödel 1986, 140-43 (english translation by S. Bauer-Mengelberg facing the original German).
- GÖDEL, K. (1931). On formally undecidable propositions... In: Gödel 1986, 144-195 (english translation by J. van Heijenoort facing the original German).
- GÖDEL, K. (1986). *Collected Works*. Oxford: Oxford University Press (edited by S. Feferman *et al.*).
- HENKIN, L. (1950). Completeness in the theory of types. *The Journal of Symbolic Logic*, 15, 81-91.
- HINTIKKA, J. (1980). Aristotelian induction. *Revue Internationale de Philosophie*, 34, 422-439.
- KEYSER, C. (1922). *Mathematical Philosophy*. New York.
- MENDELSON, E. (1987). *Introduction to Mathematical Logic*. Monterey: Wadsworth & Brooks/Cole (3rd ed.).

- PASCAL, B. (1728) (posthumous). *De l'esprit géométrique et de l'art de persuader*. In: *Oeuvres complètes*. (Edited and annotated by J. Chevalier) Paris: Gallimard 1954.
- PRESBURGER, M. (1930). On the completeness of a certain system of arithmetic... *History and philosophy of logic*, 12, 225-233 (translated by D. Jacquette).
- RUSSELL, B. (1918). *Introduction to Mathematical Philosophy*. London: Allen.
- SCANDURA, J. (ed.) (1976). *Structural Learning II, Approaches and Issues*. New York/London: Gordon/Breach Science.
- SHANKS, D. (1962). *Solved and Unsolved Problems in Number Theory*. Vol. 1. Washington, D.C.
- SHAPIRO, S. (1991). *Foundations without Foundationalism, a Case for Second-Order Logic*. Oxford: Clarendon.
- SUPPES, P. (1957). *Introduction to Logic*. Belmont Ca.: Wadsworth Int. Group.
- TARSKI, A. (1935). On the concept of truth. In: Tarski 1956.
- TARSKI, A. (1944). The semantic conception of truth. *Philosophy and Phenomenological Research*, 4, 341-373.
- TARSKI, A. (1946). *Introduction to Logic*. Oxford: Oxford University Press (second edition).
- TARSKI, A. (1956). *Logic, Semantics, Metamathematics*. Oxford: Oxford University Press (second ed. edited and introduced by John Corcoran. Indianapolis: Hackett Pub. 1983).
- TARSKI, A. (1969). Truth and proof. *Scientific American*, 220, n° 6, 63-77, revised for *L'Age de la Science*, 1 (1969), 279-301.
- TARSKI, A. & GIVANT, S. (1987). *A Formalization of Set Theory Without Variables*. Providence, R.I.: American Mathematical Society.
- VEBLEN, O. (1904). A system of axioms for geometry. *Transactions of American Mathematical Society*, 5, 343-384.
- WHITEHEAD, A.N. & RUSSELL, B. (1910). *Principia Mathematica*. Vol. 1, Cambridge: Cambridge University Press.

BOOLOS' NONCONSTRUCTIVE PROOF OF GÖDEL'S THEOREM

Richard E. Vesley

Gödel's incompleteness theorem has attracted many different statements and proofs. However, it seems that only recently are there appearing versions that are nonconstructive. Gödel's original presentations, e.g. in (1931), as well as succeeding generalizations and expositions, (Kleene 1950, 1952) and others, arose from or in response to the Hilbert program. They were carefully and explicitly formulated, both in statement and proof, to meet the metamathematical standards of that program, thus being neutral between classical and constructive (intuitionistic) mathematics.

Two recent nonconstructive proofs are those of Boolos and Chaitin. We discuss here the proof of Boolos (1989), amplifying our remarks originally published in (Vesley 1989). Chaitin's proof is different. Boolos and Chaitin themselves have not called attention to this aspect of their work.

As we hope to make clear below, we do not explore the nonconstructivity of such proofs in an effort to categorize them as acceptable according to some philosophy of mathematics, unacceptable according to another. Rather, we aim only to determine just what has been proved. In general, we can anticipate that, whenever a nonconstructive proof establishes a theorem, a corresponding constructive proof actually establishes a stronger theorem. This stronger theorem is often not made explicit. But here we emphasize the distinction by stating the stronger version(s).

It is expected that nonconstructive proofs of a theorem will be tidier in general than constructive proofs. Boolos' presentation of Gödel's theorem and proof fulfills this expectation, and was

introduced by Barwise on its publication (1989), as «the most straightforward proof of this result I have ever seen». The approach differs fundamentally from Gödel's in that an analog to the Berry paradox is developed, rather than an analog to the Epimenides paradox. Less easily noticed, but of concern to us here, are these differences from Gödel's results: (1) though the existence of an undecidable sentence of arithmetic is established, no effective means is given in the proof for actually producing such a sentence, and (2) the undecidable sentence asserted to exist is recognized to be true only classically and not constructively. The best way to explore these points is to state with some care precisely what is proved by Boolos' methods and what by Gödel's.

1. Boolos' proof

Suppose a *correct* algorithm M is one which never lists a false sentence of arithmetic. A *truth omitted by M* is a true sentence of arithmetic not listed by M .

Boolos' proof establishes the theorem:

(I) If M is a correct algorithm, then there is some truth omitted by M .

Proof. We say number n is *named* by formula $F(x)$ under algorithm M if M lists the formula $\forall x(F(x) \leftrightarrow x=[n])$, where $[n]$ is the numeral for number n , i.e. $[n]$ is 0 preceded by n occurrences of the symbol S (for successor).

Fixing any number k , we can choose formula $F(x)$ to express: x is the least number not named by any formula with fewer than $10k$ symbols. It is possible then to choose n so $\forall x(F(x) \leftrightarrow x=[n])$ is true. We can convince ourselves that there is such n by observing that there are only finitely many different $F(x)$ having fewer than $10k$ symbols and that each such $F(x)$ can name at most one number.

Now by choosing k large enough, we can construct $F(x)$ to have fewer than $10k$ symbols. The key is that expressing $10k$, i.e. writing $[10k]$, uses only a few more than k symbols (uses

exactly $k+13$ symbols in the standard alphabet employed by Boolos).

Finally, after choosing such large enough k and a corresponding n , let G be $\forall x(F(x) \leftrightarrow x=[n])$. Then G is a true sentence, and precisely because G is true and $F(x)$ has fewer than $10k$ symbols, G must be unlisted by M .

For more details see (Boolos 1989).

2. Gödel's proof and generalizations

Gödel in (1931) asserted:

(II) There is an algorithm M_1 such that, if M is a correct algorithm of a certain kind (provided by certain formal systems for arithmetic), then M_1 applied to M yields a (classical and constructive) truth omitted by M .

Sharper insight into algorithms started with Turing (cf. Gödel's note in (Gödel 1963: 616)) and eventually produced:

(III) There is an algorithm M_1 such that if M is a correct algorithm, then M_1 applied to M yields a (classical and constructive) truth omitted by M . (Kleene 1952: Sec. 60)

2.1. Providing an algorithm

The first important distinction is that Gödel provided, and intended to provide, an algorithm M_1 . His intention was made explicit in his statements that his proof is «constructive» and «intuitionistically unobjectionable» (1963: 609). Kleene conveys this intention when he writes «it is important that the strictly finitary metamathematical proof of Gödel's theorem should be appreciated» (1952: 206).

Boolos' proof does not establish (II) or (III) but only (I), as we show in a moment. A project remains to see whether Boolos' approach can give a constructive proof of (I), which would amount to a proof of (III) weakened to require only a classical truth omitted by M . A further project would be to adapt the approach to establish (III) itself. Is Boolos' new insight

somehow inherently nonconstructive? We note that the arithmetical least number principle, which may be needed in any version of the Berry paradox, is not in general constructively acceptable.

To define his true but unlisted sentence, Boolos uses an object $[n]$, the numeral for a certain number n , where the specification of n depends in its last step upon the observation: (1) for any given correct algorithm M and any given natural number y ($=10k$ in the argument), there is a least number n not named, according to M , by any formula with fewer than y symbols. But (1) is untenable constructively. Number n cannot be found by any algorithm (it cannot «effectively», or in Gödel's word «actually», be obtained), working from M and y . And so, since there is no algorithm for this step, Boolos' proof does not allow us to conclude that there is an algorithm by which to obtain the desired sentence.

We sketch an argument to show that no algorithm can compute the n claimed to exist by (1). Let $\exists tP(s,t)$ be any formula of arithmetic with $P(s,t)$ simply an equation between numerical terms (a formula whose truth for any particular numbers s and t can be determined by a finite numerical computation). Several famous unsolved mathematical problems (for example, the Fermat problem) have the form, for some particular P and s : does $\exists tP([s],t)$ hold? For any given natural number s , let M_s be the following algorithm. For each natural number input t , M_s computes whether the equation $P([s],[t])$ holds or not (for the fixed s); if so, M_s lists the formula: (i) $\forall x(x=0 \leftrightarrow x=0)$. (If we wish M_s to be completely defined, we can add: otherwise M_s lists $0=0$.) Then M_s lists (i) if and only if for the given s , $\exists tP([s],t)$ is true. M_s is correct, and the description of M_s enables us to describe easily a code number m_s for M_s . Note that (i) is of form $\forall x(F(x) \leftrightarrow x=0)$, for $F(x)$ containing 3 symbols.

Now suppose there is an algorithm L which, working on the code number m of any correct algorithm M and on any y , produces the least number not named, according to M , by any formula with fewer than y symbols. Apply L to m_s and 4. If the value obtained is 0, then (i) is not in the output of M_s ; if the value obtained is not 0, then 0, being less than this value, is named

according to M_s by a formula containing fewer than 4 symbols, and thus, since M_s allows no other way in which any number could be named, (i) must be in the output of M_s . In the first case, we know from the definition of M_s that $\neg\exists tP([s],t)$ holds; in the second case, that $\exists tP([s],t)$ holds. So L , if it existed, would enable us to solve in a standard way for any P and s , the problem: does $\exists tP([s],t)$ hold? To reiterate the method of solution, simply compute m_s , apply L to m_s and 4 and see whether the result is different from 0. However, not only are many unsolved problems of mathematics of this kind, it is even a theorem of recursion theory that no such solution method can exist.

2.2. Providing a constructively true omitted sentence

The second point to note about the Berry paradox argument concerns the nature of the omitted truth itself, whose existence (but not constructive existence, as we have just seen) the argument guarantees.

The Gödel diagonal argument provides an omitted true sentence of the form $\forall xA(x)$, with $A(x)$ quantifier free. Arithmetic sentences of this simple form which are true classically are also true constructively (intuitionistically). This observation is crucial in understanding the full force of the original Gödel theorem: not only does any consistent classical formal system for arithmetic fail to prove all classical truths, but also it must fail even to prove all constructive truths (a much smaller set). (And therefore of course a formal system for constructive arithmetic must also be incomplete).

Unfortunately, the Berry paradox argument, in the form given by Boolos, does not establish this. The omitted true sentence identified in this second argument, i.e. G above in section 1, is of a more complicated quantifier form, and there is no guarantee that this sentence will be constructively true.

(Some readers may be uncertain about the notion of «constructive truth» in arithmetic. If we agree that atomic truths are both classical and constructive truths, then the notion of a

more complicated truth can be defined in the classical case by the Tarski definition, in the constructive case by the Heyting definition. The latter requires the existence of algorithms at critical points, as e.g. in verifying a universal statement).

*Department of Mathematics
State University of New York at Buffalo
Buffalo, New York 14214 USA*

Bibliographical references

- BOOLOS, G. (1989). A new proof of the Gödel incompleteness theorem. *Notices of the American Mathematical Society*, 36, n° 4, 388-390.
- GÖDEL, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 37, 349-360.
- GÖDEL, K. (1963). On formally undecidable propositions of Principia Mathematica and related systems I. (english translation by van Heijenoort of (1931) with additional footnotes by Gödel). In: J. van Heijenoort (ed.), *From Frege to Gödel*. Cambridge (USA): Harvard.
- KLEENE, S.C. (1950). A symmetric form of Gödel's theorem. *Indagationes mathematicae*, 12, 244-246.
- KLEENE, S.C. (1952). *Introduction to Metamathematics*. New York & Amsterdam: van Nostrand.
- VESLEY, R. (1989). The last word on Boolos' proof. *Notices of the American Mathematical Society*, 36, n° 10, 1352.

ARGUMENTS GÖDELIENS CONTRE LA PSYCHOLOGIE COMPUTATIONNELLE

Jacques Paul Dubucs

Le premier théorème d'incomplétude de Gödel établit que tout système formel S suffisamment riche contient, s'il est cohérent, un énoncé élémentaire G_S improuvable mais vrai («élémentaire» signifie ici: du type $\forall x \psi x$, ψ récursif). Cet énoncé affirme sa propre improuvabilité dans S . Comme il est, en fait, improuvable dans S , il est vrai.

Je viens de montrer que l'énoncé G_S est vrai, bien qu'il ne soit pas prouvable dans S . Je suis donc capable de reconnaître la vérité de plus d'énoncés arithmétiques que le système S ne peut en prouver (ou qu'une machine de Turing $T(S)$ associée à S ne peut en énumérer). Le même raisonnement pourrait s'appliquer à tout système formel suffisamment riche, et à toutes les machines qui énumèrent les théorèmes de ces systèmes. Donc je ne suis pas une machine de Turing, et les psychologues qui affirment le contraire ont tort. Telle est, sommairement exposée, la conclusion que l'on a cru pouvoir tirer des résultats de Gödel: l'esprit humain ne «fonctionne» pas comme une machine de Turing.

1. Les arguments qui visent à établir la supériorité de l'esprit humain sur les machines peuvent pécher de deux façons: soit en attribuant à l'esprit la possibilité de performances dont il est incapable, soit en déniaut aux machines la possibilité de performances dont elles sont capables. En un certain sens, l'argument anti-mécaniste esquissé ci-dessus tombe nécessairement dans l'un ou l'autre de ces travers.

Le raisonnement qui me convainc de la vérité de G_S est le suivant:

a) S est «suffisamment riche», c'est-à-dire que les relations arithmétiques récursives y sont fortement représentables: si ϕ est un prédicat récursif, alors pour chaque entier n l'énoncé $\phi(\bar{n})$ est prouvable si n satisfait ϕ et réfutable sinon.

b) G_S exprime «correctement» sa propre improuvabilité, c'est-à-dire qu'un entier qui ne satisfait pas le prédicat ψ est bien le numéro d'une preuve de G_S ¹.

Par conséquent si G_S était faux il existerait, d'après a), un entier n tel que $\neg \psi(\bar{n})$ soit prouvable: G_S serait donc réfutable. Mais, d'après b), G_S serait également prouvable. S serait donc incohérent. Donc G_S est vrai.

L'anti-mécaniste soutient que la performance cognitive dont je viens de faire preuve n'est pas à la portée de la machine $T(S)$. Il a tort. Car ou bien il pense que je viens de prouver que G_S est vrai, et alors il surestime cette performance: j'ai simplement montré que si S est cohérent, alors G_S est vrai. Ou bien il reconnaît que j'ai seulement prouvé que si S est cohérent, alors G_S est vrai, et dans ce cas il sous-estime les capacités de $T(S)$, puisque l'énoncé qui affirme que la cohérence de S entraîne la vérité de G_S est un théorème de S.

2. Il existe de l'argument anti-mécaniste une interprétation plus charitable que celle dont on vient de montrer l'absurdité.

Je ne peux, c'est entendu, *prouver* G_S que sous l'hypothèse de la cohérence de S. Mais si S est, en fait, cohérent, je peux prouver chaque instance de G_S sans recourir à l'hypothèse que S est cohérent. En effet je suis évidemment capable de reconnaître que la phrase « n n'est pas le numéro d'une preuve de G_S » est vraie, lorsqu'elle l'est. Et si S est cohérent cette phrase est vraie, et donc *reconnaissablement* vraie pour chaque entier n . En somme si S est cohérent je suis en mesure de prouver inconditionnellement $\psi(\bar{0})$, $\psi(\bar{1})$, $\psi(\bar{2})$...

Cette nouvelle performance, pas plus que la précédente, ne permet d'affirmer ma supériorité sur la machine $T(S)$. Car le rai-

¹ Rappelons que $\psi(x)$ est la formule ouverte $\neg \text{Pr}_S(x, \ulcorner G_S \urcorner)$ (« x n'est pas le numéro d'une preuve dans S de la formule G_S »).

sonnement qui me convainc de la vérité de $\psi(\bar{n})$ consiste simplement à vérifier, au cas où n serait le numéro d'une preuve, que la dernière formule de cette preuve n'est pas G_S : rien ici qui ne soit à la portée d'une machine, puisque la notion de preuve dans S est effective. De fait toutes les instances $\psi(\bar{0})$, $\psi(\bar{1})$, $\psi(\bar{2})$... de G_S sont bel et bien prouvables dans S si S est cohérent.

Le point sur lequel mes capacités de justification semblent excéder celles d'une machine est plutôt le suivant. Si je possède, pour chaque objet d'un certain domaine, une raison probante d'affirmer que cet objet satisfait le prédicat ϕ , alors je suis autorisé, en vertu de l'interprétation familière du quantificateur universel, à affirmer que l'énoncé $\forall x \phi x$ est vrai dans le domaine considéré. Ayant prouvé que chaque entier satisfait le prédicat ψ , je suis donc justifié à affirmer que $\forall x \psi x$, c'est-à-dire G_S , est vrai dans le domaine des entiers naturels. G_S est bien un énoncé que S est incapable de prouver, mais dont je suis en mesure de reconnaître la vérité.

3. A l'argument anti-mécaniste ainsi reconstruit peut être opposée une objection bien connue: le fait que je suis en mesure de justifier G_S établit que je ne suis pas la machine $T(S)$, mais il n'établit pas que je ne suis pas une machine. En effet, puisque le système S prouve $\text{Coh}(S) \rightarrow G_S$, le système $S' = S \cup \{\text{Coh}(S)\}$ prouve G_S ². Par conséquent une machine qui énumère les théorèmes de S' est, quant à elle, capable d'engendrer G_S . L'argument anti-mécaniste qui établit ma supériorité sur $T(S)$ échoue donc à établir ma supériorité sur $T(S')$.

C'est vrai. Mais l'anti-mécaniste peut répondre que la conclusion putative de son argument n'est pas la «supériorité» de l'esprit sur toutes les machines, c'est-à-dire une thèse du type

(AM) Il existe un énoncé arithmétique que je suis capable de justifier, mais qu'aucune machine de Turing cohérente (c'est-à-dire incapable d'engendrer $\bar{0} = \bar{1}$) ne peut engendrer

2 $\text{Coh}(S)$ est la formule $\forall x \neg \text{Pr}_S(x, \ulcorner \bar{0} = \bar{1} \urcorner)$.

mais plutôt la fausseté de toute thèse du type

M(T) La machine de Turing T peut, tout en restant cohérente, énumérer toutes les vérités arithmétiques que je suis capable de justifier.

En d'autres termes l'argument anti-mécaniste peut se présenter comme un «schéma de réfutation», destiné à montrer l'inadéquation de tout modèle mécanique de l'esprit qui pourrait être effectivement proposé. Entendue en ce sens, la réfutation de l'hypothèse selon laquelle T(S) serait un modèle de l'esprit n'est nullement astreinte à contredire l'hypothèse selon laquelle T(S') en serait également un. Et la possibilité de *modifier* la machine T(S) et de la transformer en une machine T(S') capable d'engendrer l'énoncé G_S ne saurait être invoquée contre l'anti-mécaniste: ce dernier est en droit d'exiger qu'on lui présente une machine qui, *tout en restant la même*, parviendrait à simuler le comportement cognitif sous-jacent à l'inférence qui conclut de $\psi(0)$, $\psi(1)$, $\psi(2)$, ... à $\forall x \psi x$.

Naturellement la correction éventuelle du schéma anti-mécaniste de réfutation n'invalidé pas, en toute rigueur, la thèse mécaniste selon laquelle il existe une machine de Turing cohérente capable d'énumérer la totalité des vérités arithmétiques que je suis en mesure de justifier: la réfutation de chaque instance de cette thèse établit que le mécanisme est au mieux une doctrine psychologique « ω -incohérente», qui soutient que l'esprit est une machine quoiqu'il soit établi pour chaque machine qu'il n'est pas *cette* machine. Bien que la doctrine en question soit probablement digne d'attention³, je ne discuterai pas ici la portée résiduelle d'un tel affaiblissement du mécanisme, et je me limiterai dans la suite à l'examen de la correction du schéma de réfutation lui-même.

3 En un certain sens elle a été explorée par Benacerraf (1967) qui pour sa part infère des résultats d'incomplétude de Gödel que nous ne sommes pas des machines de Turing, ou que si nous en sommes nous ne pouvons en tout cas pas déterminer notre propre «table d'instructions».

4. Une grande partie de la littérature⁴ consacrée aux implications anti-mécanistes éventuelles des résultats de Gödel témoigne d'une propension fâcheuse à envisager le problème sous l'aspect purement *extensionnel* de la classe des énoncés arithmétiques susceptibles d'être certifiés respectivement par une machine et par un agent humain: la question décisive semble être de savoir qui, de l'homme ou de la machine, triomphera de l'autre dans cette compétition. Il faut sans doute reconnaître que la pertinence des résultats de Gödel pour la psychologie dépend bien d'une hypothèse de cet ordre, puisque l'argument anti-mécaniste ne mériterait aucunement d'être discuté si l'on n'admettait pas d'emblée qu'un agent humain est en mesure de reconnaître comme vrais *au moins* autant d'énoncés arithmétiques qu'une machine de Turing associée à un système formel cohérent assez puissant n'est capable d'en engendrer. Mais une fois cette hypothèse «quantitative» accordée, c'est-à-dire étant admis que la psychologie susceptible d'être affecté par les résultats de Gödel est une discipline largement «idéale», qui postule chez les agents une forme d'«omniscience logique»⁵, l'évaluation raisonnée de l'argument anti-mécaniste doit tenir compte des aspects *qualitatifs* des modes de certification à l'oeuvre chez les humains et dans les machines.

Le raisonnement par « ω -induction» décrit plus haut est particulièrement instructif à cet égard. Ce qui semble me distinguer de la machine $T(S)$, c'est la transition que j'effectue - mais pas elle - de «pour tout n : $\psi(\bar{n})$ » à « $\forall x\psi x$ ». L'improuvabilité de $\forall x\psi x$ dans le système S , jointe à la prouvabilité dans S de chaque instance $\psi(\bar{n})$, s'explique tout naturellement par l'existence d'un modèle «non standard» de S , dont l'univers N^* contient, à côté de l'univers N des entiers naturels, d'autres objets que les entiers: interprétées dans N^* , les prémisses de l'inférence sont vraies mais sa conclusion est fausse. Quant aux raisons pour lesquelles j'effectue moi-même cette inférence, elles vont de soi: engagé dans la recherche d'énoncés vrais dans l'univers N des entiers naturels, c'est *dans cet univers* que j'interprète les variables qui figurent dans les énoncés que

4 Cette littérature est pour l'essentiel issue de Lucas 1961.

5 Cf. sur ce point Dubucs 1992.

j'examine. Ayant justifié $\psi(\bar{n})$ pour tout n , je peux ajouter « $\forall x\psi x$ » à la liste des énoncés justifiés, car pour moi c'est précisément *cela*, la signification de « $\forall x\psi x$ ». Dès lors la différence «quantitative» de mes capacités recognitionnelles et de celles de la machine $T(S)$ n'est que l'indice d'une différence, beaucoup plus significative, entre la structure qualitative de ces capacités. A l'inverse de la machine $T(S)$, dont le comportement démonstratif s'explique exhaustivement en termes computationnels, c'est-à-dire par référence aux seules propriétés *formelles* des symboles arithmétiques, je suis pour ma part guidé par la référence «attendue» de ces symboles, c'est-à-dire capable d'invoquer le fait que certains de leurs regroupements reflètent éventuellement les propriétés de la structure mathématique N que je *vis*. C'est la raison pour laquelle il est impossible de soutenir, comme le défenseur du mécanisme est tenté de le faire, que le comportement de la machine $T(S')$ est identique au mien. Erronée sur le plan «quantitatif» (contrairement à moi, $T(S')$ ne parvient pas à certifier G_S), cette affirmation est en outre insoutenable sur le plan «qualitatif»: l'aptitude de $T(S')$ à inscrire G_S sur son ruban de sortie est le pur résultat de son architecture combinatoire. $T(S')$, et c'est là sa différence cruciale d'avec moi, est *intentionnellement inerte* ⁶.

5. Dans la dimension sous laquelle il peut se manifester à un observateur extérieur, le travail mathématique est une activité de traitement de symboles, dont le résultat tangible est la répartition de certaines concaténations de symboles, les «formules bien formées», en deux classes (disjointes dans le meilleur des cas): la classe des formules acceptées, et celle des formules rejetées. En

6 Il faut cependant reconnaître que la sensibilité des processus inférentiels au contenu sémantique des symboles manipulés ne découle pas, en toute rigueur, de l'hypothèse selon laquelle la vérité de G_S est reconnaissable. Il se pourrait que le raisonnement qui conduit de $\psi(\bar{0})$, $\psi(\bar{1})$, $\psi(\bar{2})$,... à $\forall x\psi x$ se fonde sur la mobilisation d'une règle formelle d'un type nouveau, la «règle ω », qui autorise précisément l'inférence en question (sur cette règle, cf. J.-Y. Girard 1987, chap. 6) Comme un raisonnement de ce type n'est pas «effectif» (il ne peut pas être simulé par une machine de Turing), ceci suffit à établir la partie négative de la thèse anti-mécaniste. La thèse de la «sémantécité», quant à elle, n'en découle pas.

vertu d'une hypothèse particulièrement séduisante par son économie et son uniformité, la partie «non visible» du travail mathématique est homogène à sa partie visible, c'est-à-dire consiste elle aussi en un traitement de symboles: le fonctionnement de l'appareil cognitif pourrait être intégralement décrit en termes de manipulations symboliques écartant toute référence à un *contenu sémantique* éventuellement attaché aux symboles traités. Cette hypothèse «computationnaliste» suggère une explication «causale» de l'activité mathématique. Car chaque réalisation d'une formule (considérée comme «type» syntaxique) est un objet physique (ou neuro-physiologique) par lequel peuvent évidemment transiter des déterminations causales: le fait que l'appareil cognitif soit à l'instant t dans une certaine relation *matérielle* R à une instance de la formule A peut être cause du fait qu'il est à l'instant t' dans la relation R à une instance de la formule B . Le computationnaliste dispose alors d'une explication naturaliste assez plausible de l'activité cognitive consistant à «accepter» ou à «justifier» certaines formules arithmétiques: ce que nous décrivons, dans un vocabulaire épistémique qui exige analyse, comme une «opération intellectuelle» conduisant à «justifier» la formule A n'est que le processus causal déterminé par lequel l'appareil cognitif se place dans une relation matérielle donnée à un *token* de « A » (la relation matérielle en question est tout à fait analogue à celle d'un processeur informatique à un opérande qui figure dans un registre déterminé de la mémoire d'un ordinateur). Compte tenu du caractère compositionnel du symbolisme mathématique, une explication de ce genre est assurément capable de rendre compte de l'essentiel des propriétés logiques satisfaites par l'activité de justification, et notamment du fait que l'ensemble des formules justifiées est stable par les règles d'inférence qui préservent la vérité: puisque l'effet causal des occurrences du symbole « $A \supset B$ » est une fonction des effets causaux des occurrences des symboles « A » et « B », une inférence comme le modus ponens, qui spécifie une relation logique entre formules, possède une contrepartie causale, c'est-à-dire que l'on peut formuler une loi de transition affirmant que si l'appareil cognitif est à l'instant t dans la relation pertinente à des *tokens* de « A » et de « $A \supset B$ » il doit être à l'instant t' dans la même relation

à un *token* de «B». En bref le computationnalisme se propose de rendre compte de l'activité mathématique en invoquant le fait que la structure combinatoire des formules arithmétiques est capable d'*encoder* la totalité des aspects de leur contenu sémantique qui interviennent dans les processus inférentiels. Et plus généralement il prétend expliquer les activités cognitives comme des processus de manipulations de symboles *cognitivement impénétrables*⁷, c'est-à-dire de symboles à la référence desquels le sujet n'a jamais accès.

La plupart des philosophes hostiles au computationnalisme sont tentés d'admettre que les choses *pourraient* bien se passer comme le computationnaliste le prétend. Ils sont convaincus que l'intelligence humaine mobilise, au moins partiellement, la référence des symboles manipulés, mais ils se résignent généralement à reconnaître que les données de l'observation sont tout à fait incapables d'attester cette particularité du répertoire cognitif humain, et ils sont donc disposés à concéder - puisque l'explication computationnaliste du comportement cognitif des ordinateurs est évidemment la seule possible - que l'on pourrait concevoir des machines dont le comportement cognitif «manifeste» («le comportement d'entrée/sortie») serait indiscernable de celui d'un agent humain. Aussi concentrent-ils leurs critiques sur le célèbre test proposé par Turing (1950) pour remplacer la question vague «Les machines peuvent-elles penser?» (pour réussir le test, une machine doit pouvoir induire en erreur sur sa vraie nature un interrogateur qui ne la voit pas, mais se contente de lire ses réponses sur un téléscripneur). Ils objectent à ce test que la similitude du comportement cognitif manifeste ne suffit pas à garantir la similitude des processus sous-jacents à ce comportement: même si une machine répondait régulièrement par les mêmes «sorties» qu'un agent humain à une classe donnée d'«entrées», nous ne serions autorisés ni à en conclure que les processus cognitifs à l'oeuvre chez l'agent humain sont de nature computationnelle, ni à qualifier sérieusement de «compréhension» ou d'«intelligence» les capacités dont fait preuve la machine. L'argument favori des anti-computationnalistes consiste ici

7 L'expression est de Z.W. Pylyshyn, dont l'ouvrage Pylyshyn (1984) est une expression particulièrement élaborée des thèses computationnalistes.

à invoquer certaines *Gedankenexperimente* censées établir qu'une activité de manipulation symbolique, quelque «performante» qu'elle puisse paraître, ne saurait être qualifiée d'«intelligente» que si elle repose sur une «pénétration cognitive» des symboles manipulés. Ainsi, pour prendre un exemple inspiré de l'argument bien connu de la «chambre chinoise» de Searle (1980), il se *pourrait* qu'un locuteur chinois répondant en chinois à des messages chinois voie ses performances linguistiques égalées par un locuteur français ignorant le chinois, mais disposant d'un ensemble de consignes écrites en français du type «Si vous recevez le message « $\P \uparrow \textcircled{R} f \partial$ » (supposé être une phrase chinoise), envoyez le message « $\textcircled{O} \ddagger \Diamond \textcircled{Y}^{\text{TM}} \Box$ » (idem)»: dans une telle situation, souligne l'anti-computationnaliste, nous dirions du locuteur pour qui les idéogrammes chinois sont impénétrables, et qui détermine son comportement sur la seule base de leur *forme*, qu'il est parvenu à *simuler* la connaissance du chinois et non qu'il comprend «réellement» cette langue. Mais de telles paraboles, et c'est là leur fragilité, supposent en partie ce qui est en question, puisqu'elles se bornent à réaffirmer que nous ne sommes en aucun cas disposés à employer les mots «compréhension» ou «intelligence» là où les performances cognitives ne résultent pas de l'accès à un *sens*.

Un théoricien anti-mécaniste pourrait être tenté de soutenir qu'il détient avec l'argument issu des résultats de Gödel une arme d'une tout autre force contre la psychologie computationnaliste: puisque le traitement de la formule G_S semble bien *manifeste* la différence entre un dispositif qui manipule les symboles arithmétiques d'après leurs seules propriétés combinatoires et un dispositif sensible à leur contenu sémantique, pourquoi ne pas *accepter* le verdict du test de Turing? Un «engin sémantique» pourrait être discriminé de n'importe quelle machine de Turing T (au programme supposé connu) au vu de ses réponses à une classe bien déterminée de questions arithmétiques: il est celui qui répond «NON» à au moins une question du type «Acceptez-vous A ?», où A est l'un des axiomes du système formel S instancié par T , ou celui des deux qui répond «OUI» à la question «Acceptez-vous G_S ?». A s'en tenir à l'argument exposé dans le paragraphe 2, une «expérience cruciale» de ce type n'a malheureusement aucune

chance d'être concluante, quand bien même on la laisserait se poursuivre indéfiniment. Car une chose est de dire que la classe des énoncés assertés par un agent qui «vise» le modèle standard doit contenir « $\forall x\psi x$ » si elle contient « $\psi(\bar{n})$ » pour tout n . Et autre chose est de dire qu'un tel agent est en mesure d'en arriver au point où il a effectivement établi « $\psi(\bar{n})$ » pour tout n , même si l'on admet qu'il dispose d'un temps de calcul illimité, et même si l'on reconnaît qu'il est, pour tout n , en mesure d'établir effectivement « $\psi(\bar{n})$ »: s'il commence ses calculs à l'instant t_0 , il possède au plus n instances de $\forall x\psi x$ à l'instant t_n , et n'est donc en mesure de conclure $\forall x\psi x$ qu'à l'«instant» $t_{\omega+1}$ où il contemple et résume l'infinité dénombrable des résultats qu'il a obtenus jusqu'«ici».

Sauf à admettre - et qu'invoquerait-on pour justifier une telle hypothèse? - que les événements mentaux prennent place dans une temporalité dense et non pas discrète, les différences éventuelles entre le comportement cognitif des «engins sémantiques» et celui des «engins syntaxiques» ne peuvent donc être attestées par les données d'une observation indéfiniment prolongée, et l'invocation de l'«argument gödélien» reconstruit dans le paragraphe 2 ne met pas le théoricien anti-computationnaliste dans une situation sensiblement plus avantageuse que celle qu'il obtiendrait par un argument purement qualitatif comme celui de la «chambre chinoise»: le schéma de réfutation qui établit l'irrecevabilité de la thèse $M(T)$ est vide de tout contenu empiriquement manifestable.

6. La thèse majeure de l'anti-mécaniste est qu'un agent qui «vise» le modèle attendu de l'arithmétique doit être capable, contrairement à la machine $T(S)$, de reconnaître la vérité de G_S . Ce processus de reconnaissance comporte nécessairement une étape non mécanisable. Dans la version discutée jusqu'ici, cette étape non mécanisable est celle qui conduit à G_S à partir d'une infinité dénombrable d'énoncés dont chacun peut être engendré mécaniquement. Mais cette version de l'argument anti-mécaniste, on vient de le voir, ne permet aucune *manifestation* de sa propre correction: bien que la vérité de G_S (dans le modèle attendu) résulte

simplement de la vérité de chacune des instances numériques de G_S , la reconnaissance de la vérité de G_S ne peut pas, quant à elle, provenir de la reconnaissance «totalisée» de la vérité de chacune de ces instances.

Il existe toutefois une autre version de l'argument anti-mécaniste, pour ainsi dire duale de la première, qui semble exempte de ce défaut. Selon cette version, la disposition à «viser» le modèle attendu de l'arithmétique confère à qui la possède la capacité de reconnaître «intuitivement» la vérité éventuelle d'un certain énoncé dont G_S découle mécaniquement. Cet énoncé est celui qui exprime la cohérence de S . Si S est cohérent, $\text{Coh}(S)$ est un énoncé dont toutes les instances $\neg \text{Pr}_S(\bar{0}, \lceil \bar{0} = \bar{1} \rceil)$, $\neg \text{Pr}_S(\bar{1}, \lceil \bar{0} = \bar{1} \rceil)$,... sont, à nouveau, vraies et reconnaissablement vraies, puisque c'est une tâche mécanique que de décider, un entier étant donné, s'il est ou non le numéro d'une preuve dans S et, dans l'affirmative, si la preuve dont il est le numéro possède « $\bar{0} = \bar{1}$ » pour formule finale. Mais le raisonnement susceptible de conduire à la reconnaissance de la vérité de $\text{Coh}(S)$, contrairement au raisonnement analogue pour G_S , n'est pas une inférence fondée sur une infinité dénombrable de prémisses: c'est un raisonnement par induction (au sens ordinaire du terme) sur la longueur des preuves dans S . Il consiste à vérifier que les axiomes de S sont satisfaits dans le modèle attendu (donc que les énoncés dont la preuve dans S est de longueur 1 sont vrais dans ce modèle) et que les règles d'inférence de S préservent la vérité. Dans ces conditions tous les énoncés prouvables dans S sont vrais dans le modèle attendu, $\bar{0} = \bar{1}$ n'est donc pas prouvable dans S , $\text{Coh}(S)$ est donc vrai, et l'on peut en déduire G_S . Un raisonnement de ce type, qui ne comporte aucun processus inférentiel «actuellement infini», est visiblement capable de conduire à une classe de réponses *effectivement discriminable* de celles de la machine $T(S)$ dans le test construit dans le paragraphe précédent. La seule étape non mécanisable est la première, dans laquelle est évaluée la correction des axiomes de S dans le modèle visé: consistant à accepter ou à rejeter une classe d'énoncés *selon leur référence* (VRAI ou FAUX) dans l'interprétation attendue, elle n'est pas plus effectuable par une machine que ne le serait une

instruction du genre «Si le symbole observé est une réalisation du type T, faire A; sinon faire B».

7. J'examinerai pour terminer une objection possible à l'argument anti-mécaniste sous sa dernière version, celle qui suggère que la vérité de G_S peut être établie au travers de la reconnaissance intuitive de la correction éventuelle de S dans le modèle attendu de l'arithmétique. Cette objection sceptique vise à mettre en doute le gain en «manifestabilité» procuré par le passage de la première version de l'argument à la seconde.

Le test du paragraphe 5 n'est certainement pas capable de manifester la différence entre la machine $T(S)$ et n'importe quel agent (cohérent et «logiquement omniscient») sensible au contenu sémantique du symbolisme arithmétique. Car il ne suffit pas à discriminer de celui de $T(S)$ le comportement cognitif d'un agent qui «viserait» pour sa part un modèle non standard de l'arithmétique: un tel agent pourrait très bien reconnaître la correction des axiomes et des règles d'inférence de S sans pour autant asserter G_S , s'il s'avérait incapable de reconnaître sur cette base la cohérence de S. En effet l'affirmation selon laquelle aucune suite finie de formules de S n'est une preuve de $\bar{0} = \bar{1}$ est elle-même libellée dans un langage («suite finie») qui met implicitement en jeu la définition des entiers naturels, de sorte qu'un agent guidé par une conception «déviante» de la totalité de ces entiers pourrait estimer insuffisante la preuve par récurrence de l'improuvabilité de $\bar{0} = \bar{1}$: cette preuve, tout juste capable d'assurer que la longueur d'une démonstration dans S de $\bar{0} = \bar{1}$ ne peut être mesurée par aucun entier pouvant être obtenu par l'itération, un «nombre fini» (au sens standard) de fois, de l'opération «successeur» appliquée à zéro, laisse ouverte la possibilité de démonstrations de $\bar{0} = \bar{1}$ qui auraient une longueur finie (au sens non standard) ...

La forme finale de la thèse anti-mécaniste est donc la suivante: c'est la fausseté de chaque énoncé du type

$M'(T)$ La machine de Turing T est à la fois cohérente et capable d'énumérer tous les énoncés dont un agent *qui vise le*

modèle standard de l'arithmétique est capable de reconnaître la vérité.

Mais une thèse qui mentionne le «modèle standard de l'arithmétique» s'expose à une forme de scepticisme dont on peut trouver chez Wittgenstein et chez la plupart de ses disciples⁸ une expression particulièrement intéressante et radicale, et en vertu de laquelle il n'y a tout simplement pas de sens à dire que G_S est vrai (le premier théorème d'incomplétude de Gödel devrait alors, en toute rigueur, être formulé de manière purement syntaxique: chaque système formel suffisamment riche contient, s'il est cohérent, un énoncé indécidable). Selon Shanker (1989: 221 sq), qui a récemment défendu ce point de vue de manière particulièrement éloquente, le défaut majeur de la formulation sémantique du premier théorème d'incomplétude (selon laquelle G_S est vraie (dans le modèle standard) mais improuvable dans S) tient à ce qu'elle rend «purements externe» la connection entre une proposition mathématique et sa preuve: il est «incohérent», écrit-il, d'affirmer que G_S est à la fois vraie et improuvable. Si on la comprend correctement, cette objection ne se limite nullement à refuser toute signification à un énoncé dont les conditions de vérité transcenderaient notre aptitude à reconnaître si elles sont ou non remplies. Car l'aptitude à «viser le modèle standard» suffit, en vertu de ce qui précède, pour reconnaître la vérité de G_S . L'objection, beaucoup plus radicale, requiert en outre que l'aptitude recognitionnelle en question puisse *publiquement* se manifester par la production d'une classe d'énoncés assez riche pour déterminer sans ambiguïté les propriétés du modèle considéré. Mais comme la mention du «modèle standard» est, nous l'avons vu, inéliminable de l'argument anti-mécaniste, et que par ailleurs le premier théorème d'incomplétude de Gödel établit précisément l'impossibilité de définir (à un isomorphisme près) ce modèle à l'aide d'une classe récursivement axiomatisable d'énoncés (du premier ordre), l'objection wittgensteinienne vide l'argument anti-mécaniste de tout contenu.

8 Cf. par exemple Goodstein 1980.

Il existe à cette objection de bonnes et de mauvaises parades. L'une des moins convaincantes, quoique des plus usitées, consiste à invoquer la possibilité de caractériser le modèle standard au sein d'un langage d'ordre supérieur: l'arithmétique de Peano du second ordre, dans lequel l'axiome de récurrence est écrit sous la forme

$$(*) \quad \forall \phi \{ [\phi(0) \ \& \ \forall x(\phi(x) \rightarrow \phi(S(x)))] \rightarrow \forall x(\phi(x)) \}$$

est une théorie catégorique, dont l'unique modèle est, à un isomorphisme près, le modèle standard⁹. Mais l'affirmation selon laquelle le modèle standard peut être décrit sans ambiguïté dans le langage de la logique du second ordre suppose que ce langage lui-même soit compris sans ambiguïté, c'est-à-dire que soit donnée au quantificateur universel de second ordre qui figure dans (*) sa signification «attendue», dans laquelle la variable ϕ parcourt la *totalité* des sous-ensembles de $\mathbb{N}$ ¹⁰.

La réponse la plus raisonnable au scepticisme wittgensteinien consiste probablement à soutenir que notre impuissance à caractériser complètement des notions comme celle d'«entier» ou de «nombre fini» à l'aide d'une classe récursivement énumérable d'énoncés ne constitue pas un réel obstacle à l'intelligibilité et à l'absence d'ambiguïté de ces notions, et qu'en définitive l'interprétation «attendue» de l'arithmétique ne dépend pas du langage utilisé pour la décrire. Car les modèles «non standard» de l'arithmétique comportent tous des éléments qui ne peuvent pas être obtenus à partir de zéro par itération de l'opération «successeur», et la question de savoir si un objet peut ou non être engendré de cette manière ne soulève en pratique aucune espèce de difficulté: comme le remarque M. Dummett (1963: 193), «nous n'entretenons aucun doute sur le point de savoir si Jules César est ou non un nombre naturel». En d'autres termes, bien qu'il nous soit impossible d'obtenir, par le biais d'un accord sur une classe «présentable» d'énoncés, une *garantie publique* de l'identité de nos notions «privées» d'«entier» ou de «nombre fini», nous avons de bonnes raisons de supposer que nous par-

9 Cf. Montague 1965.

10 Cf. Shapiro 1985: 720-721.

venons à nous entendre lorsque nous disons d'une preuve dans un système formel qu'elle est un objet «fini» (même si nous ne spécifions pas de borne supérieure pour sa longueur). Or si une telle convergence est possible, qui ne se résume ni à une disposition communément partagée à asserter les énoncés constitutifs d'un certain corpus ni, *a fortiori*, à une propension commune à entretenir avec certaines *occurrences* de ces énoncés une relation matérielle déterminée, c'est qu'il existe au moins une notion dont la psychologie computationnaliste est incapable d'expliquer la maîtrise.

Appendice. On peut donner une forme plus précise à la question de l'incidence des résultats d'incomplétude de Gödel sur la psychologie spéculative des attitudes propositionnelles. Moyennant l'hypothèse d'omniscience logique mentionnée plus haut, un opérateur épistémique K supposé rendre compte des propriétés formelles du savoir satisfait certainement les axiomes suivants:

- $\vdash_T K(A) \rightarrow A$ (ce qui est su est vrai)
 - $\vdash_T K(K(A) \rightarrow A)$ (la propriété précédente est elle-même connue)
 - $\vdash_T K(A \rightarrow B) \rightarrow (K(A) \rightarrow K(B))$ (le *modus ponens* est maîtrisé)
- Si A est une tautologie, alors $\vdash_T K(A)$ (omniscience logique)

Nous pouvons ajouter sans contradiction au système T ainsi obtenu la conjonction Q de la classe finie des axiomes de l'arithmétique de Robinson. Mais si, soucieux de conformité à l'analyse computationnaliste des processus cognitifs, nous décidons de traiter le symbole K comme un prédicat métalinguistique attaché au *nom d'un énoncé*, et non plus comme un opérateur de type modal, associant une proposition à une *proposition*, nous obtenons le système T' défini par:

$$\vdash_T Q$$

$$\vdash_T K(\ulcorner A \urcorner) \rightarrow A$$

$$\vdash_T K(\ulcorner K(\ulcorner A \urcorner) \rightarrow A \urcorner)$$

$$\vdash_T K(\ulcorner A \rightarrow B \urcorner) \rightarrow (K(\ulcorner A \urcorner) \rightarrow K(\ulcorner B \urcorner))$$

Si A est une tautologie, alors $\vdash_T K(\ulcorner A \urcorner)$

système dont Montague (1963) a montré l'incohérence (l'opérateur K ainsi défini possède toutes les propriétés de l'opérateur de prouvabilité de Gödel, augmentées du «principe de réflexion» $\text{Pr}(\ulcorner A \urcorner) \rightarrow A$).

Signalons enfin que cette impossibilité logique de reconstruire la psychologie des attitudes propositionnelles sur une base purement computationnaliste a été récemment étendue par Thomason (1980) aux attitudes «solipsistes» comme la croyance, qui n'entraînent pas la vérité de leur corrélat.

*C.N.R.S. Institut d'Histoire et
de Philosophie des Sciences et des Techniques
Université de Paris-I Sorbonne
13, rue du Four; F - 75006 Paris*

Références bibliographiques

- ANDERSON, A.R. (ed.) (1964). *Minds and Machines*. Englewood Cliffs, N.J.: Prentice Hall (trad. fçse *Pensée et Machine*. Seyssel, 01420: Champ Vallon, 1983).
- BENACERRAF, P. (1967). God, the devil, and Gödel. *The Monist*, LI-1, 9-32.
- DUBUCS, J.-P. (1992). Omniscience logique et frictions cognitives. In: D. Andler *et al.* (éds), *Philosophie et Cognition*. Bruxelles: Mardaga.

- DUMMETT, M. (1963). The philosophical significance of Gödel's theorem. In: *Truth and Other Enigmas*. London: Duckworth, 1978.
- GIRARD, J.-Y. (1987). *Proof Theory and Logical Complexity I*. Napoli: Bibliopolis.
- GOODSTEIN, R.L. (1963). The significance of incompleteness theorems. *British Journal for the Philosophy of Science*, XIV, 208-220.
- LUCAS, J.R. (1961). Minds, machines, and Gödel. *Philosophy*, XXXVI (repr. in: Anderson 1964, 43-59).
- MONTAGUE, R. (1963). Syntactical treatment of modality, with corollaries on reflection principles and finite axiomatizability. In: R. Montague, *Formal Philosophy*. Yale U.P., 1974, 286-302.
- MONTAGUE, R. (1965). Set theory and higher-order logic. In: J. Crossley & M. Dummett (eds.), *Formal Systems and Recursive Functions, Proceedings of the Eighth Logic Colloquium, Oxford, July 1963*. Amsterdam: North-Holland Pub., 131-148.
- PYLYSHYN, Z.W. (1984). *Computation and Cognition*. M.I.T.: Bradford Books.
- SEARLE, J. (1980). Minds, brains, and programs. *The Behavioral and Brain Sciences*, III: 417-457 (repr. in: J. Haugeland (ed.), *Mind Design*. M.I.T.: Bradford, 3e éd., 1985, 282-306).
- SHANKER, S.G. (1989). Wittgenstein's remarks on the significance of Gödel's theorem. In: S.G. Shanker (ed.), *Gödel's Theorem in Focus*. London/New York: Routledge, 155-256.
- SHAPIRO, S. (1985). Second-order languages and mathematical practice. *The Journal of Symbolic Logic*, L, 714-742.
- THOMASON, R.H. (1980). A note on syntactical treatments of modality. *Synthese*, XLIV, 391-395.
- TURING, A.M. (1950). Computing machinery and intelligence. *Mind*, LIX (repr. in: Anderson 1964, 4-30).

REMARQUES SUR L'INTERPRETATION DU THEOREME D'INDECIDABILITE DE GÖDEL

Henri Lauener

Ce n'est pas seulement dans des cercles de mathématiciens que le théorème d'incomplétude de Gödel a été salué comme un résultat ouvrant de larges perspectives, mais aussi par des philosophes. Cependant l'interprétation philosophique du célèbre théorème ne me semble pas aussi irréprochable que sa démonstration purement formelle. On attribue entre autres à Gödel le mérite d'avoir ébranlé le programme de Hilbert en apportant la preuve que dans les mathématiques - à la différence de la logique - la classe des thèses (axiomes et théorèmes) ne coïncide pas avec la classe des propositions vraies et que, pour cette raison, la première ne peut être ramenée à la seconde. On affirme généralement que le résultat de ses travaux démontre l'incomplétude de tout système formel de l'arithmétique qui soit suffisamment fort et exempt de contradictions. Mes difficultés à comprendre cette affirmation proviennent du fait que, dans la partie formelle du célèbre article de 1931, on trouve des propositions mathématiques que l'on ne peut pas, à mon avis, transformer sans autre en des évidences philosophiques. Comment parvenons-nous des expressions purement arithmétiques aux interprétations, que l'auteur lui-même place en tête de son article? Pour les justifier, il faudrait préalablement expliciter les raisons pour lesquelles une sémantique formelle doit passer pour la (seule) interprétation correcte.

Qu'un philosophe soit prêt à accepter l'interprétation habituelle ou non, cela dépend de sa conception fondamentale des mathématiques. Gödel était un platonicien déclaré. Pour lui les objets mathématiques existent indépendamment du fait que nous autres, êtres humains, devons utiliser, pour en parler, une langue

spécifiquement élaborée dans ce but. Dans «What is Cantor's continuum problem?» (1947, il postule, à l'instar de Kant, une forme d'intuition pure, à l'aide de laquelle nous percevons des objets abstraits. Le fait que les axiomes s'imposent à nous comme vrais constitue pour lui une preuve de l'existence de ces objets. L'interprétation traditionnelle du théorème ne me semble en effet compréhensible qu'en faisant une présupposition de ce genre. Pour Gödel la propriété d'être une thèse était visiblement dépendante de la langue, mais pas celle d'être vrai; ainsi sa conception débouche en ultime analyse sur la thèse kantienne affirmant que certaines vérités *a priori* sont synthétiques. Celui qui ne partage pas cette conviction n'accordera pas la même signification philosophique aux résultats de Gödel. L'enthousiasme initial avec lequel certains membres du Cercle de Vienne ont accueilli le premier théorème reposait sur un malentendu, car celui-ci, en fait, entraînait moins la fin de la métaphysique que la restauration du platonisme.

Si l'on admet des vérités absolues, indépendantes de la langue et des théories, l'affirmation qu'il existe dans l'arithmétique une proposition vraie, mais non démontrable, peut paraître intuitivement plausible. Par contre, si l'on considère la question du point de vue de ma philosophie transcendantale pragmatiquement relativisée, qui exige une relativisation systématique de toute vérité à un contexte (théorique), cela n'est pas le cas. En effet, selon ce point de vue, on produit les objets mathématiques en utilisant un système linguistique fixé par des règles. En adoptant des conventions qui déterminent les conditions d'utilisation des termes logiques et mathématiques, nous *créons* par des définitions les vérités de la logique et des mathématiques («truth by fiat»). Par le choix du cadre conceptuel, nous fixons aussi une ontologie qui, dans le cas des mathématiques, consiste en objets abstraits, c'est-à-dire en objets produits par la seule utilisation d'un système linguistique; ceci a pour conséquence qu'en mathématiques, contrairement à ce qui se passe dans les sciences naturelles, il faut considérer les affirmations d'existence comme analytiques. Or, si la signification d'une expression mathématique est déterminée par l'ensemble des règles qui régissent son usage dans le cadre de la théorie en

question et s'il faut limiter le domaine des valeurs des variables en conséquence, il ne peut y avoir aucune vérité *externe* dans le sens du réalisme platonicien. La méthode de la relativisation en fonction des contextes aboutit à un concept purement interne, relativisé - conformément à la définition de Tarski - à un langage particulier.

Le fait qu'on peut conclure à la vérité d'énoncés mathématiques au moyen de méthodes purement analytiques exclut certes une éventuelle réfutation empirique, mais il ne signifie pas pour autant que toute révision soit exclue. Si, par exemple, quelqu'un se décide à utiliser la logique ou les mathématiques intuitionnistes au lieu de leurs versions classiques, il procède à une altération de tout l'appareil conceptuel en redéfinissant les termes en question. Comme aucune preuve objective de la vérité externe du système, considéré comme un tout, n'est possible, les raisons qui le poussent à faire ce choix doivent être de nature pragmatique. Dans la mesure où les axiomes de la géométrie se bornent à stipuler comment nous devons parler des objets spatiaux et de leurs relations, il ne faut pas considérer les différents systèmes géométriques comme des théories rivales qui rendraient compte d'un même domaine de réalité donné de façon neutre, mais bien plutôt comme des systèmes conceptuels différents. Ces deux exemples illustrent la relativité de la notion d'analyticité telle que je la conçois; j'insiste sur le fait que celle-ci ne repose nullement sur la présupposition de significations immuables, comme entités platoniciennes; elle met bien au contraire un accent particulier sur le fait que des expressions homonymes peuvent avoir des significations différentes selon leur usage dans des contextes différents. Ainsi, il peut se produire que la même phrase soit utilisée dans une théorie scientifique comme définition (analytique) et, dans une autre théorie, comme loi naturelle (synthétique).

Ces indications générales sur ma conception des mathématiques devraient déjà faire apparaître les raisons qui suscitent en moi un doute quant à la portée philosophique du théorème de Gödel. Dans ce qui suit, j'aimerais mettre en évidence les points particuliers, où résident à mon avis les difficultés. Afin d'éviter certaines erreurs d'interprétation que les mathématiciens ont re-

prochées aux philosophes, il est important de distinguer les trois systèmes (linguistiques) que Gödel introduit. D'une part, nous avons un système formel, P , qui représente pour l'essentiel une version (non interprétée) de l'axiomatique de Peano exprimée dans la notation de la logique des prédicats (avec des nombres comme individus et le symbole de fonction pour la relation de succession, en tant que terme primitif); les formules de ce système sont déductibles ou non déductibles - par contraste avec les formules vraies ou fausses d'un système interprété. Le but visé est de prouver que ce système contient des propositions non-décidables. On utilise la langue naturelle comme métalangue M , dans laquelle on discute sur les expressions de P et dont les énoncés sont vrais ou faux. Enfin, nous avons encore besoin du système non formalisé de l'arithmétique récursive A , dans lequel nous formulons des propositions (vraies ou fausses) sur les nombres naturels et leurs relations. Or la célèbre trouvaille de Gödel, qu'on a baptisée la numérotation de Gödel, a consisté en la découverte d'un procédé permettant d'attribuer à chaque expression de P un nombre naturel dans la métalangue M . On affirme parfois - en donnant au terme un sens quelque peu mystérieux pour moi - que l'on obtient ainsi une *interprétation* mathématique de P . La prudence nous commande pourtant de souligner qu'il ne s'agit pas d'une réduction ontologique - on ne déclare pas que les nombres naturels et les expressions métalinguistiques sont identiques! - mais seulement d'une correspondance biunivoque, qui doit servir à démontrer l'identité de la structure des deux domaines. La correspondance se limite au fait qu'à chaque symbole on associe un nombre naturel, à chaque formule une suite de nombres naturels et à chaque preuve une suite de suites de nombres naturels. Il s'agit donc d'une interprétation dans un sens purement formel. Pour démontrer l'incomplétude de l'arithmétique, il nous faut d'abord montrer que toutes les expressions de l'arithmétique récursive primitive A peuvent être corrélées à des expressions dans P de telle façon qu'une formule (respectivement, la négation d'une formule) soit démontrable dans P , si la proposition correspondante dans A est vraie (respectivement, fausse). Il nous faut de plus prouver que le système P , pour autant qu'il soit exempt de contradiction, doit

être incomplet en ce sens qu'il doit contenir au moins une formule qui, de même que sa négation, n'est pas démontrable. On peut ensuite prouver que l'incomplétude de P est liée à l'incomplétude de A .

Grâce au procédé de la numérotation de Gödel, nous sommes en mesure de produire un pendant arithmétique du concept méta-mathématique de la démonstration. Au prédicat de la métalangue ' v_1 est une démonstration dans P pour v_2 ' [v_1 désignant une variable dont le domaine est constitué de suites de suites d'expressions et ' v_2 ' une variable dont le domaine est constitué de suites d'expressions] correspond dans A la fonction propositionnelle ' xBy ', que l'on transforme en une proposition vraie en substituant à ' y ' le nombre de Gödel d'une formule et à ' x ' le nombre de Gödel de sa preuve. Nous pouvons alors définir pour P

$$Dem\ x \leftrightarrow (Ey)\ yBx ,$$

de telle façon que ' $Dem\ x$ ' soit vraie précisément lorsque ' x ' a la valeur du nombre de Gödel d'une formule démontrable dans P .

Pour formuler la proposition-clef de la preuve, c'est-à-dire V , qui exprime la correspondance, il nous faut définir quelques autres fonctions récursives primitives. Le pendant arithmétique du concept métalinguistique de substitution est l'expression

$$Sb\ (x \frac{y}{z}) ,$$

qui désigne le nombre de Gödel de la formule qui résulte de la substitution, dans la formule dont le nombre de Gödel est x , de la variable libre dont le nombre de Gödel est y par le terme dont le nombre de Gödel est z . Posons en outre que

$$Z\ (x)$$

désigne le nombre de Gödel du symbole représentant x dans P^1 et que

$$\text{Neg } (x)$$

désigne le nom du nombre de Gödel de la négation de la formule dont le nombre de Gödel est x . Enfin, nous aurons encore besoin de l'opération de la généralisation universelle:

$$x \text{ Gen } y$$

désigne le nombre de Gödel de la formule qui résulte de ce que nous plaçons le quantificateur universel ' $(\forall)$ ' devant la formule dont le nombre de Gödel est y (où x est le nombre de Gödel de la variable ' v '). Nous obtenons ainsi l'outillage nécessaire pour exprimer l'affirmation suivante, qui est à mon avis difficile à saisir intuitivement:

«Le fait intuitif que toute relation récursive est définissable dans le système P (si l'on donne leur signification habituelle aux formules de ce système) se trouve exprimé précisément et *sans* qu'il soit fait référence à aucune interprétation des formules de P , dans le théorème suivant:

Théorème V. Pour toute relation récursive $R(x_1, \dots, x_n)$, il existe un SIGNE DE RELATION r à n places (avec les VARIABLES LIBRES $u_1, u_2, \dots, u_n$), tel que pour tous les n -uples de nombres $(x_1, \dots, x_n)$, on ait:

$$R(x_1 \dots x_n) \rightarrow \text{Dem } [Sb(r \overset{u_1}{Z(x_1)} \dots \overset{u_n}{Z(x_n)})]$$

$$\overline{R(x_1, \dots, x_n)} \rightarrow \text{Dem } [\text{Neg } Sb(r \overset{u_1}{Z(x_1)} \dots \overset{u_n}{Z(x_n)})]$$

(Traduction française 1989: 125-126; édition allemande 1931: 186).

¹ On interprète Z comme une fonction qui représente des nombres par le nombre de Gödel de leur représentation dans P .

Avec cette proposition, nous avons atteint le point d'articulation qui est déterminant pour la conception selon laquelle la notion syntaxique de la démontrabilité dans P doit être un pendant du concept sémantique de la vérité dans l'arithmétique récursive. La démonstration que P , pour autant qu'il est exempt de contradiction, contient une formule qui n'est ni démontrable, ni réfutable, s'appuie essentiellement sur cette proposition. Si nous considérons le cas particulier:

$$\overline{x B (Sb (y_{Z(y)}^{19}))} \rightarrow Dem (Sb (q_{Z(x) Z(y)}^{17 \ 19}))$$

et si nous admettons qu'est vrai l'antécédent de toutes les conditionnelles qui résultent de la généralisation universelle des variables ' x ', nous obtenons alors la proposition arithmétique:

$$G_A: (x) (x B (Sb (p_{Z(p)}^{19})))$$

[p est la généralisation de q : $p = 17 \ Gen \ q$]

Si G_A est vraie, la formule qui a le nombre de Gödel

$$G: Sb (p_{Z(p)}^{19})$$

n'est pas démontrable dans P , et par conséquent la proposition métalinguistique suivante:

$$G_M: G \text{ n'est pas démontrable dans } P$$

doit être également vraie. Nous arrivons donc à la conclusion que ' G_A est vraie', ' G_M est vraie' et ' G n'est pas démontrable dans P ' sont équivalentes. Gödel a prouvé que l'affirmation que G est démontrable et l'affirmation que la négation de G est démontrable conduisent toutes deux à une contradiction.

J'ai déjà laissé entendre que pour moi la difficulté réside en mon incapacité à comprendre la relation exacte entre l'incomplétude de P et l'incomplétude de A . Mes doutes ne por-

tent pas sur la preuve formelle démontrant que la proposition ' G_A doit être vrai, si P est exempt de contradiction', mais bien sur l'argument selon lequel G devrait signifier la même chose que G_A , parce que nous interprétons P arithmétiquement. Que doit-on entendre ici par 'signifier la même chose'? On ne saurait penser à une relation extensionnelle, puisque en ce cas les objets seraient identiques, de sorte qu'il ne nous serait pas possible d'établir une correspondance biunivoque entre deux domaines *différents*. Que faisons-nous exactement lorsque nous *interprétons* des énoncés métalinguistiques portant sur des expressions comme des énoncés arithmétiques portant sur des nombres? La confusion est créée par les affirmations que G_M parle de G , que G_A parle des nombres naturels, que G , en tant que formule d'un système non interprété, ne parle de rien du tout et enfin que G est interprété comme G_A , affirmations qui sont incompatibles à mon avis. Car si elles étaient toutes vraies, G_M serait ambigu, dans la mesure où l'énoncé devrait à la fois traiter de rien et des nombres naturels. Nous aurions donc d'une certaine façon deux versions du théorème - une syntaxique et une sémantique -, qui n'ont rien à voir l'une avec l'autre. La seconde version recèle d'ailleurs en elle le risque d'un paradoxe du fait de son caractère auto-référentiel. Je ne pense pas que l'on puisse conjurer ce risque de manière convaincante en arguant qu'il ne s'agit pas d'une identification, mais bien d'une correspondance biunivoque.

Ma tentative d'interpréter intuitivement le premier théorème de Gödel échoue finalement sur le doute que j'ai quant à la nature mathématique des énoncés censés exprimer la correspondance. Je n'arrive pas à me persuader que l'antécédent et le conséquent de ces conditionnelles sont de même nature. L'interprétation d'après laquelle la formule G exprime sa propre indémontrabilité ne pourrait acquérir son éventuelle crédibilité intuitive qu'en présupposant l'interprétation standard de l'arithmétique et en concédant que B exprime numériquement la relation de preuve '*Dem*'. Si toutefois on n'exige pour B rien de plus que la satisfaction des conditionnelles exprimant la correspondance, une telle utilisation du mot 'preuve' paraîtra arbitraire du fait qu'on ne présente aucune raison pour la justifier. Il me

semble insuffisant de rétorquer que des systèmes formels pour lesquels la proposition V n'est pas valable ne pourraient pas raisonnablement être interprétés comme des systèmes mathématiques. D'ailleurs, comme le modèle standard n'est pas le seul, on se demandera s'il ne pourrait pas y avoir d'autres systèmes, dans lesquels cette condition serait satisfaite sans qu'ils ne contiennent de formule indécidable.

Mon impuissance à comprendre le contenu philosophique du théorème de Gödel provient de ma conception normative de la logique et des mathématiques, qui se distance fondamentalement de la vision platonicienne et descriptive de Gödel, d'après laquelle nous découvrons et décrivons de prétendues vérités (au moyen d'une intuition intellectuelle). Accepter que des formules dépourvues de sens puissent être équivalentes à des énoncés vrais de la métalangue me semble déraisonnable; ceci reviendrait à affirmer qu'il existe des vérités inexprimables, qui se trouvent au-delà d'un système des mathématiques constitué par des règles. Pour pouvoir utiliser le prédicat sémantique 'vrai' dans le sens de Tarski, il nous faut déterminer un domaine fixant les valeurs des variables qui ne peut être donné que de façon interne, c'est-à-dire dans le cadre d'une théorie établie et axiomatiquement construite. Pour cette raison, il est impossible de définir un concept *externe* de la vérité, autant pour les mathématiques que pour les sciences naturelles; ceci vaut d'ailleurs également pour le concept de la démontrabilité, qu'on ne peut introduire que relativement à un système donné. Selon ma conception, la signification des énoncés mathématiques est fixée par des règles intra-linguistiques, qui ne décrivent pas une concordance existante entre des expressions et une «réalité» platonicienne ou empirique, mais qui, en tant que normes, *prescrivent* l'usage des termes dans le système en question. De mon point de vue, il faut insister sur le fait que toute connaissance dépend de nos moyens d'expression et que toute vérité *a priori* est analytique, dans le sens qu'elle est produite par l'adoption d'un système linguistique. La perspective platonicienne pèche précisément par le fait qu'elle présuppose des vérités synthétiques *a priori*.

En vue du problème que pose l'interprétation philosophique des théorèmes de Gödel, il est indispensable d'élucider exactement ce que leur auteur obtient par son procédé. On nous assure que l'arithmétisation nous permet d'«interpréter» comme des expressions portant sur des relations arithmétiques les expressions de la métalangue dans laquelle nous parlons du calcul. Dans la mesure où l'on n'entend pas par là opérer une simple traduction dans une autre langue ou introduire une nouvelle notation, on prétend convertir des propositions de la métalangue en des propositions (nécessairement vraies) exprimant des relations purement mathématiques. Comme d'autre part on nous assure qu'il ne s'agit pas d'une réduction ontologique, il ne me reste plus qu'à jeter l'éponge. En effet, je ne vois pas comment s'opère la transition de la simple correspondance biunivoque à une interprétation dans le sens propre du terme, qui confère un contenu aux symboles. Si l'on transforme, par arithmétisation, en une proposition mathématique vraie une règle disant qu'une suite de formules représente la preuve d'une formule déterminée si et seulement si le nombre de Gödel de la démonstration se trouve dans la relation «*Dem* (x, y)» avec le nombre de Gödel de la conclusion, alors on commet - me semble-t-il - une erreur analogue à celle connue sous le nom de «*naturalistic fallacy*». Je suspecte précisément cette transmutation subreptice d'être à l'origine de mon incompréhension de l'interprétation que l'on donne du théorème.

Dans les remarques informelles qu'il place au début de son article, Gödel constate que du fait que ' $R(q); q$ ' affirme sa propre indémontrabilité découle immédiatement sa vérité, puisque, en effet, ' $R(q); q$ ' n'est pas démontrable². Mais comment la construction d'une proposition arithmétique qui «représente» l'affirmation métalinguistique de sa propre indémontrabilité est-elle possible? On prétend échapper au paradoxe sémantique parce que l'énoncé dans P dont le nombre de Gödel est i n'est pas auto-référentiel, malgré le fait que son prédicat soit corrélé avec la notion métalinguistique de la démontrabilité. Delong,

2 Cf. *op. cit.*, p. 176. Il me semble caractéristique de l'utilisation vague du mot 'vrai' avant la célèbre définition de Tarski que Gödel lui-même emploie l'expression 'correct' ('richtig').

par exemple, tout en admettant qu'il a en effet le nombre de Gödel i , avance qu'il n'exprime pas lui-même ce fait³. Cet argument me semble pourtant spécieux, car je ne vois pas comment l'énoncé en question pourrait ne pas être auto-référentiel, s'il doit passer pour une proposition authentiquement mathématique.

Bien que je ne prétende pas connaître l'«essence» des mathématiques, je crois qu'il y a de bonnes raisons de repousser le platonisme. On peut résumer la conception relativiste que je lui oppose en insistant sur le fait qu'en pratiquant les mathématiques nous ne décrivons pas des états de choses mathématiques, mais que nous les créons nous-mêmes par l'utilisation de certains systèmes linguistiques. Dans la mesure où j'attribue un rôle constitutif à la langue dans l'ensemble des activités scientifiques et que je considère comme inéluctable la différenciation entre la langue-objet et la métalangue, il m'est difficile de trouver un sens à «l'interprétation» d'énoncés méta-mathématiques par des énoncés arithmétiques. Comme les similitudes de structure mises en lumière par Gödel me semblent être de nature externe, je doute que ses théorèmes ouvrent réellement toutes les perspectives philosophiques qu'on leur attribue. En effet, ma conviction profonde que le sens des expressions linguistiques est établi de façon strictement interne, c'est-à-dire à l'aide de règles d'utilisation fixées dans un contexte théorique, m'empêche de voir un rapport significatif entre M et A . Je pense que les énoncés mathématiques ne peuvent être compris que relativement à un système donné ou une théorie établie. Considéré de ce point de vue 'vrai, mais indémontrable (à l'intérieur d'un système!)' ne peut que donner l'impression d'une contradiction.

Mes remarques critiques ne concernent pas la construction formelle de la démonstration, qui est - autant que je peux en juger - irréprochable. Il me semble, en revanche, que les interprétations philosophiques qu'on y a ajoutées ne sont pas vraiment

3 Cf. entre autres Howard DeLong, *A Profile of Mathematical Logic*. Reading (Mass.): Addison-Wesley Pub. Co 1971, p. 176 sq. et S.G. Shanker, Wittgenstein's Remarks on the Significance of Gödel's Theorem, in: *Gödel's Theorem in Focus*. London/New York: Routledge, 1988, p. 223.

satisfaisantes si on les juge d'après les critères de ma propre philosophie.⁴

*Séminaire de Philosophie
Université de Berne
Falkenplatz 16, CH - 3000 Berne*

Références bibliographiques

- GÖDEL, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, *Monatshefte für Mathematik*, 38, Vienne. (Traduction française in: E. Nagel *et al.* (eds), *Le théorème de Gödel*. Paris: Seuil, 1989, 105-143.
- GÖDEL, K. (1947). What Cantor's continuum problem? *American Mathematical Monthly*, 54.

4 Je remercie Markus Sprenger et Jürg Freudiger, qui m'ont rendu attentif à une erreur et à des formulations inexactes dans la première version de ce texte. Dans une communication personnelle, le premier souligne l'importance du théorème de Gödel pour les mathématiques, en signalant qu'en informatique il permet en particulier d'exclure *a priori* de nombreux problèmes comme insolubles pour l'ordinateur. Je remercie également Alain Perrinjaquet qui m'a grandement facilité la tâche en établissant une excellente traduction de mon texte écrit en allemand. Je me suis permis de modifier un certain nombre de formulations qui me semblaient déformer ma pensée.

GÖDEL'S INCOMPLETENESS THEOREM AND THE PHILOSOPHY OF OPEN SYSTEMS

Carlo Cellucci

1. Introduction

In recent years a number of criticisms have been raised against the formal systems of mathematical logic. The latter, qualified as *closed systems*, have been contrasted with systems of a new kind, called *open systems*, whose main feature is that they are always subject to unanticipated outcomes in their operation and can receive new information from outside at any time (cf. Hewitt 1991). While Gödel's incompleteness theorem has been widely used to refute the main contentions of Hilbert's program, it does not seem to have been generally used to point out the inadequacy of a basic ingredient of that program - the concept of formal system as a closed system - and to stress the need to replace it by the concept of formal system as an open system.

A partial exception seems to be provided by van Heijenoort who states:

«The notion of formal system, introduced by Frege in 1879, had become by then the accepted standard of precision in the foundations of mathematics. It seemed to embody the Aristotelian ideal of a perfect deduction from first principles. Gödel's results, by showing that mathematics cannot be completely and consistently formalized in one system, shattered this ideal. The bounds of mathematics cannot be those of one formal system. Since mathematics has often been regarded as the standard of rational knowledge that other sciences should strive to attain, Gödel's theorems seem to acquire significance for the whole body of human knowledge; they certainly establish that the old ideal of a deductive system cannot be maintained» (1967: 356).

Here van Heijenoort rightly points out the impact of Gödel's result on the concept of formal system as a closed system but does not suggest any alternative to such a concept.

In this paper, on the one hand, I want to stress the role of Gödel's incompleteness theorem in showing the inadequacy of the concept of formal system as a closed system, and, on the other hand, I want to point out the interest of the concept of formal system as an open system, which is essential both for current developments in artificial intelligence and for the emergence of a new paradigm of logic, alternative to mathematical logic: *computational logic* (see also Cellucci 1990, 1992).

2. The logicist concept of formal system

At least in its most elementary form, the notion of open system is not a new one. One of its ancestors is provided by Frege's concept of formal system, but it has also much older ancestors (see section 9 below).

It is usually claimed that «Frege was the first to present, with all the necessary accuracy, the notion of formal system» (van Heijenoort 1985: 11). The value of such claim depends on how it is interpreted. It is true if it is meant in the sense that Frege was the first to present a notion of formal system. It is false if it is meant in the sense that he introduced the notion of formal system which has been used in mathematical logic since the Thirties.

The latter interpretation is given, for instance, by Dummett in his book on Frege. Dummett acknowledges that Frege's aim in constructing his formal systems was not to make proofs themselves the objects of mathematical investigation, but only «to attain the ideal of that rigour to which the whole of nineteenth-century mathematics had been striving» (1981: xxxiv-xxxv).

On the other hand Dummett also claims that «Frege's calculus was a formal system in the modern sense» (1981: xxxiv). This view, that Frege's notion of formal system was the same as the notion of formal system of current mathematical logic, seems to me quite misleading. A basic difference stems from the fact that it

is an important part of Frege's conception of logic that nothing can be, or has to be, said outside the system: since logic is the system, anything that can be said must be said within the system. This was called by Sheffer the 'logocentric predicament' of logicism (cf. Sheffer 1926). A consequence of this position is that, since logic is the universal system within which every rational discourse must be made, there is no law court outside logic by which logic can be judged: no metasystematic question on logic can be raised (cf. van Heijenoort 1985: 13; Goldfarb 1979; Cellucci 1987).

Indeed, no metasystematic question on logic was ever raised by Frege. Of course Frege was fully aware that, in order to introduce a formal system, one needs formal rules that are not expressed in the language of the system, but considered such rules as void of any intuitive logic. Indeed, in the *Begriffsschrift* he introduces a number of fundamental principles of thought

«in order to transform them into rules for the use of our signs. These rules and the laws whose transforms they are cannot be expressed in the ideography because they form its basis» (1967: 28).

Since such rules cannot be expressed in the ideography, they are just rules for the use of signs, void of any intuitive logic (cf. van Heijenoort 1985: 13).

A similar position was held by Russell. But, as regards the concept of formal system, Russell took a backward step with respect to Frege. This was stressed by Gödel maintaining that the system of *Principia Mathematica* is greatly lacking in formal precision, not only because it provides no precise statement of the syntax of the formalism, but also because it omits syntactical considerations, even in cases where they are necessary for the cogency of the proofs (Gödel 1964).

In one respect, however, Russell went even further than Frege: not only he never raised any metasystematic question on logic, he even considered impossible to raise such a question.

For example, in the first edition of *The Principles of Mathematics*, Russell claims that the method used in geometry to establish the independence of the axiom of parallels cannot be

used to establish the independence of his axioms for propositional calculus, because all such axioms

«are principles of deduction; and if they are true, the consequences which appear to follow from the employment of an opposite principle will not really follow» (1937: 15).

Furthermore, in the introduction to the second edition of *The Principles of Mathematics*, Russell says:

«It is difficult to see any way of proving that the system resulting from a given set of premisses is complete, in the sense of embracing everything that we would wish to include among logical propositions» (1937: xii).

To appreciate the interest of this statement, note that it was made a number of years after the discovery of Gödel's result. The reason for Russell's statement was that, from the viewpoint of Frege and Russell by which logic is the universal system within which every rational discourse must be made, the question of completeness, in the metasystematic sense of current mathematical logic, simply did not arise.

The idea that logic is the system and that anything that can be said must be said within the system, together with its corollary that no metasystematic question can be raised, was also shared by Wittgenstein. Indeed it is one of the main themes of the *Tractatus* where it is stated that, since the limits of the world are also the limits of logic, one cannot say in logic that this there is in the world while that there is not since «otherwise logic must get outside the limits of the world: only in this way one could consider these limits from the other side also» (Wittgenstein 1961: 5.61). Here Wittgenstein makes fully explicit the implications of Frege's conception of logic as a total system for which no metasystematic question can be raised, and carries them to extremes: raising metasystematic questions would entail getting outside the limits of the world.

3. The logicist view of logic as an open system

The view of Frege, Russell and Wittgenstein on logic as the universal system implies that logic is an open system. This can be seen as follows.

Frege's first motivation in the *Begriffsschrift* was to establish the fundamental principles of mathematics: «I became aware of the need for a *Begriffsschrift* when I was looking for the fundamental principles or axioms upon which the whole of mathematics rests» (1969: 1). This raises the problem: once a certain number of principles have been laid out as the fundamental principles or axioms upon which the whole of mathematics rests, how can one make sure that such principles are complete?

Russell's answer to this problem was that the only question of completeness that could be raised was completeness in an empirical sense: do the rules exhaust the intuitive modes of reasoning actually used in science?

All the works of Frege and Russell, from the *Begriffsschrift* to *Principia Mathematica*, can be regarded as steps towards establishing completeness experimentally. This concept of completeness in an empirical sense is stated in the preface to *Principia Mathematica* by claiming that the evidence in favour of any theory on the principles of mathematics «must always be inductive, i.e. it must lie in the fact that the theory in question enables us to deduce ordinary mathematics» (Whitehead and Russell 1927: v).

The problem of completeness of *Principia Mathematica* had for Russell this empirical meaning, not the metasystematic meaning familiar from current mathematical logic. For him, existing mathematics was a datum to be explained in logical terms: one need only show that, starting from the axioms of *Principia Mathematica*, «such datum can be reconstructed» (Russell 1937: xii). What worried Russell was not that the system of *Principia Mathematica* could turn out to be incomplete in the metasystematic sense, but only that it could turn out to be incomplete in an empirical sense, i.e. that some parts of ordinary

mathematics could not be reconstructed from the axioms of *Principia Mathematica*.

The discovery of parts of mathematics not derivable from these axioms would raise the problem: are such parts incorrect, or are the axioms inadequate? In the latter case the axioms would have to be modified in order to accomodate also such parts: here lies the character of open system of *Principia Mathematica*. From this viewpoint Gödel's incompleteness theorem raised no problem: a formally undecidable Gödel sentence simply did not belong to ordinary mathematics.

On the other hand, while the existence of formally undecidable metasystematic sentences, like Gödel sentences, was irrelevant to the question of completeness of *Principia Mathematica*, the existence of formally undecidable *mathematical* sentences would be quite relevant to that question. Now, in the last decade several examples of mathematical sentences formally undecidable in elementary number theory have been discovered, such as Goodstein's theorem, the Paris-Harrington variant of the finite Ramsey's theorem and the Friedman-Takeuti miniature version of Kruskal's theorem (see e.g. Takeuti 1987: 120-147).

These examples show elementary number theory to be incomplete in an empirical sense. It has been objected that such sentences, while being mathematical, are not strictly number-theoretical, because there would be a sense to 'number-theoretical' different from expressibility in the language of elementary number theory, but the alleged alternative sense that has been proposed does not seem to be quite clear or convincing (cf. Isaacson 1987). In any case, these sentences would be no problem for Russell because they are provable in *Principia Mathematica*.

4. The formalist concept of formal system

I have already mentioned that, as regards the concept of formal system, Russell took a backward step with regard to Frege. The notion of formal system was again brought into the forefront by Hilbert in the Twenties, but with an essential twist: it

was no longer assumed that there is a universal system in the sense of Frege, Russell and Wittgenstein, and formal systems were no longer conceived as open systems. This was stated by Hilbert as follows:

«No more than any other science can mathematics be founded by logic alone; rather, as a condition for the use of logical inferences and the performance of logical operations, something must already be given to us in our faculty of representation, certain extralogical concrete objects that are intuitively present as immediate experience prior to all thought (...) All the propositions that constitute mathematics are converted into formulas, so that mathematics proper becomes an inventory of formulas (...) Wherever the axiomatic method is used it is incumbent upon us to prove the consistency of the axioms» (1967: 464-465, 472).

By stating that, no more than any other science, can mathematics be founded by logic alone - and hence denying the system of *Principia Mathematica* the status of the all-embracing logic - Hilbert claims that no universal system is possible. By stating that, as a condition for the use of logical inferences and the performance of logical operations, something must already be given to us in our faculty of representation, he denies that the inference rules are void of any intuitive logic. By stating that all propositions that constitute mathematics are converted into formulas, Hilbert claims that the content of every mathematical theory can be given by its representation in a formal system. Finally, by considering the task of giving a consistency proof, he appears to have a metasystematic standpoint.

While for Frege and Russell the rules of inference were the rules of conduct for the study of mathematics, for Hilbert they became the object of study. Thus he was led to a proof theory which deals with the proofs themselves. As in the concrete intuitive number theory the numbers are something which is objective and can be exhibited, in proof theory «it is just proof that is something which is concrete and can be exhibited; contentual arguments are only about proofs» (Hilbert 1922: 169-170).

Making proofs the object of study was Hilbert's way of implementing his metasystematic standpoint.

5. The formalist view of logic as a closed system

Hilbert's approach entails that formal systems are no longer conceived as open systems. They are formula games carried out according to certain definite rules which «form a closed system that can be discovered and definitively stated» (1967: 475).

Hilbert's influence was such that the concept of formal system became synonymous with closed system. Thus Hertz uses 'closed system of propositions' instead of 'formal system' (cf. Hertz 1929). Zermelo uses 'logically closed system' in the same sense (cf. Zermelo 1929). Tarski follows on: «Every set of sentences which contains all its consequences is called a deductive system, or possibly a closed system, or simply a system» (1956: 69-70).

It is true that Hilbert admits that formal systems are constantly extended: «The axiom system is constantly extended and the formal construction, corresponding to our constructive tendency, becomes more and more complete» (1922: 169). In particular, formal systems are constantly extended by adding new concepts. However Bernays, in his 1930 paper on Hilbert's program, stresses that an extension of a formal system by new concepts may be conservative, i.e. it may not lead to new results in the original language of the system.

This condition is satisfied when the formal system is deductively closed, i.e. complete in the syntactical sense. Now, «as regards number theory, as determined by Peano's axioms with the addition of recursive definitions, we believe that it is, in such sense, deductively closed» (Bernays 1976: 59). This shows that Bernays considered the possibility of constantly extending a formal system to be perfectly compatible with its deductive closure and with its character of closed system.

Indeed he had to admit such a compatibility if he wanted to preserve Hilbert's view of formal systems as closed systems. Such a view was implicit in the fact that Hilbert had a

metasystematic standpoint: asking metasystematic questions about a formal system S involves referring to the collection of theorems of S as to a given totality, and hence considering it as a closed system.

Gödel's incompleteness theorem shows that the concept of formal system as a closed system is inadequate for mathematics. Indeed it establishes that each formal system for number-theoretical truth must admit proper extensions, and hence the choice of any particular formal system for number-theoretical truth would be intrinsically provisional, subject to an eventual need to go beyond it. Therefore one may conclude that the concept of formal system as a closed system is inadequate for number-theoretical truth and hence is inadequate for mathematics.

As anticipated by the remarks of Bernays already mentioned, this conclusion seems to be in conflict with an authoritative interpretation of Gödel's incompleteness theorem: the one provided by Bernays in his 1930 paper on Hilbert's program and taken over by Hilbert and Bernays in their joint work *Grundlagen der Mathematik*.

6. The Hilbert-Bernays interpretation of Gödel's incompleteness theorem

In the second volume of the *Grundlagen der Mathematik* Hilbert and Bernays state that, by Gödel's result,

«the idea of characterizing the whole of mathematics as a deductive formalism, as sometimes suggested by the logistic systems, appears inadequate (...) We have avoided introducing the idea of a total system for mathematics (...) Instead we have contented ourselves with characterizing the actually existing system of analysis and set theory as providing an adequate framework for accomodating the geometrical and physical disciplines. A formalism may correspond to this aim even without having the property of full deductive closure. Our conception (...) very well agrees with the deductive openness of this system: the inference modes in the system are oriented in accordance with the representation of a closed, totally determined reality and give formal expression to this representation; but from this it does not

follow that the deductive structure (...) resulting from these inference modes should have that property of total closure» (1970: 289-290).

Such an interpretation seems inadequate in at least two respects. First, Hilbert and Bernays claim that Gödel's incompleteness theorem refutes the logicist view that logic is the system and that anything that can be said must be said within the system. Now, as we have already seen, the logicist view has as corollary that no metasystematic question can be raised and that the only question of completeness that can be raised is completeness in an empirical sense: hence the logicist view is unaffected by Gödel's result.

Secondly, while reasserting that formal systems are closed systems, Hilbert and Bernays claim that their incompleteness is compatible with their adequacy for mathematical practice, at least relative to the part of mathematics used in geometrical and physical applications. This amounts to saying that such systems, while incomplete in the metasystematic sense, may be complete in an empirical sense, which clashes with Hilbert's aim «to eliminate once and for all the questions regarding the foundations of mathematics» (1967: 489).

In order to achieve such an aim no empirical approach seems to be possible: how could one make sure that the next theorem discovered in mathematical practice would be provable in the formal system? In order to obtain a definitive solution, nothing less than a reduction of all mathematical reasoning to the mathematically secure finitary reasoning would be required.

Hilbert felt that such a reduction could be obtained, thus definitively eliminating the foundational problem: indeed, in 1930 he even felt that he had already achieved such an aim:

«I believe that, through my proof theory, I have completely achieved what I wanted and had promised: the problem of foundations of mathematics, as such, is thereby definitively eliminated» (Hilbert 1931: 494).

But, in order to obtain such a definitive solution, finitary reasoning must involve only concrete-intuitive objects, i.e. objects such that they can be surveyed

«completely in all their parts, and the fact that they occur, that they differ from one another, and that they follow each other, or are concatenated, is immediately given intuitively, together with the objects» (Hilbert 1967: 464-465).

Thus, finitary reasoning must not involve abstract objects. Depending only on concrete-intuitive objects, Hilbert's definitive solution to the problem of foundations of mathematics would achieve that separation of mathematics from philosophy which would allow him to claim that «mathematics is a presuppositionless science» (Hilbert 1967: 479).

The currently prevailing interpretation of the Gödel-Hilbert-Bernays theorem on unprovability of consistency is that, while all finitary proofs can be expressed in the formalism of elementary number theory, the proof of its consistency cannot be carried out within that formalism and hence cannot be finitary. However, in his 1931 paper, speaking about the system P for which he established his theorem, Gödel suggested that «it is conceivable that there exist finitary proofs that cannot be expressed in the formalism of P » (1986: 195).

Such conclusion was considered to be somewhat paradoxical by Hilbert and Bernays who stated that Gödel's incompleteness theorem entails that,

«in case of success of a finitary consistency proof for the formalism of analysis and set theory, at the same time also a finitary proof of a theorem of recursive number theory would be made possible which is not derivable in that formalism. It appears paradoxical that the methods of finitary proof theory should, under a certain respect, be superior to those of analysis and set theory in proving number-theoretical propositions. We are thus led to the question of the range of finitary methods» (1970: 290).

No convincing solution to this question was provided by Hilbert and Bernays. On this subject Bernays' position seems to be wavering. In 1935 he acknowledged that by Gödel's result one cannot prove by elementary combinatorial methods the consistency of a formal system in which every elementary combinatorial proof of an arithmetical proposition can be

represented. An example of such a system is provided by elementary number theory. Indeed,

«no attempt made up to now has given us any example of an elementary combinatorial proof which cannot be expressed in this formalism, and the methods by which one can, in the cases considered, translate a proof into the aforementioned formalism, seem to suffice in general» (Bernays 1964: 285).

Here Bernays seems to be rejecting the possibility, suggested by Gödel 1931, that it is conceivable that there exist finitary proofs which cannot be expressed in the formalism of elementary number theory. Indeed he recognizes that «means more powerful than elementary combinatorial methods are necessary to prove the consistency of the axiomatic theory of numbers» (Bernays 1964: 285). In his view such a more powerful method is provided by the Gödel-Gentzen result that the consistency of intuitionistic number theory entails the consistency of classical number theory. In order to conclude from this result that classical number theory is consistent, it suffices to assume the consistency of intuitionistic number theory. This proof of the consistency of classical number theory shows us that «intuitionism, by its abstract arguments, goes essentially beyond elementary combinatorial methods» (1964: 285-286).

Here Bernays acknowledges that, in order to prove the consistency of elementary number theory, abstract methods are required and seems to be willing to accept the intuitionistic methods to this effect. However, in the second volume of the *Grundlagen der Mathematik*, Hilbert and Bernays reject the proof of the consistency of elementary number theory via translations into intuitionistic number theory as totally removed from Hilbert's methodological views on proof theory (cf. Hilbert and Bernays 1970: 372). On the other hand, in the Introduction to that volume Bernays stresses the necessity to extend the previous delimitation of the 'finitary standpoint' and claims that «the discussion of the extension of the finitary standpoint leads to consider Gentzen's new consistency proof for the number-theoretical formalism» (1970: vii). Here, then, Bernays seems to

be willing to extend the finitary standpoint so as to include Gentzen's proof.

In no sense, however, Gentzen's proof can be considered to be finitary, if finitary reasoning must involve no abstract objects. As stressed by Gödel in his 1958 *Dialectica* paper, the validity of recursion up to the first epsilon number cannot be made immediately intuitive because it is impossible to visualize the many different ways in which descending sequences may be structured, and hence it is impossible to intuitively know that each such sequence must break off. In particular, «if one comes to know it by moving step by step from lower to higher ordinals, this will not constitute knowing it by *inspection*. It will merely be an abstract knowledge» (Gödel 1980: 134).

The fact that Gentzen's proof involves essentially abstract knowledge entails that it cannot be considered as finitary. Indeed, this opinion seems to be shared by Gödel in his *Dialectica* paper, where he implicitly rejects his previous 1931 suggestion that it is conceivable that there exist finitary proofs which cannot be expressed in the formalism of elementary number theory. Indeed, apparently approvingly, he mentions Bernays' 1935 view that, in order to prove the consistency of the number-theoretical formalism, it is necessary to extend Hilbert's finitary standpoint by admitting certain abstract concepts in addition to the concrete-intuitive ones. With the only difference that, while the abstract concepts considered by Bernays 1935 in order to go beyond finitary reasoning were those of intuitionism, Gödel instead intended to use the concept of primitive recursive functional of finite type.

Both Gödel's 1931 view and *Dialectica* view have a devastating impact on Hilbert's program. The 1931 view raises the problem that, for any formal system S which is supposed to be adequate for elementary number theory, there must be a finitary proof in the extended sense which cannot be carried out in S . Thus the finitary methods in the extended sense cannot be fixed once and for all, but must be suitably chosen for any specific formal system S considered, i.e. they are dependent on S . This contradicts Hilbert's assumption that, in order to obtain a

definitive solution to the question of foundations of mathematics, the finitary methods must be absolute, not relative to S.

On the other hand, the *Dialectica* view raises the problem that the question of the consistency of any specific S which is supposed to be adequate for elementary number theory - and hence adequate for finitary reasoning - cannot be settled conclusively, because the consistency of S can be expressed in the language of S but cannot be decided by S. Hence the choice of S becomes problematic because, by Gödel's incompleteness theorem, the completeness criterion is not available in all the essential cases. This contradicts Hilbert's aim to obtain a definitive solution to the question of foundations of mathematics.

Quite rightly, Hilbert and Bernays considered the possibility suggested by Gödel 1931 - that there exist finitary proofs which cannot be expressed in the formalism of elementary number theory - to be somewhat paradoxical. An obvious alternative is to interpret Gödel's incompleteness theorem as providing a conclusive refutation of Hilbert's program and of the underlying assumption that formal systems must be closed systems. By denying this natural conclusion, Hilbert and Bernays were lead to consider the former paradoxical possibility and were unable to produce a sense of 'finitary' which, while substantiating that possibility, would be compatible with Hilbert's aim of providing a definitive solution to the problem of foundations of mathematics.

7. Logic programming and open systems

The proper interpretation of Gödel's incompleteness theorem seems to be that the choice of any particular formal system for number-theoretical truth is intrinsically provisional, subject to an eventual need to go beyond it: hence the concept of formal system as a closed system is inadequate for number-theoretical truth and must be replaced by that of formal system as an open system.

This conclusion, however natural, did not find its way easily. For a long time people preferred to preserve the concept of

formal system as a closed system and tried to cope with the situation raised by Gödel's result by pushing the openness requirement outside the formal system - using Turing's ordinal logics or Feferman's recursive progressions of axiom systems. This, however, was no escape from incompleteness because, for any recursive progression satisfying certain natural conditions, there exists a true number-theoretical sentence which is not provable in the progression (see Feferman and Spector 1962).

The notion of formal system as an open system has been brought again into the forefront only recently, thanks to the emergence of a new paradigm of logic, alternative to mathematical logic: *computational logic*. The origin of such a paradigm can be dated back to the early Seventies, when a programming language of an entirely new kind was developed: *Prolog*. Although it was not realized at the time, an accessory and somewhat marginal feature of Prolog introduced an essentially new concept of proof.

According to the concept of proof underlying Hilbert's proof theory, proofs start from given axioms and lead to theorems by given rules of inference: axioms and rules of inference are fixed beforehand and do not change in the course of proof. The concept of proof introduced in Prolog deviates from Hilbert's in one essential respect: thanks to predicates for manipulating the data base, such as **assert** and **retract**, axioms and rules of inference may change in the course of proof. Such predicates modify the program, so relations holding at a certain time may no longer hold at another time, hence the same question can have different answers at different times.

This capability of Prolog of managing the data base dynamically, is essential for simulating the real world and is one of the main features which distinguish Prolog from the formal systems of mathematical logic. Prolog makes fully explicit what was only implicit in the logicist concept of formal system: the possibility of modifying the system whenever required. While in the logicist concept of formal system this possibility was allowed but was, so to say, accidental and exceptional, in Prolog it becomes part of the definition of formal system itself.

As already mentioned, the fact that computational logic represents a change of paradigm with respect to Hilbert's concept of proof was not realized, however, from the very beginning. The mathematical logic paradigm was so deeply rooted that the innovative changes of Prolog have long been mistaken for its defects. E.g., according to what it is claimed in the first and by now classical book on Prolog:

«These operations violate the simple self-contained nature of Predicate Calculus propositions (...) the use of **assert** means that the rule is talking about adding something to the set of axioms. In logic, each fact or rule states an independent truth, independent of what other facts and rules there may be. Here we have a rule that violates that principle. Also, if we use this rule, we will be in a position of having a different set of axioms at different times of the proof! (...) The ultimate goal of a logic programming language has not, then, been achieved with Prolog» (Clocksin and Mellish 1987: 241-242).

Here, what is an important theoretical innovation of Prolog is rightly seen as a deviation from the mathematical logic paradigm, but such a deviation is considered dangerous and to be amended as soon as possible to comply with that paradigm. On the contrary, the proper interpretation of the situation seems to be: almost accidentally and unintentionally, Prolog ran into a concept of proof which is essentially innovative with respect to the concepts of proof of Euclid and Hilbert.

8. Concurrent object-logic programming and open systems

Introducing a new concept of proof, however, was not a main aim of Prolog, but only a byproduct of an accidental and somewhat marginal feature of the language: the inclusion of predicates for manipulating the data base such as **assert** and **retract**. On the other hand, this aim became central in various attempts to combine logic, object-oriented and concurrent programming which have been made in the last decade. In such attempts two distinct approaches can be identified.

The first one is exemplified by languages, like *Concurrent Prolog*, in which concurrency occurs at the micro-level of the single Horn clause (cf. Shapiro and Takeuchi 1987). Such languages, while designed for fast execution on parallel architectures, do not however provide a good syntax for expressing the abstractions of object-oriented programming. At best they may be considered as machine-oriented object-logic languages (cf. Kahn, Tribble, Miller and Bobrow 1987).

The second approach is exemplified by languages, like *logObjects*, in which concurrency occurs at the macro-level of logical systems (cf. Welsch and Barth 1989; see also Mello and Natali 1987). Such languages seem to provide a better integration between logic, object-oriented and concurrent programming.

In languages like *logObjects* a program is intended as a collection of logical objects. Each logical object consists 1) of a database which represents a chunk of knowledge about a particular domain and 2) of a theorem prover which is able to answer proof requests. Logical objects can communicate by sending messages: a message sent to a particular logical object may be conceived either as a request to modify the database or as a request to prove a goal using the given database. Both kinds of requests trigger a proof search procedure. However, while the answer to a request to modify the database consists either in an acceptance or in a refusal, the answer to a request to prove a goal consists either in the proof of the goal or in the notification of a failure. Thanks to the possibility of making requests to other logical objects, each logical object can be modeled as an open world.

At least potentially, this results into a concept of proof which is even more innovative than that provided by *Prolog*. In order to show such an innovative character let us consider two examples.

Our first example is given by the flight reservation system, which consists of several concurrent objects, e.g. flights and agents. Flights will register the reservation of seats and ensure that no seats are reserved by two agents at the same time. Both flights and agents may be constructed as logical objects. Now, suppose that two agents are concurrently attempting to reserve seats in the same flight but the number of seats is insufficient to

satisfy both requests. The operation will be serialized so that one of the requests will be accepted and the other will be rejected. However, it is indeterminate which of them will be accepted and which will be rejected. The outcome cannot be deductively decided even from complete knowledge of all the circumstances at the time when the two agents send their respective requests. In order to manage conflicts between requests, a negotiation is needed to establish an order of acceptance.

This example illustrates a situation which is typical of concurrent systems in general: they have the property that often the behavior of the system is critically affected by the arrival order of communications. Generally, the decisions on the arrival order in concurrent systems are not deductively derivable, hence logical deduction does not provide an adequate basis for decision making in computations in concurrent systems (see Hewitt 1991).

Our second example is provided by setting the price of an industrial product. The following two principles suggest themselves: 1) greater profitability is better than lower profitability; 2) greater market share is better than lower market share. Such two principles, however, are incompatible: increasing the price generally increases profitability but decreases market share. Two formal systems including 1) and 2), respectively, among their axioms would be inconsistent. Logical deduction provides no mechanism for resolving this conflict: what is needed is a negotiation yielding restrictions on 1) and 2), but such a negotiation cannot be based on logical principles only.

This suggests the need for a concept of formal system far removed from Hilbert's. The correctness of proofs in Hilbert's formal systems can be algorithmically checked without having to make any observations on the external world or to consult any external sources of information. The algorithm is fixed in advance and no other computation is needed other than to read the proof and apply the algorithm. Thus proof checking proceeds in a closed world in which the axioms and the rules of inference have been laid out explicitly beforehand.

Hilbert's concept of formal system has the advantage that proofs in the system have a timeless and acontextual character

and can be checked only from the text of the proof. This allows them to be checked by many people at different times and places. However, just because of their timeless and acontextual character, such proofs do not provide an adequate basis for decision making in concurrent systems because decision making is usually crucially time dependent.

For example, in the flight reservation system mentioned earlier, the capability to process conflicts in which multiple agents concurrently attempt to reserve seats in the same flight goes beyond the capabilities of logical deduction. Of course a record can be made of what happened, but this record is not the same as a proof: proofs are produced in advance of any conflict and their correctness can be algorithmically checked.

Whenever conflicts arise, logical deduction does not allow to infer the outcome, because the processes which actually produce the outcome are left out. Decisions in concurrent systems may be justified by a negotiation: such is the case both in the flight reservation system and in setting the price of an industrial product. But justifying a decision by negotiation is essentially different from justifying it by logical proof.

Logical deduction is concerned only with the internal structure of Hilbert's formal systems. As we have already mentioned, its great strength stems from the fact that proofs in such systems have a timeless and acontextual character and can be checked only from the text of the proof. But, by concentrating only on the internal structure of formal systems, logical deduction leaves out something which plays a fundamental role in concurrent systems: communication. Thus the very features of formal systems that give them their great strength turn out to also be their greatest weakness.

The development of a theory of communication has already started (see e.g. Hoare 1985; Milner 1989). In the theory of communication one deals with conflicts, cooperation, negotiation, i.e. concepts which are essential for modeling the real world, as shown by the above examples of the flight reservation system or setting the price of an industrial product (cf. Hewitt 1991). There is, moreover, the problem of integrating

logical deduction with communication, which requires further elaboration.

An adequate concept of formal system must include, in addition to logical deduction, also communication: logical deduction is an important component of the concept of formal system but is only a component and has inherent limitations in concurrent systems (cf. Hewitt 1987).

9. Open systems: past and present

In the *Republic*, Plato makes a fierce attack against contemporary geometers. What is the reason for that attack, given that in his previous works, up to and including *Phaedo*, geometry was a research model for him? Most likely, the reason is to be found in the circumstance that, between *Phaedo* and the *Republic*, geometry underwent a radical transformation: the axiomatic revolution.

Geometry, as described in Plato's works before the *Republic*, especially in *Meno* and *Phaedo*, is an open system, based on the analytic method in the form already used by Hippocrates of Chios in his work on the quadrature of lunes. For Plato the axioms of geometry are not the starting point of the proof: the starting point is the proposition to be proved. The aim of the proof is to reduce such a proposition to another one already proved.

While the axiomatic method assumes that axioms are fixed beforehand and hence the system is a closed system, the analytic method does not make such an assumption. Proofs in the axiomatic method establish a global ordering between propositions. On the contrary, the aim of proofs in the analytic method is to establish a local ordering between propositions through a backwards search. The ordering is local because it is only relative to the proposition to be proved: different propositions to be proved may require different orderings. The axioms are not fixed beforehand. They are unknown at the beginning of the backwards search and are found only at the end of the process. It may very well turn out that, in order to prove a

given proposition, new axioms must be introduced: thus geometry is an open system (cf. Cambiano 1967, 1991).

Notwithstanding Plato's fierce opposition, the view of geometry as a closed axiomatic system prevailed and was elaborated by Aristotle in his *Posterior Analytics*. Apparently this depended on a misjudgement: Aristotle believed that most sciences, including geometry, were complete or nearly complete, hence the question was not how to acquire new knowledge but how best to arrange and impart this vast and perfect body of knowledge (cf. Barnes 1975). In any case, as a result of the axiomatic viewpoint prevailing, the analytic method of Hippocrates of Chios and Plato was abandoned and replaced by a different form of the method, subordinate to the axiomatic method, as described by Pappus. On his own part, after the *Republic*, starting from *Phaedrus*, Plato replaced geometry by medicine as his research model (see Cambiano 1967).

An attempt to revive the analytic method in its original form, and hence the concept of open system, was made by Galileo, Descartes, Newton at the beginning of the modern age (see e.g. Engfer 1982). But in the Nineteenth century, as a result of the non-euclidean revolution, the view of geometry as a closed axiomatic system again prevailed.

Ultimately, by his celebrated result, Gödel showed that the concept of formal system as a closed system is inadequate for mathematics. But the logical community failed to learn Gödel's lesson - that Hilbert's concept of formal system is inadequate - and continued to use it as if there were no incompleteness theorem. Contemporary work on computational logic tries to reverse this trend, by developing a new concept of formal system as an open system, capable of handling information in continuous change and evolution. One may hope that this work will succeed where Gödel failed, convincing people to replace Hilbert's concept of formal system, still used in mathematical logic, by a richer concept more adequate for applications to the real world.

Contrary to what was suggested by Turing's ordinal logics or Feferman's recursive progressions of axiom systems, there is no global axiomatic theory of the real world that gradually becomes

more and more complete. More modestly, each problematical situation is dealt with by establishing ways of negotiation between the available, usually conflicting, theories concerning that situation. This seems to be the only possible approach to the real world.

*Dipartimento di Studi Filosofici ed epistemologici
Università di Roma «La Sapienza»
Via Nomentana 118 - 00161 Roma*

Bibliographical references

- BARNES, J. (1975). Aristotle's theory of demonstration. In: J. Barnes, M. Schofield & R. Sorabji (eds.), *Articles on Aristotle: 1. Science*. London: Duckworth, 65-87.
- BERNAYS, P. (1964). On platonism in mathematics. In: P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics: Selected Readings*. Englewood Cliffs, N.J.: Prentice-Hall, 274-286.
- BERNAYS, P. (1976). Die Philosophie der Mathematik und die Hilbertsche Beweistheorie. In: P. Bernays, *Abhandlungen zur Philosophie der Mathematik*. Darmstadt: Wissenschaftliche Buchgesellschaft, 17-61.
- CAMBIANO, G. (1967). Il metodo ipotetico e le origini della sistemazione euclidea della geometria. *Rivista di filosofia* 58, 115-149.
- CAMBIANO, G. (1991). *Platone e le tecniche*. Roma: Laterza.
- CELLUCCI, C. (1987). Il logicismo di Russell. In: R. Simili (ed.), *L'epistemologia di Cambridge 1850-1950*. Bologna: il Mulino, 305-328.
- CELLUCCI, C. (1990). La logica e la rappresentazione delle conoscenze. *Rivista di filosofia*, 81, 19-55.
- CELLUCCI, C. (1992). Dalla logica teoretica alla logica pratica. *Rivista di filosofia*, forthcoming.

- CLOCKSIN, W.F. & MELLISH, C.S. (1987). *Programming in Prolog*. Berlin: Springer-Verlag.
- DUMMETT, M. (1981). *Frege: Philosophy of language*. London: Duckworth.
- ENGFER, H.J. (1982). *Philosophie als Analysis*. Stuttgart-Bad Cannstatt: Frommann-Holzboog.
- FEFERMAN, S. & SPECTOR, C. (1962). Incompleteness along paths in progressions of theories. *The Journal of Symbolic Logic*, 27, 383-390.
- FREGE, G. (1967). *Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought*. In: J. van Heijenoort, *From Frege to Gödel*. Cambridge, Mass.: Harvard University Press, 1-82.
- FREGE, G. (1969). On Herr Peano's Begriffsschrift and my own. *Australasian Journal of Philosophy* 47, 1-14.
- GÖDEL, K. (1964). Russell's mathematical logic. In: P. Benacerraf & H. Putnam (eds), *Philosophy of mathematics*. Oxford: Blackwell, 211-232.
- GÖDEL, K. (1980). On a hitherto unexploited extension of the finitary standpoint. *Journal of Philosophical Logic*, 39, 133-142.
- GÖDEL, K. (1986). On formally undecidable propositions of Principia Mathematica and related systems I. In: K. Gödel, *Collected works*, vol. I. Oxford: Oxford University Press, 144-195.
- GOLDFARB, W.D. (1979). Logic in the twenties: the nature of the quantifier. *The Journal of Symbolic Logic*, 344, 351-368.
- HERTZ, P. (1929). Über Axiomensysteme für beliebige Satzsysteme. *Mathematische Annalen*, 310, 457-514.
- HEWITT, C. (1987). Metacritique of McDermott and the logicist approach. *Computational Intelligence*, 3, 185-189.
- HEWITT, C. (1991). Open information systems semantics for distributed artificial intelligence. *Artificial Intelligence*, 347, 79-106.
- HILBERT, D. (1922). Neubegründung der Mathematik (Erste Mitteilung). *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität* 1, 157-177.

- HILBERT, D. (1931). Die Grundlegung der elementaren Zahlentheorie. *Mathematische Annalen*, 104, 485-494.
- HILBERT, D. (1967). The foundations of mathematics. In: J. van Heijenoort (ed.), *From Frege to Gödel*. Cambridge, Mass.: The MIT Press, 464-479.
- HILBERT, D. & BERNAYS, P. (1970). *Grundlagen der Mathematik II*. Berlin: Springer-Verlag.
- HOARE, C.A.R. (1985). *Communicating Sequential Processes*. Englewood Cliffs: Prentice Hall International.
- ISAACSON, D. (1987). Arithmetical truth and hidden higher-order concepts. In: The Paris Logic Group (eds), *Logic Colloquium '85*. Amsterdam: North-Holland, 147-169.
- KAHN, K., TRIBBLE, E.D., MILLER, M.S. & BOBROW, D.G. (1987). Vulcan: Logical Concurrent Objects. In: E. Shapiro (ed.), *Concurrent Prolog*, vol. II. Cambridge, Mass.: The MIT Press, 274-303.
- MELLO, P. & NATALI, A. (1987). Objects as communicating prolog units. In: *ECOOP '87*, LNCS 276. Berlin: Springer-Verlag, 181-191.
- MILNER, P. (1989). *Communication and Concurrency*. Englewood Cliffs: Prentice Hall International.
- RUSSELL, B. (1937). *The Principles of Mathematics*. Cambridge: Cambridge University Press.
- SHAPIRO, E. & TAKEUCHI, A. (1987). Object oriented programming in concurrent prolog. In: E. Shapiro (ed.), *Concurrent Prolog*, vol. II. Cambridge, Mass.: The MIT Press, 251-273.
- SHEFFER, H. (1926). Review of A.N. Whitehead and B. Russell, *Principia Mathematica*. *Isis*, 8, 226-231.
- TAKEUTI, G. (1987). *Proof Theory*. Amsterdam: North-Holland.
- TARSKI, A. (1956). Fundamental concepts of the methodology of the deductive sciences. In: A. Tarski, *Logic, Semantics, Metamathematics*. Oxford: Clarendon Press, 60-109.
- VAN HEIJENOORT, J. (1967). Gödel's theorem. In: *The Encyclopedia of Philosophy*. New York: MacMillan.
- VAN HEIJENOORT, J. (1985). *Selected Essays*. Napoli: Bibliopolis.

- WELSCH, C. & BARTH, G. (1989). Reasoning objects with dynamic knowledge bases. In: *EPIA '89*, LNAI 390. Berlin: Springer-Verlag, 257-268.
- WHITEHEAD, A.N. & RUSSELL, B. (1927). *Principia Mathematica*, vol. I. Cambridge: Cambridge University Press.
- WITTGENSTEIN, L. (1961). *Tractatus Logico-philosophicus*. London: Routledge & Kegan Paul.
- ZERMELO, E. (1929). Über den Begriff der Definitheit in der Axiomatik. *Fundamenta Mathematicae*, 314, 339-344.

SOME PHILOSOPHICAL IMPLICATIONS OF GÖDEL'S THEOREM

Evandro Agazzi

Some internal dynamics of formalism

If we reflect on the complex history of formalism, we can see that the acceptance of a more or less substantial degree of formalism was a feature common to a lot of trends of modern mathematics. This explains why what is termed 'modern mathematics' (practically speaking the mathematics of our century) is so deeply characterized by 'abstract' theories, by 'formal' procedures and tools, by stress being laid on 'structures', and by an indifference towards 'intuition'. All this (which has been programmatically expressed and concretely realized e.g. in the famous *Eléments de mathématique* of the Bourbaki group) has even led to a strictly formal reconstruction - by means of abstract axiomatic systems analogous to those of algebra and topology - of traditionally constructive and intuitive mathematical theories, such as arithmetic and analysis. Moreover, the considerable success of formalization in mathematics strongly encouraged the extension of this methodology to many other fields. R. Carnap (1966), for instance, developed formalism as a critical methodology for a linguistic treatment of *philosophy*, while axiomatizations of *physical* theories were promoted especially regarding quantum mechanics. In this field not only were formal axiomatizations of various kinds soon elaborated at the view to coping with the non-'visualizability' of micro-phenomena, but the special logical calculi of many-valued logics were used in order to cope with some well known difficulties

related to the uncertainty principle¹. But one could also mention mathematical *psychology*, where different formal axiom systems have been proposed e.g. by P. Suppes (1969), for a precise analysis of human behaviour. Also the elaboration of the notion of preference in the field of *economic* and *social* activity has received this kind of treatment by J. von Neumann and O. Morgenstern (1944), not to mention the attempts to apply this methodology in the field of *juridical* science (see Klug 1951). But we are not interested in multiplying these examples: we want rather to explore the main motivations which encouraged this general acceptance of the 'formal way of thinking'.

Some of these motivations have something of a 'negative' flavour, in the sense that they derive from a kind of frustration in the attempt to attain truth by gaining command of the 'content' of intuitive knowledge, this content being at the same time both the configuration of a meaning and the relation to a referent. This unreliability of intuition, however, must be further analyzed; and if we take into account the fact that it emerged in the field of mathematics, it appears that the mistrust regards *intellectual intuition*, (possibly) leaving *sensory* intuition untouched. But if this is so, we immediately see that the origin of this phenomenon goes back to a level deeper than the history of non-Euclidean geometry. Indeed one should at least consider Kant, the transition of whose thought from the so-called pre-critical to the critical stage is characterized by the fact that in the *Critique of Pure Reason* (1781) intuition is confined to the senses and denied to the intellect, which was not the case at least until Kant's famous dissertation of 1770². But Kant himself was, in this sense, the terminus of a longer itinerary, the first part of which had been travelled by other philosophers (especially the British empiricists), and even more significantly by the founders of modern science. When Galilei was rejecting the grasping of the

1 This trend originated with Birkhoff/von Neumann (1936), and was taken up again e.g. by v. Weizsäcker (1955), Mittelstaedt (1968) and Scheibe (1964), and in more recent years has given rise to a rich branch of studies known as *quantum logic*.

2 *De mundi sensibilis atque intelligibilis forma et principiis dissertation*.

'essences' of natural substances as a «desperate enterprise»³, he was actually expressing a mistrust in the possibility of having an intellectual intuition. And when modern science, following his conceptual revolution, accepted that natural *laws* should constitute the immutable structure of physical reality, it attributed to these laws the character of being *forms* of the phenomena, in this way breaking the identification of 'form' and 'essence' which had governed philosophy from the time of the Greeks. This amounts to saying that 'formal knowledge' (though in a sense which was not at all identical with 'formalistic knowledge') was considered to be reliable and rigorous. In this sense, Kant's 'Copernican Revolution' and transcendental philosophy may be seen as a powerful generalization of knowledge as such, and even of all forms of the activity of reason - a generalization of that view which modern science had advocated for a more restricted field. Indeed, transcendental philosophy tries to determine the *a priori* formal conditions of every judgment, be it cognitive, moral, or aesthetic.

It is interesting to note that, when a crisis of certitude affected the *physical* sciences at the end of the 19th and beginning of the 20th century (a crisis which concerned the validity of the 'formal' concepts and laws of physics themselves), an issue was found again in an intensification of the formal point of view, i.e. in a formalization of the second level, which was very close to a 'formalistic' perspective. It would lead us too far to explore this development, but this means that we could just as well have used the case of the physical sciences (instead of that of mathematics) as an illustration of the rise of modern formalism. However this path would have been much less transparent and would have needed a much more elaborate and sophisticated analysis. Along it we would have arrived essentially at the same conclusion concerning the 'negative' motivations which promoted the pre-eminence of formalism: a mistrust with regard to both truth and the 'intensional' sense of 'meaning'.

³ This claim is made in the third letter to M. Welser on the sunspots. See Galilei (1929-1939), vol. V, pp. 187-188.

But there were also several 'positive' reasons for the success of formalism, which might be condensed into one fundamental feature: the satisfaction of the requirement of generality. Generality resides in form and not in content: on this point a remarkable continuity exists in the history of philosophy. So too when 'form' was equated with 'essence' was it stressed that generality (or even better 'universality') was connected with form, while individuality was the specific contribution of 'matter'. The whole history of philosophy might be reconstructed in the light of the different ways in which the complex relation between form and matter has been conceived (the distinction between form and 'content' being nothing but a particular instance of this dichotomy). We shall certainly not sketch such a reconstruction here, but satisfy ourselves with hinting only at a few salient aspects of this question.

The first deserving mention is probably the Aristotelian invention of *formal logic*, the aim of which was to find general rules capable of granting the correctness of inferential discourse, i.e. discourse in which the consequences 'necessarily follow' from the premisses. In spite of the fact that this inquiry was moved by the desire to grant the discourse permanent *truth*, it turned out that, in order for this permanence to be really granted with necessity, one was to disregard the 'accidental' fact of the propositions involved being true (and consequently disregard also their *meaning*). This means that, in the case of a correct inference, the truth of the conclusion can be stated not so much because of the truth of the premisses, but *by force* of the cogency of the link, a cogency which remains intact if the link is applied to false premisses, but which *necessarily* ensures the transmission of truth, when it is applied to true premisses. This is why the treatment of the *Prior Analytics* is not only 'formal' in this sense, but also largely symbolic. Here we have a double level of universality: formal logical rules are valid not only for whatever 'content' or meaning the propositions might have, but also for whatever their truth-value might be, which means that they also apply to discourses in which the truth of the premisses is uncertain. Of course, in the *Posterior Analytics* Aristotle fully recovers the requirements of a knowledge endowed with

uncontroversial truth, and proposes the structure of the 'classical' axiomatic method which we outlined above. Within this structure, the 'formal' tool only plays the role of an instrument, while an intellectual intuition is required to establish the absolute truth of the premisses (but then the point of view is no longer formal).

As a second prominent example we shall mention again the philosophy of Kant, in which the most central tenet is perhaps that no universality or necessity can be linked to what is empirical, so that we must inevitably look for something *a priori* whenever we want to understand or justify any universal judgment of our reason. However, it turns out that for Kant the term «empirical» is applied to any 'content' whatever, while *a priori* is equivalent to *formal*, so that *a priori* or transcendental *forms* are proposed for identifying all possible kinds of universals, from those of sensory knowledge (space and time) to those of intellectual knowledge (the categories), and the moral law (the categorical imperative).

A remarkable quality of the universality granted by *form* is that it has nothing to do with the 'equivocation' or 'ambiguity' which is the price that most common sense notions pay when they are applied to different referents. On the contrary, in spite of their being applicable to (or interpretable on) a potentially infinite number of single cases, all formal notions are strictly 'univocal', and if one looks for the reason for this *prima facie* astonishing property, one finds that it is because they are free from the ambiguities which are *implicit* in the meanings of many intuitive concepts. Therefore the first result of formalization, which is so decisive that it deserves being chosen as its defining characteristic, is total *explicitness*. It is this requirement that suggests, as a practical measure, that we leave meaning aside, since meaning may be charged with hidden, uncontrolled and implicit elements that could produce errors in the long run. A classical example is the *quaternio terminorum* in a syllogism, which occurs when a certain concept is taken with two different meanings in two different propositions, so that the terms really involved are four instead of three, thus making the conclusion incorrect. If two different symbols are applied to denote the two

different meanings (which corresponds to a process of formalization), the error becomes patent and may be avoided. This feature explains first of all why formalization turned out to be such a powerful tool in making rigorous all kinds of disciplines (mathematical or not). It did this by permitting the elimination of ambiguities via the analysis and separation of several components in the meanings of basic notions, thereby making explicit presuppositions which were tacit in the immediate intuitions of several traditional disciplines. (In this regard consider e.g. that it was only with Pasch (1882) that the ordering properties of the points on a line - which had remained hidden in the Euclidean no less than in the non-Euclidean axiomatizations of geometry - were made explicit and axiomatized, so that the consequences of admitting or not admitting them could also be investigated). Secondly, this also explains the reason for the multi-valency of the formalisms: their being uncommitted to any *particular* meaning (which is quite different from being devoid of any possible meaning) leaves them open to a plurality of *distinct* (and therefore unambiguous) interpretations.

We may have said enough to present some of the most important factors which have determined the high ranking of formal thinking in our time. Of course, several more might be investigated, such as the impressive power of *unification* with which it is endowed, but it would lead us too far if we should continue this inquiry.

The justification of formalism

We have seen that the predominance of the formalistic attitude in the field of foundational research (especially regarding mathematics) was a result of a multitude of factors (some of a 'negative' and some of a 'positive' character), none of which in itself demanded that mathematical 'content' be fully discarded, but rather that it be kept in the background in order not to interfere with the obtaining of rigour. In fact, the working mathematician is used to availing himself of a special kind of intuition (he 'sees - so to speak - with the eyes of his mind'

numbers, functions, structures, sets, large cardinals, abstract spaces and their properties), and has therefore a natural tendency to consider them as 'existing' at least in some vaguely determinable sense. This impression is far from being a simple naive belief, and could be rigorously analyzed and justified by resorting to skilful distinctions which have been proposed in the history of philosophy, from the Medieval doctrines of the *ens rationis*, to the investigations of Meinong⁴ in our century. This feeling is so spontaneous and deeply rooted that one really has to be 'forced' to abandon or even reject it. Now the reason which many mathematicians felt to be 'forcing' them to take this step was the explosion of contradictions. Before that moment, formalization and symbolization (including the use of symbolic logic) were mainly conceived as *instrumental*, but not as *indispensable* for foundational research. This means that formal and symbolic axiomatizations were primarily intended either to provide extremely powerful tools for the logical *analysis* of concepts, assumptions and proofs occurring in mathematical theories; or to show how the content of a certain mathematical theory could be expressed by means of quite different axiomatizations (i.e. of axiomatic systems based on different sets of 'primitive notions'); or finally to reduce to the smallest possible number the primitive notions and assumptions necessary for reconstructing the content of a certain theory.

This was the chief interest of G. Peano and his school, to which we owe the famous conception of mathematics as a collection of *hypothetico-deductive* systems, as well as the idea

4 It would be out of place to go into details here concerning Meinong's doctrine of the objects of thinking. Let us simply recall that, according to him, there are different kinds of *objects*, which are characterized by the different mental acts through which we perceive them. So the objects of sense perception are different from those of thinking, but the latter are no less 'objective' than the former: they are 'apprehended' through thinking, but not constituted by it. Meanings and judgements are examples of this second kind of object: according to his terminology, they 'subsist' (*bestehen*), while individual things and qualities 'exist'. In this sense, objects of thinking may be *real* without 'existing' in the technical sense given by Meinong, and mathematical objects are such. In other words, an intellectual investigation is not bound to the empirical existence of its objects, no *existential presupposition* is needed in order for its objects to be given and 'real' in an exactly specifiable sense. The object *as such* is indifferent to the character of existence. *Theory of the Object* (see Meinong 1904) is probably the work where the most central claims of his doctrine are to be found.

that axioms provide 'implicit definitions' of their primitive concepts, and the first characterization of this kind of definition⁵. In this context the chief metatheoretical problem taken into consideration and widely investigated was that of the *independence* of concepts and axioms, while those of *consistency* and *completeness* were hardly perceived. But this was true even of Hilbert, whose famous book *The Foundations of Geometry*, first published in 1899 (ten years after the first writings of Peano in this field), is formal without being 'formalistic' (no symbolization is used and not even the logic employed is made explicit), and pays great attention to problems of independence, while consistency is only marginally treated, and completeness is not considered at all. Moreover, he says of his axioms that they «express certain fundamental homogeneous facts of our intuition», while stressing that 'the exact and complete description of these relations for the purposes of mathematics' follows from the axioms and nothing else⁶. But this is no wonder: at that time he was still quite Kantian, and in the introduction of his book he writes that «the problem in question (i.e. of setting up axioms for geometry and investigating the way in which they are connected) amounts to logical analysis of our intuition of space», while this intention was made clear again by his placing at the head of the introduction, as a motto for the whole book, these famous words from the *Critique of Pure Reason*: «Thus all human knowledge begins with intuitions, goes from there to concepts, and ends with ideas». It was only at a later stage that he fully concentrated upon the formalisms which are used in the formulation of mathematics, but even at that time he certainly did not intend to *reduce* the essence of mathematics to formalisms, as could easily be shown⁷. But why was Hilbert

5 Besides several articles devoted by Peano to the problem of definition, one should mention at least the contributions of Pieri (1899) and Padoa (1900).

6 See the first section of the first chapter of Hilbert (1899).

7 Hilbert certainly considered formalism to be an indispensable passage through which the consistency of even the most audacious mathematical theories was hopefully to be proved. But the 'content' of these theories was by no means trivialized, not even the «Paradise (of the infinite) which Cantor has created and from which nobody will ever chase us» (see Hilbert 1926) At the beginning of *The Foundations of Mathematics* (see Hilbert-Bernays 1934-39) an important distinction is drawn between 'concrete' (*inhaltlich*) and 'formal' axiomatics, where the former are by no means considered

(and not only Hilbert) led to concentrate almost entirely on formalisms? The reason is that no other way seemed to be suitable for establishing the consistency of mathematics. Indeed, the problem of consistency had already emerged in the discussions concerning non-Euclidean geometries, and was solved by an *indirect* proof, i.e. by constructing 'Euclidean models' of these geometries. This move contained several implicit requirements of which the geometers who constructed these models (Beltrami, Cayley, Klein, Poincaré) were unaware, and which became clear only in the much more sophisticated metamathematics of our century. Yet their basic intuition was correct: if we are able to *interpret* the axioms of a non-Euclidean geometry so that they become true of some structure of geometrical entities which, in spite of being 'artificial' still obeys the axioms of Euclidean geometry, then we can say that, in case a contradiction should follow from the axioms of the non-Euclidean geometry, this would be reflected in the model, and therefore it would follow also from the Euclidean axioms. Hence, *if* Euclidean geometry is consistent, non-Euclidean geometry must also be so. The obvious question is then: is Euclidean geometry consistent? One could adopt the same strategy and construct a model of this geometry in real number analysis (by using the familiar tools of analytic geometry), but in this way the problem would simply be shifted, as it would be again if one should go on and 'reduce' it to the consistency, say, of elementary arithmetic or of set theory (i.e. of two theories which have been indicated as possible grounds for the 'construction' of all number systems, and therefore of the whole of mathematics). In short, one must in the end be able to prove the consistency of one of the basic mathematical theories *directly*. But which one, and in what way?

Traditional mathematics did not know of and *could not* know of this problem, for it was considered to be based upon *immediately true* axioms. Since the use of correct inference necessarily implies that correctly deduced consequences share the

spurious. Finally it is certainly not insignificant that in the very years when he was working on his formalistic programme, Hilbert published a book on «intuitive geometry» (see Hilbert/Cohn-Vossen 1932).

truth of their premisses, it follows that the consequences of such axioms must be necessarily true. Therefore a contradiction, which is always *false*, can never be deduced from true axioms, and the problem does not even exist. But modern mathematics, as we have seen, is characterized by a distrust of the possibility of affirming the *immediate truth* of any axiom system, so that the traditional path was barred. On the other hand, by considering axioms as purely formal, i.e. neither true nor false, it was even more fundamentally impossible to prove their consistency by resorting to the notion of truth and truth consequences. This was the reason why the only hope seemed to many to reside in the possibility of taking a given axiom system - conceived as a combination of pure symbols devoid of meaning and truth and submitted only to material transformation rules - and trying to show its *formal* consistency, i.e. that a contradiction could not be *proved* in it.

Hilbert's programme

This displacement of the conceptual center of gravity was drastic, since it meant that the warranty of consistency was no longer to be looked for by exploring the nature of the axioms, but the nature and structure of the mathematical proofs. This move was explicitly taken by Hilbert when he characterized as *Beweistheorie* (or *proof theory*) his new approach to the problem of consistency. Since a proof is formal, and moreover since it was now intended to apply to purely formal systems of axioms, a complete formalization was required: this means that not only the specifically mathematical axioms, but also the logical calculus used for producing the proofs, be explicitly and symbolically given. Symbolism becomes now not only useful, but strictly necessary: if our task is that of investigating proofs, they must be susceptible of inspection, and therefore be given in a materially testable way, which only symbolization can afford. Another obstacle may also be avoided: a consistency proof for an axiomatic system had traditionally been believed to be impossible because it seemed to entail an infinite task (our having deduced a

hundred or a thousand theorems from the axioms without finding any contradiction does not ensure that we shall never find one). But a *formal* law already known to Medieval logicians is that from a contradictory set of premisses *whatever* conclusion may be correctly deduced, and this law is preserved in standard logical calculi. This means that a consistency proof may reduce to showing that even *one single* arbitrarily selected (and particularly simple) formula, such as $0=1$, cannot be formally derived in our axiom system. Now, the axioms are *finite* in number (and each of them consists of a *finite* sequence of symbols), and the transformation rules of the logical calculus used in the formalization (which permit the construction of proofs) are *also finite*. Thus it seems reasonable to hope that by manipulating the axioms according to ways of reasoning resembling those of the combinatorial calculus (which are confined to the consideration of the finite grouping and permutations of symbols) one might be able to show that that particular formula could never be derived. This non-hypothetical reasoning, which is halfway between the material manipulation and the intuitive perception of graphic symbols, was called by Hilbert the *finitistic* method, or «finitary inference» (*finites Schliessen*), and it was prescribed as the method to follow to obtain a direct proof of the consistency of formalisms. We have thus found the answer to the question 'how' to provide this proof.

The second question concerns which axiomatic system should be submitted to this kind of examination. In principle any system whatever could be chosen, but considering that the aim was that of providing a 'foundation' for mathematics, and that the investigations performed in the 19th century had already shown many relationships among the various mathematical theories, it was reasonable to choose the simplest of the known systems. Thus *elementary arithmetic* was considered, with the hope of being able to prove its consistency directly, and then extending this consistency to the other basic axiomatic systems of mathematics (essentially to analysis and set theory). This was in short the celebrated 'Hilbert Programme' which was first announced in the early years of the century, took a precise and

very explicit shape around 1920, and was thereafter developed by Hilbert in cooperation with several disciples and collaborators⁸.

Gödel's theorem

In spite of its reasonableness, Hilbert's programme failed to meet with the success expected: after some partial positive results, which concerned certain 'weakened' formal systems of elementary arithmetic, in 1931 K. Gödel (1931) proved a theorem which is deservedly considered one of the most outstanding results in the whole of 20th century science, and whose impact on the philosophical discussion was probably equalled only by Heisenberg's uncertainty principle in quantum mechanics. This theorem expresses the intrinsic impossibility of realizing Hilbert's programme in its original form. Indeed, Gödel proved that, if a formal system is *assumed* to be consistent (for if it is inconsistent anything may be proved in it), and is powerful enough to formalize elementary arithmetic, then it is impossible to prove its consistency by means of tools which are formalizable in the system itself. In a shorter form it may be said that any formal system satisfying the very minimal requirements just mentioned is unable to prove its own consistency 'internally'. Now, since the 'finitistic methods' admitted by Hilbert's programme are certainly formalizable in any axiomatic system for elementary arithmetic (and *a fortiori* in any more powerful mathematical system whatever), it was clear that Hilbert's programme had been condemned. This does not mean that some 'modified Hilbert's

8 This programme was first made known by Hilbert in a paper presented in 1904 at the International Congress of Mathematics in Heidelberg (See Hilbert 1905), where he even used the neologism *metamathematical* to characterize his method for investigating the consistency of a totally symbolized axiomatic system). This proposal remained uninfluential and was almost not understood at that time. Hilbert took it up again and developed it in a succession of papers between 1922 and 1928. (See Hilbert 1922, 1923, 1926, 1928); and some of his outstanding disciples, such as W. Ackermann, J. von Neumann and P. Bernays contributed to the effort of applying it concretely. In spite of the (at least partial) failure of this programme due to Gödel's theorem, the spirit of Hilbert's metamathematics and proof theory continued in a clearly defined line of foundational research, which also found expression in certain standard handbooks such as Hilbert-Bernays (1934-39) and Schütte (1960).

programme' could not work, but only by resorting to more powerful methods which, though not 'finitistic', could nevertheless be considered 'sure' or reliable. Since that time these methods have come to be called 'constructive', and by using them the consistency of arithmetic (and some - but not all - more complex mathematical theories) was actually proved⁹.

A very extensive literature has been devoted to the critical analysis of Gödel's result, and we are certainly not able to mention its various aspects here. However a few comments must be made in connection with the specific questions addressed in this paper. First of all we have to note that this result meant a definitive refutation of the *formalistic* view of mathematics, a view which claims that mathematical theories are 'nothing but' formal axiomatic structures devoid of meaning and truth. This claim, as we have seen, is tenable only if formalisms are able to provide at least the internal proof of their own consistency, which is not the case. A consistency proof is only possible by resorting to some tools or methods belonging to some other more powerful system (practically speaking, set theory), and this replicates the procedure of 'discharging' the consistency of one system onto that of some other, a strategy we recognize from the case of non-Euclidean geometries. This of course only expresses relative *consistency*, while proof of *absolute consistency* was being sought. But since the need to use 'external' tools is common to every formal system, not even the most powerful of them (e.g. set theory) could escape this condition. This means that, to the extent that a strictly formal view is accepted, only relative consistency can be established in mathematics - i.e. one establishing mutual relations among different mathematical theories - while it is impossible to obtain an absolute consistency proof either for one single theory or for mathematics as a whole.

One could assume another attitude, and say that the tools accepted for proving the consistency of, say, arithmetic, are

9 Let us only mention that already in 1936 G. Gentzen was able to obtain a consistency proof of elementary arithmetic by using as a metatheoretical tool (i.e. as a 'materially' accepted procedure) the transfinite induction of Cantor's set theory up to a 'constructively' definable ordinal number (the 'first number'). See Gentzen (1934). In more recent years this constructivist proof theory has experienced a rich variety of developments.

certainly formalizable in some mathematical theory, but are nevertheless sure and reliable *in themselves*, quite independently of that accidental circumstance. This position corresponds to the attitude Hilbert adopted towards his finitary methods, but how should we consider it? It undeniably expresses a return to *intuition* and, in this sense, equally disclaims that programme of eliminating intuition which we have seen to be at the very root of the formalistic view. This trend is even more clearly perceivable in a practice which has become very usual for granting consistency after Gödel's result: the practice of finding a *model* in which the axioms are satisfied, i.e. are *true*. This strategy was encouraged after Tarski (1936) had shown a rigorous way of interpreting formal languages, and has been developed in that important branch of mathematical logic known as *model theory*. In this case we are in the presence of a vindication of the 'classical' view: if the axioms are 'true' of something then we feel safe, since no contradiction may be deduced from true sentences. Of course, one can still remark that, after all, models are constructed by using methods that are formalizable in the last analysis within set theory, so that they ultimately depend on it. But if we accept this view, we are back again to the impossibility of proving consistency.

There is another way out, when Hilbert decided to accept the finitary methods as unquestionable, he was explicitly underlining their material and empirical nature, and this enables us to say intellectual intuition (a position which we have already met). Unfortunately such 'materially intuitive' tools proved to be insufficient. But now we may ask whether the 'constructive' methods advocated by the supporters of proof theory after Hilbert can be credited with being 'materially intuitive', and this is at least highly debatable.

Other remarks are related to a consideration of the methods used in Gödel's proof (as well as in analogous metatheoretical investigations). These methods consist in mapping the symbols of the formal system into the natural numbers, so that all the metatheoretical properties of the system are automatically mapped into concrete numerical properties. In this way these metatheoretical properties considered to be 'material' and no

longer formalistic (up to the point at which the formally undecidable sentence of Gödel's proof turns out to be universally true in arithmetic). This means that not only must the material symbols of the formal system be accepted as objects of our sensory intuition in performing such an investigation, but so must the natural numbers with their properties; and this too oversteps the strict limitations of the formalistic approach.

As a conclusion of this survey we must say that the programme of employing a strictly formalistic conception of mathematics proved to be unworkable, and that the necessity of relying upon some kind of intuition - related to meaning and reference - was recognized, in spite of the reasons which had led to mistrusting these factors in the past.

The adequacy of formalism

What we have been considering in the above section are sometimes called the internal limitations of formalism, and may be seen as disclaiming the pretension of the 'self-sufficiency' of formalism. However, a set of interesting problems remains after this renunciation and concerns the adequacy of formalism with regard to those tasks which it has often been allotted, and may be summarized by saying that formalism should free us of the servitude of meaning and reference. As we have seen in our historical account, this has to a great extent been a consequence of our having constructed *artificial* symbolic languages according to the model of abstract algebraic calculi, which are totally free of any *reference* to particular objects. This fact was soon expressed by saying that these calculi are 'devoid of *meaning*', and in such a way an identification of meaning and reference was implicitly, and even unconsciously, introduced. Moreover, since formalization appeared to be an excellent tool for eliminating the logical difficulties which seemed to nestle in several familiar 'intuitions', and since these intuitions may be qualified as 'meanings' (to the extent that meaning is often understood as being the content of some intellectual representation), the idea that formalisms are abstract constructions without meaning was

strongly reinforced, this time without implicitly identifying meaning and reference. As a result, the undeniable success of formalization was easily attributed to the fact of its having eliminated the encumbering burden of meaning and reference. But meaning and reference had always been considered the basic conditions for producing any true discourse, so that their elimination also eliminated *truth* from the formalized discourse, as we have actually seen.

The first impression one might have is that the elimination of truth was to involve total destruction, leaving only a desert, rather than the more solid knowledge which was aimed at. But the actual situation was not that negative: indeed the realm of the abstract is not a vacuum, and the systems of symbols can be 'understood', which means that they remain endowed with some kind of meaning after all. This meaning is at the same time purely intellectual and essentially operational, and since it has to be attributed to symbolic structures which are considered as 'languages', it may be reasonably qualified as a 'syntactic' meaning. What the formal point of view discards is the meaning that we could call 'eidetic', and that corresponds (as this terminology expresses through its links with a long philosophical tradition going from Plato and Aristotle to Husserl) to the 'intension' of concepts and propositions, to a mental content or representation.

These remarks show that the move towards formalization was by no means a loss, but rather a gain of *intelligibility*: the abstract turns out to be more intelligible than the intuitive, than the 'concrete', and the intellectual effort which is demanded for overstepping the eidetic intuitions is compensated by a higher degree of clarity, by an elimination of ambiguities and inconsistencies. However, one might fear that these advantages are nothing but the consequence of a terrible impoverishment, of a closure of the intellect upon itself. Of course, once this closure is in operation, everything is clear: indeed how could the intellect not 'intelligere' its own constructions? But what could be the point of such an 'intellection'? What kind of knowledge would it provide?

The point of departure had been an attempt to reach a more reliable form of knowledge of reality, while the point of arrival risks being knowledge of certain of our intellectual games. It is therefore clear that the problem is that of not making intelligibility so banal as to reduce it to emptiness, and this concretely means that a formalistic programme - if it is stripped of its possible polemic and iconoclastic shadings - appears as a challenge to *replace* the traditional features of reference, meaning and truth with something that is able to play their role in spite of being confined within the limits of what we have called 'syntactic' meaning.

Two quite different ways could be, and have been, adopted to escape the making of intelligibility banal: we have already mentioned them, and we may now add that they could be labelled as the 'realistic' and the 'idealistic' solutions respectively. The first is usually expressed by saying: formal systems do not possess any meaning, but may *receive* one, and even more than one, through suitable interpretations. In order to be fully consistent with the formalistic point of view, such an interpretation must be conceived extensionally, and not intensionally, i.e. the interpretation should provide referents, sets of referents, sets of ordered n-tuples of referents, etc. to be directly connected with the symbolic expressions of the formal language, without the mediation of any intensional or eidetic meaning. We could say that according to this view there is, apart from the syntactic meaning, only a 'referential meaning', which is sufficient to avoid banality and also to recover the notion of truth via the notion of 'satisfaction'. The second solution is usually expressed by saying: the formulae of the formal systems *determine* the meaning of their primitive notions contextually, and these notions become at the same time the objects of the discourse. We have called this position idealistic, because it shares with idealism the basic tenet that there are no things or referents 'outside' thought, that they are determined by thought and nothing more. One could say that in the formalistic context it is said of language and not of thought; however, since this language is not seen as simply a heap of material signs, but rather as a structure endowed with 'syntactic meaning', the real

situation is that this meaning is claimed to be able to exhaust the roles of the eidetic and referential meanings, and this escapes banality only within the idealistic claim that there is no reality other than that constructed through thinking.

Let us now try to evaluate the success of these two strategies in avoiding the use of eidetic meaning. Following the 'realistic' approach, the adequacy of a formalism should be determined on the basis of its ability to fully characterize a universe of 'given' referents or objects. Now certain well known theorems of mathematical logic show that this ability is not very great. On the one hand we know since the time of Gödel's theorem that even such a simple theory as elementary arithmetic is 'semantically incomplete', i.e. that *any* formal system designed with the aim of describing the structure of the natural numbers will be unable to capture *all* their properties (since there are propositions about these numbers which are true of them, but which are not provable in the formal system). On the other hand the 'isomorphism theorem' of mathematical logic proves that if a formal system admits a model in a given universe of referents, it automatically admits an infinity of other models isomorphic with the first. This fact indicates that we can at best formally characterize only the *structure* of our domain of referents. Still this happens to be a rather exceptionally fortunate situation, for it is known already from a theorem of Skolem¹⁰, that in general formal systems are

10 The theorem, technically known in the literature under the name of «Löwenheim-Skolem», since it was partially anticipated by L. Löwenheim in 1914, says that if a set of formulae of first-order logic is simultaneously satisfiable in any non-empty domain (i.e. if the set has a model at all), then it is simultaneously satisfiable in a denumerable domain (i.e. it has a denumerable model). Since an axiom system is a set of formulae, it follows that every first order theory that has a model has a denumerable model (see Skolem 1919). This result already implies that if our intention with our axiomatization were that of characterizing a domain of objects having a finite, or a more than denumerable cardinality, we would not succeed in our effort, since that axiomatization would also be satisfied in a domain with a denumerable cardinality (it is also provable that it would have a model of any arbitrary infinite cardinality). By developing this line of thought, Skolem (1933, 1934) proved that if a first order theory of arithmetic (with identity) has its intended model, then it also has a normal model that is not isomorphic with its intended model. This was the origin of what has been called *non-standard arithmetic*. A. Robinson, by using similar methods for a formal system of the theory of real numbers, has developed a theory of *non-standard analysis* (see Robinson 1966).

not 'categorical', i.e. that they admit different models which are not even mutually isomorphic, and this is tantamount to saying that they cannot even univocally characterize the 'structure' of a given domain of referents (in the sense that what they say about this domain is equally valid for some non-standard or unintended models). It would be out of place to enter here into technical details concerning the strategies for recovering categoricity at least to certain degrees; and in any case it would appear that these strategies imply giving up completeness, so that these two features, which should *jointly* grant the ability of a formal system to perform referentially, are hardly compatible, therefore making such a performance not really adequate.

Coming now to the 'idealistic' strategy, its adequacy would consist, as we have seen, in a kind of capacity of the formal system to 'produce' a domain of referents, in order that it not be simply an empty language, but one which, at least in a minimal sense, speaks 'about' something. This requirement is imposed by the fact that a formal system which is supposed or even intended to express a certain specific theory (e.g. elementary arithmetic) in a strict sense. Indeed even in mathematics a difference is admitted (except by logicians) between 'logical truths' and 'mathematical truths', and if one agrees to qualify as logically true those propositions which are true in all possible models (or in all 'possible worlds'), one should say that mathematically true propositions are true only in some possible worlds. Our problem may now receive the following more precise formulation: given a set of propositions satisfying the minimal 'formal' requirement of being consistent, can we be sure that there exists at least one possible world in which these propositions are true? We know that several mathematicians have been and are of this opinion, including some who were not ready simply to identify mathematical existence with consistency (let us only mention Poincaré, who was a pre-intuitionist in certain respects while nevertheless being inclined to accept the view that every consistent set of mathematical propositions admits a model).

What does mathematical logic tell us about this claim? It says that if we can prove, for whatever formal language, that *every* consistent set M of sentences of that language admits a model,

then it is possible to formulate in this language a logical calculus that is 'semantically complete'. This result already shows how the above claim is untenable: if we really maintain that *every* consistent set of mathematical sentences has a model, we should conclude that *every* logical calculus is semantically complete, while it is known that semantical completeness is a property which does not hold in general for logical calculi beyond the level of first-order logic. This conclusion is already interesting because it shows at any rate that consistency and truth are not totally identifiable even in mathematics, since there is at least some possibility of discriminating between them. A 'realist' would say that all that is obvious, for it was actually too pretentious to suggest that consistency be a sufficient warranty for the 'existence' of objects. However, we shall try to be fair and see whether there is at least a sense in which a consistent set of sentences may 'create' its own model, without pretending this to be a 'creation *ex nihilo*', but rather a construction obtained by using the ingredients provided by the language itself in which the consistent set of sentences is formulated. This is clearly quite a generous concession, but it is only given this concession that it is possible to prove that every consistent set of sentences formulated in the language of *first-order* logic possesses a model (this model being constructed out of the sentences' linguistic ingredients), while this is no longer the case if the axioms are formulated in a higher-order language¹¹. This fact is of

-
- 11 One could object that, by using the same techniques introduced by Henkin for the construction of the model of any consistent set of first-order sentences, it is also possible to obtain the same result for consistent sets of sentences of higher order, as is shown e.g. in Henkin (1949, 1950). But this fact entails that semantic completeness also holds for second order logic and for the theory of types. This is contrary to what is commonly said in textbooks of mathematical logic, and contrary to the fact that the semantic incompleteness of second order logic derives directly, e.g., from Gödel's completeness theorem for first-order logic with his incompleteness theorem for elementary arithmetic and the categoricity of second order arithmetic. Without going into technicalities, it is clear that semantic completeness and the existence of a model for any consistent set of sentences are understood in a *particular and partially different* sense in these enlarged versions, so that they are not incompatible with the preceding claims and results. In particular, the domains of objects which are introduced are *non-standard*, i.e. they respect the condition that transformation rules preserve the truth from the premisses to the consequences in such models, which, however, are not maximal. Moreover the condition of extensionality

importance especially if we keep in mind the philosophically relevant point of the 'idealistic' claim, i.e. that the referents of a formal theory, such as the natural numbers, are created by the formal system itself. Now, according to the result just mentioned, the warranty for the existence of such referents would depend not so much on the skilful choice of consistent set of axioms, but basically on the expressive power of the language used. This is already strange, but it becomes more puzzling if we consider that this warranty decreases with the improvement of the richness and expressiveness of the language. This fact contradicts the spirit of any idealistically inspired approach, for which the mark of 'truth' or soundness is typically represented by the comprehensiveness of a system, and its capability of embracing the widest variety of properties, relations and their levels (which corresponds to an increase in the expressive power and the complexity of the language)¹².

However, this is still less important than another fact. In the construction of the said model, the usual methods are applied which characterize the Tarskian procedure for defining interpretations and the satisfaction of the expressions of a formalized language, and these methods presuppose that a certain domain of objects be given which is distinct from the language that will be interpreted on it and 'speak about' it. Now this

is not preserved. Let us note that the criteria which preside over the construction of such models and make reasonable the weakening of the standard semantical conventions are dictated by the goal of assuring the strictest possible correspondence (which however cannot in general be one-to-one) between the components of the language and those of the 'linguistic' domain on which they are interpreted. In this way the components of the language replicate the features of the 'linguistic' model obtained for the consistent sets of first-order language.

- 12 It would be instructive, but would take us too far, to investigate the strict links which exist between the idealistic thesis of the 'wholeness' of truth, with its consequent doctrine of 'internal relations', and the idea of the self-containing nature of formal systems, which 'internally' determines all the meaning and truth (these notions being taken 'formally') of all their components. This idealistic doctrine has known a variety of developments and modifications, from Hegel to Hamelin, Bonsaquet, Bradley, Gentile, Joachim and others. An account of it may be found e.g. in Ewing (1934). A heated polemics was engaged in against this line of thought in the first decades of our century by Moore and Russell, whose criticism of the 'axiom of internal relations' was expressed particularly in Russell (1906, 1910). However, a significant family resemblance with this doctrine may be found in the 'coherence theory of truth' recently elaborated by N. Rescher. See in particular Rescher (1973).

distinction is eliminated in the construction of such a model, since the domain of objects is constituted by the set of closed terms of the language, by a kind of double 'self-reference': every term is taken as the metalinguistic name of itself, and every sentence is declared to be 'true' of the terms it contains simply if and only if it belongs to the consistent set M^{13} . Hence one should at least be conscious of the very particular situation here, which we might describe as follows. We can take the above result to mean that, at least in the case of first-order languages, every consistent set M of sentences describes a 'possible world'. Then we ask: «what possible world?», and the answer we receive is: «the world described by M , of course». The situation seems quite ironical, but actually depends on the fact that the 'referential' feature of truth, which is contained in the Tarskian methods, is reduced to a pure appearance, so that no possibility of 'falsifying' the sentence of M really exists. In fact, since they do not speak about any independent structure of objects, such sentences are protected against any extra-linguistic or referential refutation, and they are also protected against the only remaining source of 'linguistic' refutation, i.e. contradiction, since the set M is *assumed* to be consistent.

At this point one might ask why such a result is taken seriously in mathematical logic. There are at least two reasons. The first is that the procedure of interpreting a language in terms of itself, while certainly 'artificial', is not absurd, and is adopted with profit in certain branches of abstract mathematics (e.g. when as a representation of a group one takes the permutations on a domain whose elements are the elements of the group itself):

-
- 13 It is perhaps worth mentioning that models of consistent sets of first order sentences may be constructed using as a domain of objects not the ingredients of the language, but e.g. natural numbers. However this does not really change the situation, for in the construction of these models the strategy of numerically 'coding' the individual variables of the sentences is usually followed, so that the model is constructed using these coding numbers. This move does not imply any reference to the structure of natural numbers proper, since, when the sentences are interpreted, they are again 'declared' true if and only if they belong to M and not if they express some *numerically* true property in a real sense. This is possible thanks to the extensionalist semantics, which makes properties and relations in the model entirely dependent upon the language, without any 'reference' to the structure of the domain of objects. For a more detailed discussion of these issues see Agazzi (1978).

what matters is to understand what this procedure actually means. The second reason is that, after all, even this 'artificial' model is not granted for logical languages of a higher order, and this fact is strictly bound to such an important requirement as the 'semantic completeness' of the logical calculi, which does not hold in general, in a proper sense, beyond first-order logic.

We can derive two morals from the above reflections. The first is that formalisms may perform a kind of weak truth-granting role (which we may call 'truth by consistency') at the level of first-order logic, while they are not generally in the position of performing even this role at higher levels. The second is that the existence of these results seems to impose a moderation upon the formalistic trends in mathematics and upon a certain excess of importance given to axiomatizations of all possible kinds, from which (under the assumption that they are consistent) many curious theorems are deduced. The question is: are these axiom systems really mathematical? In order to show that they are, one should show that they can apply to some recognized 'mathematical objects' or (if one prefers to remain ontologically uncommitted) to be of help in the solution of 'serious' mathematical problems (the requirement of seriousness being expressed by the fact that such problems are not just 'internal' to the formalism which is intended to solve them). This is the gist of the reaction of certain outstanding mathematicians of our time against 'gratuitous' abstractions and 'meaningless and uninteresting' axiomatizations.

Formal and non-formal

The different reasons for the inadequacy of the radical formalism which we have considered above may probably be clarified through an analysis of the way in which a language may be used in speaking about a given domain of objects. The first condition is, obviously, that these objects be there, and this is not realized through a speech act, but through the *presence* of these objects to our thought. We shall here call this presence 'phenomenological', in order to cover by this term all possible

'ways' of being present, and to suspend any judgment about the ontological status of what is present. It also seems legitimate to call this phenomenological situation *truth*, for it is such that an object, simply by being present, offers to thought an irrefutable (and perhaps the only irrefutable) witness of itself. Therefore the phenomenological truth is 'unstable' and 'private': it is unstable because it does not allow us to remain within the realm of truth when we leave the immediate presence; it is private because the presence of certain objects is such only for that individual for which they are actually and instantaneously present. Now it can be said that a fundamental function of language is that of overcoming these shortcomings, enabling us to 'preserve' in some way truth also outside the instant of its immediateness, and to make it *intersubjective*. In such a way the characteristic of truth, which is intrinsic to the phenomenological situation, is transferred to language, as is shown by the fact that the most common (and most appropriate) use of the term "true" concerns the propositions of a language, and is primarily attributed to them when they 'denote' a state of the objects which is phenomenologically present.

But language itself can enter the sphere of presence, and this under two aspects. On the one hand it is present with its structures and ways of functioning; on the other its 'denoting' (i.e. its referring to another sector of presence) is also present. Formal logic may be seen as belonging to this stage and, more specifically, to the investigation of those functions of the language capable of leading from propositions denoting a presence to other propositions also expected to denote a presence. This is tantamount to saying that logical inference itself develops within the phenomenological sphere (i.e. so-called 'logical evidence' is a special kind of phenomenological evidence). And thanks to this, by means of a logically evident inference we can bring a proposition which does not phenomenologically denote a presence into the domain of truth, provided we can derive it from propositions which are phenomenologically evident.

This quick description actually involves several delicate steps. First of all, the language must rely upon tools for 'retaining' the presence of objects even when they are actually no longer there,

and these tools are the *meanings*, which in this way appear to be 'extracted' from the *referential situation* (which is the situation of presence), but do not coincide with it, since they hold even outside this situation. However, owing to this original link, they may be seen as a permanent and open 'possibility of referring'. It is worth noting that meanings are already only partially 'faithful' with regard to any particular phenomenological presence or referential situation they might denote. For example, the concept 'man' does not contain all the details of every single man who might be denoted through it, but its not doing so is part of the condition for it to denote men who do not share all these details.

Concepts or meanings are mental entities and, as such, are private. In order to become public they must be associated with the expressions of a language, and this determines the passage to the *formal* level, «formal» being here understood (in conformity with an already explained view) as the realization of conditions for explicitness and non-ambiguity, so that the correct application of these conditions should enable man to understand what another 'means' by using a certain expression. This stage entails the creation of a complex structure, since the compactness of the 'presence' is not only analyzed through a complex net of mutually interwoven meanings, but the language itself must then replicate to some extent the complexity of this meaning-structure in order to express it. This is why the *semantic* performance of the language (i.e. its ability to express meaning) necessarily presupposes its possessing a certain *syntactic* structure. On the other hand a syntactic structure, the different components of which can be seen as having the 'possibility of expressing a meaning' in a sense analogous to that of meanings having the possibility of denoting referents. Meaningful sentences which also denote a presence are said to be true, and a conspicuous part of syntax, as we have said, consists in exploring the structure of the domain of true sentences.

Let us now consider some consequences of certain technical results which have been surveyed in the preceding sections. Certain of them say that a formal system (i.e. a language according to our last remarks) on the one hand says 'more', and on the other says 'less' than what is true of its intended models

(or referents). This means that no language can express the 'presence' completely adequately. This is already the case with any private use of the language (my linguistic description is unfaithful with respect to what is present to my thought). Moreover, an interlocutor receiving my linguistic communication will translate it to denote a presence for him, and an additional inadequacy will be added. All this is tantamount to saying that absolute truth (which we have seen to coincide with the situation of phenomenological presence) is not intersubjective. Another limitation we have seen is that in general formal logical calculi cannot deductively cover the whole domain of the true sentences expressible in a formal language (semantic incompleteness).

These facts may be restated by saying that it is not possible to fully characterize a structure of objects, through a language, or to fully master even that portion of 'truth' which the language is able to express about these objects through the deductive operation of the language. The first fact indicates the inadequacy of the semantic dimension with respect to the referential or phenomenological one; the second indicates the inadequacy of the syntactic dimension with respect to the semantic one. Moreover, it must be stressed that these conclusions could be reached within the phenomenological presence in which the analysis of the denotation and of the syntactic structure of the language were performed. This links with what we have noted when speaking of the impossibility of escaping 'intuitive' reasoning of different kinds in metatheoretical investigations.

These remarks are by no means intended to discredit formalism. They only reject the opposed pretensions of 'reducing' one level to another, on the one hand, and of 'disconnecting' them on the other. But once these mistaken attitudes are avoided, the great merits and advantages of formalism become clear. Indeed, the fact that we have spoken of 'inadequacies' may suggest the idea of a negative appreciation, but we have actually seen *why* it was not only reasonable but necessary to pay this price in order to overstep the authentic limitations of the original situation of phenomenological truth. Two of these reasons have already been indicated: proceeding towards formalization allows one to overcome the instability and

privacy of the original truth. But something else may be added. In the first place, the advancement in the construction of a more and more complex net of meanings, which is made possible partly by the increasing formal complexity of languages, corresponds to the achievement of higher levels of universality. Therefore the trend towards formalization does not express a 'flight' from reference and meaning, but a search for universality, which allows for a unification of our experience.

From this point of view, formalization is the latest expression of our aims at unifying the core of our rational understanding of reality, and has been the constant driving force of philosophy since the earliest times. This effort has to do with the fact that the immediate 'presence' is instantaneous while at the same time we feel ourselves to be in the presence of 'one' reality; and this challenges us to find 'general' tools for connecting the instantaneous witnesses of the presence, so as to express the wholeness of reality. Formalization, as we have seen, offers a very efficient means to this end.

A second reason is perhaps even deeper. The maturation of human knowledge (both individual and collective) corresponds to the awareness that the content of the immediate presence is only a part of the much richer dimensions of reality, so that the great cognitive adventure of man consists in going behind immediateness to enlarge the domain of truth. But then it is clear that the elaboration of instruments consisting precisely in 'detaching' us from immediateness is of paramount importance. These instruments belong to the sphere which we could broadly call that of *abstraction*, and they are to be used very delicately, since we are not sure where they will lead us, once we leave (as we *must*) the sound ground of the immediate presence. From this point of view it is undeniable that the development of formalization in the history of mankind corresponds to a powerful process of *controlled and reliable use of abstraction*, the eloquent results of which may be found in the different sciences and in the great amount of knowledge they have reached by overstepping immediateness (and it is not by chance that this process has been accelerating with the increase of the use of formal tools in most recent times).

The impulse towards formalization, as we have seen, has been especially powerful in the last two centuries, when it has also expressed the manifestation of the free (though not fully arbitrary) creativity of the human mind. From this point of view, the formal way of thinking is a manifestation of the typically 'modern' awareness of the complexity of knowledge, and of the great role which the subject plays in it. The active intervention of the subject in the construction of knowledge has been stressed in several ways and it has sometimes led to the annihilation of the idea of an object of knowledge. Here again we can see in the process of formalization a very valuable expression of many ways in which this constructive aspect of human subjectivity intervenes in the knowing process; and investigations on the 'limitations' of formalism may also be understood as quite objective indications of the impossibility of eliminating the objects of knowledge. This is perhaps one of the most significant contributions we have reached on the clarification of the intimate complementarity of these two aspects, just to give an example: it is often (and correctly) claimed that a mathematician must first intuitively 'see' a theorem before formally proving it. However it is equally true that what he 'sees' will not become a theorem until it has been formally proved. More generally, we must recognize that our knowledge of reality has known astonishing advance, e.g. in the domain of the natural sciences, thanks to the application of more and more abstract conceptualizations and formal tools.

*Séminaire de Philosophie
Université de Fribourg
Miséricorde; CH 1700 Fribourg*

Bibliographical references

- AGAZZI, E. (1978). Non-contradiction et existence en mathématique. *Logique et analyse*, 21, 459-481.
- BIRKHOFF, G. & von NEUMANN, J. (1936). The logic of quantum mechanics. *Ann. of Math.*, 37, 823.
- EWING, A.C. (1934). *Idealism. A Critical Survey*. London: Methuen.
- GALILEI, G. (1929-1939). *Opere*, Ediz. Nazionale. Firenze: Barbera.
- GENTZEN, G. (1936). Die Widerspruchsfreiheit der reinen Zahlentheorie, *Mathem. Annalen*, 112, 493-565. (Engl. transl. in: M.E. Szabo (ed.), *The Collected Papers of G. Gentzen*. Amsterdam: North-Holland, 1969).
- GÖDEL, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I, *Monatshefte für Mathematik und Physik*, 38, 173-98. (Engl. transl. in: van Heijenoort 1967, 595-616).
- HENKIN, L. (1949). The completeness of the first-order functional calculus. *The Journal of Symbolic Logic*, 14, 159-166.
- HENKIN, L. (1950). Completeness in the theory of types. *The Journal of Symbolic Logic*, 15, 81-91.
- HILBERT, D. (1899). *Die Grundlagen der Geometrie*. Leipzig: Teubner, 10th edition with supplements by P. Bernays, Teubner, Stuttgart, 1968. (Engl. transl. by E.T. Townsend, Chicago 1902).
- HILBERT, D. (1904). Über die Grundlagen der Logik und Arithmetik. *Verhandlungen des Dritten Internationalen Mathematikerkongresses*. Leipzig: Teubner. (Engl. transl. in: van Heijenoort 1967).
- HILBERT, D. (1922). Neubegründung der Mathematik. Erste Mitteilung. *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität*, vol. I, 157-177. (Repr. in: *Gesammelte Abhandlungen*, vol. III, 178-191).
- HILBERT, D. (1923). Die logischen Grundlagen der Mathematik. *Mathem. Annalen*, 92, 151-15. (Repr. in *Gesammelte Abhandlungen*, vol. III, 178-191).

- HILBERT, D. (1926). Über das Unendliche. *Mathem. Annalen*, 95, 161-190.
- HILBERT, D. (1928). Die Grundlagen der Mathematik. *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität*, vol. VI, 65-85. (Repr. as *Die Grundlagen der Mathematik*. Leipzig: Teubner. Engl. transl. in van Heijenoort (1967), 464-79).
- HILBERT, D. & BERNAYS, P. (1934 - 39). *Die Grundlagen der Mathematik*. Berlin: Springer. (2nd ed. revised by P. Bernays, Berlin: Springer, 1968-70).
- HILBERT, D. & COHN-VOSSEN, S. (1932). *Anschauliche Geometrie*. Berlin: Springer. (Engl. transl. by P. Nemenyi, *Geometry and imagination*. New York: Chelsea, 1956).
- KLUG, U. (1951). *Juristische Logik*. Berlin: Springer.
- MEINONG, A. (1904). *Über Gegenstandstheorie*. Leipzig: J.A. Barth.
- MITTELSTAEDT, P. (1968). *Philosophische Probleme der modernen Physik*. Mannheim: Bibliographisches Institut. The VI chap. of this book contains results of previous research published by the author in 1959-1961.
- NEUMANN, J. von & MORGENSTERN O. (1944). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.
- PADOA, A. (1900). Essai d'une théorie algébrique des nombres entiers, précédé d'une introduction logique à une théorie déductive quelconque, *Bibliothèque du Congrès International de Philosophie (Paris 1900)*. Paris: Colin, 1901, 309-365.
- PASCH, M. (1882). *Vorlesungen über neuere Geometrie*. Leipzig: Teubner.
- PIERI, M. (1899). Della geometria elementare come sistema ipotetico-deduttivo; monografia del punto e del moto. *Memorie della Reale Accademia delle Scienze di Torino. Classe di Sc. Fisiche, Matematiche e Naturali*, Serie 2, 49, 173-222.
- RESCHER, N. (1973). *The Coherence Theory of Truth*. Oxford: Clarendon Press.
- ROBINSON, A. (1966). *Non-standard Analysis*. Amsterdam: North-Holland.

- RUSSELL, B. (1906). The Nature of Truth. *Mind*, 15, 528-533.
- RUSSELL, B. (1910). The monistic theory of truth. *Philosophical Essays*. London: Longmans.
- SCHEIBE, E. (1964). *Die kontingenten Aussagen in der Physik*. Frankfurt-Bonn: Athenäum.
- SCHÜTTE, K. (1960). *Beweistheorie*. Berlin: Springer. (Engl. transl. 1977).
- SKOLEM, T. (1920). Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theoreme über dichte Mengen. *Videnskapsselskapets skrifter. I: Matematisk-naturvidenskabelig klasse*, n. 4, Kristiania, 36ff. (Engl. transl. in: van Heijenoort 1967, 254-263).
- SKOLEM, T. (1933). Über die Unmöglichkeit einer vollständigen Charakterisierung der Zahlenreihe mittels eines endlichen Axiomensystems. *Norsk matematisk forenings skrifter*, series 2, n. 10, 73-82, Oslo.
- SKOLEM, T. (1934). Über die Nicht-charakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschliesslich Zahlenvariablen. *Fundamenta Mathematicae*, 23, 150-61.
- SUPPES, P. (1969). Stimulus-response theory of finite automata. *Journal of math. Psychology*, 6, 327-355.
- TARSKI, A. (1936). Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia philosophica*, 1, 261-405. (Engl. transl. in: A. Tarski, *Logic, Semantics, Metamathematics. Papers from 1923 to 1938*. (Transl. by J.H. Woodger. Oxford: Clarendon Press, 1956).
- WEIZSÄCKER von, C.F. (1955). Komplementarität und Logik. *Naturwissenschaft*, 42, 521, 545. (Repr. in *Zum Weltbild der Physik*. Stuttgart: Hirzel, 1976).

IMPORTANCE OU INSIGNIFIANCE MATHÉMATIQUE DES INDECIDABLES DE GÖDEL?

Jean-Paul Delahaye

1. Introduction

«95% des mathématiciens se moquent éperdument de ce que peuvent faire tous les logiciens et tous les philosophes» (Dieudonné 1982: 16).

«The feeling was that Gödel's theorem was of interest only to logicians». Robert Solovay. Cité par Gina Kolata (Kolata 1985).

«La proposition indécidable A décrite par Gödel paraît très artificielle, sans lien avec aucune autre partie de la théorie des nombres actuelle; sa principale utilité était d'établir l'impossibilité d'une preuve de la non contradiction de l'arithmétique. Parmi les nombreuses questions classiques non résolues de la théorie des nombres, on n'a pas encore à ma connaissance établi que l'une d'elles est indécidable» (Dieudonné 1987).

«What people would really like is to take a big unsolved problem like Fermat's Last Theorem and show it is undecidable. That would be spectacular». Joel Spencer. Cité par Gina Kolata (Kolata 1985).

«La croyance que la théorie de la démonstration de Hilbert fait partie intégrante de la mathématique (...) ne nous paraît pas justifiée, et nous considérons que l'intervention de la métamathématique dans l'exposé de la logique peut et doit être réduite à la partie élémentaire qui traite du maniement des symboles abrégiateurs et des critères déductifs» (Bourbaki 1969: 57).

1.1. Importance philosophique et importance mathématique

Les indécidables dont les théorèmes de Gödel de 1931 nous démontrent l'existence et qu'ils nous exhibent, ont-ils une signification mathématique réelle, ou au contraire sont-ils des énoncés pathologiques sans intérêt autre que logique ou philosophique?

C'est cette question que nous voulons nous poser. Nul ne conteste l'intérêt logique ou philosophique de ces indécidables, seul leur intérêt mathématique est en cause. Et en effet il se trouve que bien des mathématiciens les considèrent avec prudence et les jugent comme des curiosités. N. Bourbaki par exemple ne les mentionne que dans ses notes historiques.

Tenter de comprendre ce que peut vouloir dire «les indécidables sont sans véritable importance pour les mathématiques» est un défi tant il est évident que ce qui est important pour la philosophie des mathématiques l'est aussi pour les mathématiques elles-mêmes. Notre méthode sera de parcourir les différentes classes d'indécidables que les logiciens ont mis à jour depuis 1931. A propos de chacune nous imaginerons un dialogue contradictoire sur l'importance mathématique des indécidables. Parfois aussi nous envisagerons le problème de la physique. Notre conclusion sera que si soutenir l'insignifiance mathématique des indécidables est encore possible, cela se fait toujours au prix d'un réductionnisme qui possède plusieurs formes dont les deux plus extrêmes sont le formalisme et le finitisme. Certaines variantes du réductionnisme comme le prédicativisme de Feferman, se nourrissant des récents résultats de la théorie de la preuve, refusent de dire leur dernier mot et prétendent même faire revivre le programme de Hilbert qu'on pensait enterré justement par les théorèmes d'indécidabilité de 1931.

Nous verrons finalement que le problème de l'importance mathématique des indécidables de Gödel concentre en lui tout le problème de la philosophie des mathématiques, en un mot que le problème de l'importance mathématique des indécidables de Gödel est un problème important ... philosophiquement!

1.2. Comment définir un indécidable de Gödel

Nous commençons par proposer une définition des «indécidables de Gödel». La définition que nous donnons n'est pas mathématique, elle a simplement pour but de préciser un usage établi qui refuse que n'importe quel énoncé indépendant de n'importe quel système formel puisse être considéré comme un «indécidable de Gödel». Plus loin nous serons d'ailleurs amenés à préciser que parmi les indécidables de Gödel il faut encore distinguer ceux qui renforcent réellement les systèmes auxquels on les ajoute (comme les énoncés de consistance) de ceux qui sont conservatifs (comme l'axiome du choix, ou le lemme de König).

Un indécidable de Gödel c'est: un énoncé mathématique vrai et non démontrable, plus précisément: un indécidable de Gödel c'est un énoncé mathématique E tel que:

(1) E n'est pas démontrable dans un système formel S_1 qui permet de l'exprimer.

(2) On peut montrer (dans un système formel S_2) la non-prouvabilité de E .

(3) On peut montrer (dans un système formel S_3) que E est vrai.

(4) Le système S_1 est une **axiomatisation raisonnable** d'un champ mathématique donné et reconnu comme intéressant, et on a pu penser que cette axiomatisation formalisait bien le champ en question, (comme par exemple l'arithmétique de Peano du premier ordre vis-à-vis de l'arithmétique). Cette condition est réalisée si on a cru à un moment donné que S_1 est une axiomatisation complète du champ en question. Cette condition est aussi satisfaite si certains soutiennent que tous les énoncés intéressants (dans un sens à préciser) du champ en question sont prouvables avec S_1 . Beaucoup de mathématiciens pensent que $ZF + AC$ par exemple permet d'exprimer et de démontrer tout énoncé mathématique vraiment intéressant. La condition (4) sert à éviter qu'on appelle «indécidable de Gödel» n'importe quel axiome indépendant d'un système S_1 clairement trop faible: l'axiome des parallèles par exemple n'est pas considéré comme un indécidable de Gödel, de même l'axiome de l'infini n'est pas un indécidable de $ZF - \{\text{axiome de l'infini}\}$.

(5) Le système S2 est un système dont on a certaines raisons de penser qu'il est consistant (par exemple ZF + AC).

(6) Le système S3 est un système raisonnable dont on a aussi certaines raisons de penser qu'il est consistant. Dans le cas de certains axiomes des grands cardinaux en théorie des ensembles, les raisons de croire à la consistance de S3 sont assez ténues (S3 est obtenu en ajoutant un axiome de grand cardinal à ZF + AC), il y a même des cas où la consistance de S3 est douteuse.

Remarques:

- On pourrait peut-être supprimer (3) mais alors il faudrait préciser «dont la négation n'est pas démontrable».

- Le plus souvent on ne précise que S1. Les systèmes S2 et S3 sont plus forts que S1, et par exemple il est fréquent que S2 et S3 soient équivalents à S1 + consistance(S1).

2. Classification des indécidables de Gödel

Dans ce paragraphe nous présentons une classification des indécidables de Gödel et pour chaque type d'indécidables nous imaginons un petit dialogue entre **Monsieur Insignifiance** qui ne trouve aucun intérêt mathématique réel à ces indécidables, et **Monsieur Importance** qui lui leur trouve un sens et défend l'idée qu'ils sont d'authentiques résultats mathématiques. Les dialogues sont quelque peu naïfs, sauf à propos des derniers éléments de la classification où ils prennent un tour un peu plus technique.

2.1. Les indécidables de Gödel donnés par la démonstration du premier théorème d'incomplétude de Gödel

Ils ont un sens simple, moyennant des considérations métamathématiques sur la prouvabilité.

Ils signifient en effet: «je ne suis pas démontrable dans S1».

S1 peut être n'importe quel système extension (primitive réursive) du système Q de Robinson qui est un système plus faible que PA (l'arithmétique du premier ordre de Peano) et même que PRA (l'arithmétique primitive réursive). Voir par exemple (Boolos Jeffrey 1980) pour des définitions précises, voir aussi (Girard 1987) où Q est noté EA. La restriction aux extensions primitives réursives n'est pas une véritable restriction (voir Girard 1987: 58).

Ils sont démontrés exister, mais mieux que cela, ils sont produits effectivement: si S1 est donné, la démonstration de Gödel permet d'explicitier un indécidable de S1.

Pour S2 on peut prendre S1+consistance(S1). Dans la démonstration originale de Gödel S3 est plus fort que S1+consistance(S1) puisque Gödel utilisait l' ω -consistance de S1 (l'impossibilité dans S3 de démontrer $\exists n P(n)$ en même temps que $\text{non}P(0)$ $\text{non}P(1)$ etc. C'est Rosser (1936) qui en modifiant l'énoncé de Gödel (qui n'a donc plus un sens aussi simple) a permis de prendre S2=S3= S1+consistance(S1).

Les formulations (données par Kleene) du premier théorème de Gödel utilisant les concepts de la théorie de la réursivité sont très intéressantes, elles s'appliquent à tout système formel et d'une certaine façon éclairent le phénomène de l'indécidabilité: l'ensemble des vérités de l'arithmétique du premier ordre n'est pas réursivement énumérable, l'ensemble des théorèmes que peut démontrer un système formel est toujours réursivement énumérable (à cause des conditions d'effectivité qu'on impose aux systèmes formels: il faut qu'un procédé mécanique puisse vérifier si une suite donnée de formules est bien une démonstration) donc **aucun système formel ne peut produire que des vérités de l'arithmétique du premier ordre et les produire toutes.**

Monsieur Insignifiance: ces énoncés étranges qui affirment au travers d'un codage épouvantable leur propre indémontrabilité sont des divertissements mathématiques amusants et non des mathématiques.

Monsieur Importance: le paradoxe du menteur, et les paradoxes de la théorie des ensembles qui se présentent sous des formes très proches ont touché les mathématiciens du début du siècle, qui y ont vu plus que des divertissements insignifiants, alors pourquoi refuser de considérer comme authentiquement mathématiques ces premiers indécidables. Il est dans la nature des mathématiques ces premiers indécidables. Il est dans la nature des mathématiques de s'intéresser à toutes les conséquences qu'on peut tirer des méthodes de raisonnement dont elles disposent. Ce que fait le premier théorème de Gödel, ce n'est que cela.

2.2. Les indécidables de Gödel donnés par la démonstration du second théorème d'incomplétude de Gödel

«I regard Hilbert as a mathematician and thus his beloved consistency as mathematical» (Smorynski 1982).

«I do not know of any case where anything of clear mathematical interest can be proved from $T + \text{cons}(T)$ that cannot be proved already from T (except $\text{cons}(T)$ itself)» (Drake 1985: 30).

Les indécidables du second théorème d'incomplétude (qui sont équivalents à ceux donnés par le premier théorème, (Smorynski 1977: 829)) ont un sens encore plus simple, toujours moyennant des considérations métamathématiques sur la prouvabilité.

Ils signifient en effet: «le système S_1 est consistant».

Comme précédemment ils sont produits effectivement.

Le système S_1 peut être n'importe quelle extension (primitive récursive) du système PRA de l'arithmétique primitive récursive (plus forte que le système Q) (Boolos Jeffrey 1980; Girard 1987).

Comme précédemment pour S_2 et S_3 on peut prendre $S_1 + \text{consistance}(S_1)$, mais bien sûr démontrer la consistance de S_1 dans un système qui contient comme axiome la consistance de S_1 ne paraît pas très intéressant, aussi pour S_3 il est plus naturel de penser qu'on fait appel à une théorie puissante comme

l'arithmétique du second ordre ou ZF, ou une théorie contenant un principe d'induction.

Dans le cas où S1 est ZF pour se convaincre de consistance(ZF) on peut faire appel à des axiomes de grands cardinaux qui en un certain sens sont naturels et plausibles, mais bien évidemment plus on considère des théories fortes plus la consistance de ces théories devient problématique. Comme nous l'avons déjà indiqué certains axiomes forts concernant les grands cardinaux sont suspectés d'être inconsistants! (voir par exemple Maddy 1988).

Monsieur Insignifiance: Si on ne s'était pas intéressé à la formalisation des démonstrations aucun mathématicien n'aurait jamais rencontré ces énoncés étranges, dont on ne peut comprendre le sens littéral tant ils sont compliqués et tant est indéchiffrable la codification qui les lie à leur interprétation métamathématique.

Monsieur Importance: Si on peut prétendre que les indécidables produits par le premier théorème de Gödel sont d'intérêt mathématique douteux, cela est plus difficile pour les indécidables du second: la consistance de l'arithmétique de Peano est un problème qui avait été étudié avant qu'on sache qu'elle était indécidable (pour PA). Par Hilbert lui-même! (Cf. la citation de Smorynski). Le codage est désagréable certes mais il est possible dans tout système contenant un peu d'arithmétique et c'est justement parce qu'il est possible qu'on rencontre cette difficulté. Ou bien il faut le rendre impossible (comme en calcul propositionnel ou en géométrie élémentaire) ou bien il faut en accepter les conséquences. A ce propos, notons que N. Bourbaki qui juge par ailleurs d'intérêt mathématique secondaire la théorie de la démonstration (et donc les théorèmes de Gödel) dit de la démonstration de Gödel qu'elle est ingénieuse et que le grand nombre de signes nécessaires pour écrire les indécidables de Gödel bien que les rendant pratiquement impossibles à expliciter est sans gravité car «aucun mathématicien ne pense que cela diminue en rien la valeur de ces constructions» (Bourbaki 1969: 60).

2.3. Les indécidables de Gödel produits par la théorie de la récursivité

Développée à partir de 1936 la théorie de la récursivité (c'est-à-dire de la calculabilité) a produit un grand nombre d'indécidables de Gödel, soit directement, soit indirectement.

A chaque fois qu'on démontre qu'un ensemble A n'est pas récursivement énumérable cela signifie que pour tout système formel S une instance au moins de la formule « n appartient à A » est un indécidable de Gödel pour S . A chaque fois en particulier qu'on établit qu'un problème $P(n)$ (dépendant d'un paramètre n) est indécidable, c'est-à-dire qu'il n'existe pas d'algorithme qui pour chaque donnée n puisse produire la solution à $P(n)$ (attention à ne pas confondre 'problème indécidable' qui est une notion absolue, avec 'énoncé indécidable de Gödel' qui est une notion relative à un système formel), cela implique que pour tout système formel correct S au moins une instance du problème est un indécidable de Gödel pour S . (Preuve: Soit S un système formel. Supposons que S donne tous les théorèmes $P(n) = R$ et qu'il ne se trompe jamais, c'est-à-dire supposons que S est correct et complet pour toutes les instances de $P(n)$. L'algorithme d'énumération des théorèmes de S qui existe par définition d'un système formel, fournit alors une procédure de résolution pour $P(n)$ ce qui contredit l'indécidabilité du problème $P(n)$).

Le premier exemple de problème indécidable est le problème de l'arrêt des machines de Turing, donc: pour tout système formel correct S , il existe au moins une machine de Turing bien précise (et qu'on peut expliciter) qui ne s'arrête pas et dont S ne peut pas démontrer le non-arrêt.

De la même façon pour tout système formel fixé correct S , il existe un couple de programmes d'ordinateurs ($Pr1$, $Pr2$) (par exemple écrits en Pascal) calculant la même fonction de N dans N et tel que la preuve de leur équivalence échappe à S (indécidabilité de l'équivalence des programmes).

De la même façon encore pour tout système formel correct S , il existe une configuration finie du jeu de la vie de Conway donnant lieu à une croissance indéfinie et qui échappe au pouvoir de démonstration de S . (Indécidabilité du devenir ultime d'une con-

figuration du jeu de la vie (Berlekamp Conway Guy 1982). Définition du jeu de la vie de Conway: un damier infini; certaines cases occupées par des pions; pour passer d'une configuration à la configuration suivante on applique les règles si une case est vide et possède trois voisins (parmi les huit possibles) alors elle devient occupée, si une case est occupée et possède deux ou trois voisins elle reste occupée, dans tous les autres cas la case sera vide dans la configuration suivante).

Des résultats d'indécidabilité en algèbre, en topologie donnent aussi naissance à des indécidables de Gödel pour tout système formel. Voir par exemple Davis (1977).

Hartmanis et Hopcroft ont tenté d'établir que le fameux problème $P=?NP$ était indépendant des axiomes de ZF, et bien qu'ils n'aient pas réussi, ils ont quand même montré que certaines versions relativisées $P^I=?NP^I$ sont indépendantes de ZF. En fait leur résultat est plus général: pour tout système formel S, ils construisent explicitement un oracle récursif I pour lequel $P^I=NP^I$ et $P^I\neq NP^I$ ne sont pas démontrables dans S (Hartmanis & Hopcroft 1976). Dans le même travail ils explicitent aussi la construction d'un algorithme tel qu'on peut l'ajouter à ZF sans introduire d'inconsistance qu'il fonctionne en temps n^2 ou qu'il fonctionne en temps 2^n .

Monsieur Importance: Cette indécidabilité est concrète puisqu'on peut réaliser mécaniquement des systèmes simulant des machines de Turing, ou le jeu de la vie de Conway, etc. L'indécidabilité des énoncés évoqués pour les machines de Turing ou le jeu de la vie de Conway, ou la complexité des algorithmes, signifie donc qu'aucune théorie physique (formalisée) ne sera jamais en mesure de tout dire du monde mécanique, et donc *a fortiori* du monde physique: quelle que soit la théorie (formelle) qu'on proposera il existera toujours un système physique fini (simulant une machine de Turing ou le jeu de la vie, etc.) dont le devenir ne pourra être prévu par la théorie.

Monsieur Insignifiance: Il s'agit du devenir 'à l'infini', or qui en physique s'intéresse au devenir de quoi que ce soit à

l'infini? Les énoncés indécidables produits à partir des problèmes indécidables ne nous concernent pas vraiment. Par exemple les machines de Turing dont le comportement échappe à $ZF + AC$ sont sans intérêt pour nous et nous ne les construirons jamais, ni ne nous y intéresserons jamais.

Monsieur Importance: Pas d'accord, je peux très raisonnablement écrire un programme qui produise une énumération des théorèmes de $ZF + AC$ et s'arrête s'il rencontre la démonstration de $0=1$. Le non-arrêt de ce programme (si $ZF + AC$ est consistant) est un indécidable de Gödel (pour $ZF + AC$). Ce programme est-il si déraisonnable?

Monsieur Insignifiance: Oui, car comme je viens de le dire, ce programme s'arrêtera à cause des limitations de ressource du système informatique que vous utiliserez, ou à cause de sa non fiabilité parfaite, et donc son comportement 'à l'infini' n'est pas un problème du monde physique. De toutes façons il est ridicule d'écrire un programme comme celui que vous évoquez car on sait bien que les techniques de démonstration automatique (et celle que vous utilisez en particulier) sont impuissantes à produire des énoncés mathématiques ayant un quelconque intérêt en temps raisonnable.

Monsieur Importance: Ce n'est pas exact et par exemple le second théorème de Gödel a été produit par un système de démonstration automatique. (Beeson 1988).

Monsieur Insignifiance: Justement cela ne prouve rien puisque le second théorème n'a pas d'intérêt mathématique authentique!!!

Monsieur Importance: Il n'empêche que les résultats d'indécidabilité sont utiles en informatique: il est arrivé bien des fois qu'on découvre qu'un problème dont on cherchait la solution, était indécidable (par exemple le problème de l'utilité d'un morceau de code dans un programme) et qu'on renonce à le résoudre complètement à cause de l'indécidabilité.

Monsieur Insignifiance: C'est très exagéré, car ce qui compte en informatique c'est la possibilité de résoudre pratiquement un problème. Si on accepte d'identifier «pratiquement» avec «en temps polynomial» ce qui est communément fait (mais est discutable Gurevich 1989) alors ce ne sont pas les questions d'indécidabilité qui sont importantes mais l'analyse d'algorithme.

2.4. Les indécidables de Post, Barzdin, Chaitin de la théorie des ensembles immunes

2.4.1. «Tous indécidables sauf un nombre fini»

Que la famille infinie des vérités de l'arithmétique élémentaire soit telle qu'un système formel en laisse nécessairement échapper certaines, ce n'est peut-être pas très grave. Mais il est certainement gênant de penser qu'il existe des familles d'énoncés mathématiques dont tout système formel (correct vis-à-vis d'eux, c'est-à-dire qui n'en démontre jamais de faux) ne peut en produire qu'un nombre fini. Pour de telles familles et quel que soit le système formel donné, tous les énoncés sont indécidables sauf quelques-uns.

L'existence de telles familles est liée à ce qu'on appelle les ensembles **immunes** (en termes précis: infinis et ne contenant aucun sous-ensemble infini récursivement énumérable). Les premiers exemples explicites de tels ensembles ont été produits dans les années 40 et 50 par Post et les théoriciens des fonctions récursives (voir par exemple Rogers 1967). Leurs définitions ne sont pas très compliquées mais aujourd'hui on en connaît d'autres assez extraordinaires venant de la théorie algorithmique de l'information.

2.4.2. Les immunes de la théorie algorithmique de l'information

La théorie algorithmique de l'information définit la complexité d'un objet fini par la taille du plus petit programme capable de

l'engendrer. Si on prend comme ordinateur de référence un ordinateur assez puissant (équivalent à une machine de Turing universelle) alors cette définition ne dépend du choix de l'ordinateur que par une constante. Autrement dit pour tout couple d'ordinateurs universels U, V , il existe une constante c_{UV} telle que pour tout objet fini (assimilé à une suite de digits 0 ou 1) on ait:

$$|K_U(s) - K_V(s)| \leq c_{UV}$$

Un ordinateur universel U étant fixé il est naturel de s'intéresser à la famille d'énoncés:

$$K_U(s) = n \quad \text{si } s \text{ suite finie de 0 ou de 1} \quad n \text{ entier}$$

Il a été établi par Barzdin et Chaitin (voir Zvonkin Levin 1970; Chaitin 1987a, 1987b) que dans tout système formel correct S , seul un nombre fini d'énoncés de cette famille pouvait être démontré. Donc pour tout système formel correct S , tous les énoncés sauf un nombre fini concernant la complexité algorithmique sont des indécidables de Gödel pour S . Jamais une théorie mathématique formalisée ne pourra dire plus qu'un nombre fini de choses sur la complexité algorithmique des objets finis. Quel que soit S , tous les énoncés de complexité, sauf un nombre fini sont des indécidables de Gödel pour S .

Sur ce sujet on pourra consulter Kolmogorov Uspenskii (1987); Chaitin (1987a, 1987b); van Lambalgen (1989); Delahaye (1989).

2.4.3. Le nombre oméga de Chaitin

Le nombre oméga de Chaitin (défini comme la probabilité pour qu'un ordinateur universel à programmes autodélimités s'arrête lorsqu'on lui donne un programme tiré au hasard) est un nombre réel qu'on peut écrire:

$$0, a_1 a_2 \dots a_n \dots \quad a_i = 0 \text{ ou } 1$$

Il est tel qu'un système formel correct S ne peut fournir qu'un nombre fini de digits a_n de ω , ce nombre de digits étant majoré par la complexité du système formel lui-même, plus une constante (indépendante du système formel S). Quel que soit S , tous les énoncés donnant les digits de ω sauf un nombre fini sont des énoncés indécidables de Gödel pour S . Voir Chaitin (1987b, 1987a), Bennett (1979), Delahaye (1989).

Ce nombre ω apparaît quelque peu magique puisque la connaissance de ses mille premiers digits donnerait un algorithme pour résoudre la plupart des conjectures mathématiques célèbres et en tout cas toutes celles de la forme: dans tel système formel S tel énoncé est-il démontrable? (par exemple: le grand théorème de Fermat est-il démontrable dans l'arithmétique de Peano?) pourvu que la conjecture s'exprime assez brièvement, ce qui est le cas de toutes les conjectures jugées intéressantes! (Bennett 1979). La raison de ce pouvoir du nombre ω provient de ce qu'il contient sous une forme compressée la solution du problème de l'arrêt de toutes les machines de Turing. Mais remarquons (Bennett 1988) que même si on connaissait les mille premiers digits de ω le temps nécessaire pour en extraire la solution des conjectures intéressantes serait tellement grand qu'on ne pourrait vraiment l'exploiter (l'information est trop compressée dans ω , le temps de décompression est donc très long).

Il semble avec ces résultats qu'on ait atteint un étonnant «concentré» d'indécidabilité.

Monsieur Importance: Que la complexité d'un objet fini soit intéressante est évident et donc le résultat sur l'impuissance d'un système formel à traiter plus qu'un nombre fini de cas du problème de la détermination de complexité donne bien des indécidables de Gödel intéressants et simples.

Monsieur Insignifiance: D'abord l'intérêt pratique de la théorie algorithmique de l'information peut être mis en doute, à cause de la fameuse constante c_{UV} qui, il faut le dire, peut devenir aussi grande qu'on veut lorsqu'on choisit mal sa machine universelle. Ensuite cette théorie est tellement ineffective que bien qu'elle indique qu'il y a beaucoup d'indécidables,

elle n'en fournit aucun explicitement. Le théorème de Gödel lui est constructif et permet de produire effectivement les indécidables dont il indique l'existence. Van Lambalgen dit même du résultat évoqué plus haut qu'à cause de cette ineffectivité, les résultats de Chaitin sont plutôt «a weak form rather than an extension, of the first incompleteness theorem» (van Lambalgen 1989: 1394).

Monsieur Importance: Le point de vue de van Lambalgen est lui-même discuté et nombreux sont ceux qui s'accordent à trouver les résultats d'indécidabilité produits par la théorie algorithmique de l'information comme de véritables aggravations des résultats d'indécidabilité de Gödel (Davis 1978; Tymoczko 1986; Delahaye 1989; Chaitin 1987a). Et puis, comment prétendre que cette théorie est sans intérêt pratique alors que justement les physiciens se sont mis à l'utiliser pour reformuler les fondements de la thermodynamique (Bennett 1982; Zurek 1989b; Delahaye 1991). L'entropie d'un micro-état d'après eux ne peut être définie rigoureusement par les conceptions classiques, fondées sur l'entropie statistique, car celle-ci ne peut définir l'entropie que pour des ensembles d'états. La solution proposée (qui en même temps permet de retirer un peu de subjectivité aux fondements de la thermodynamique) est d'utiliser la théorie algorithmique de l'information. Si cette théorie devait être adoptée, non seulement les indécidables produits par la théorie algorithmique de l'information ne pourraient plus être déclarés sans intérêt, mais de plus ils se trouveraient recevoir un sens physique.

Par ailleurs à propos des digits de ω , on peut faire une remarque: Si on considère que les indécidables des systèmes forts comme $ZF + AC$ sont sans intérêt mathématique alors les conjectures usuelles des mathématiques doivent avoir leur solution dans $ZF + AC$ (car on peut simplement les formuler par: C ou non-C résulte-t-il de $ZF + AC?$), et donc dans les premiers digits de ω . Soutenir que les indécidables ne concernent pas vraiment les mathématiques revient donc à enfermer les mathématiques dans les premiers digits de ω . C'est là un réductionnisme involontaire et extrême.

2.5. Les indécidables purifiés. Equations diophantiennes

«Impressive as these results are, the problems they concern are still very distant from the bread-and-butter problems of every day number theory ('everyday' over the last 300 years), because the number of variables is so much larger than is considered in such problems, and the complexities of the polynomials involved is so great (according to various measures of complexity). There is no evidence that these undecidability and incompleteness results have any relevance to the classic unsettled problems they have challenged generations of number theorists. ...» (Feferman 1987: 196).

2.5.1. Le dixième problème de Hilbert

Les problèmes les plus purs des mathématiques sont ceux de l'arithmétique élémentaire, et parmi eux sans doute les plus simples de tous sont ceux liés aux équations diophantiennes comme le fameux Grand Théorème de Fermat. C'est pourquoi les logiciens désireux de convaincre les mathématiciens que l'indécidabilité est bien présente au centre même des mathématiques ont recherché des problèmes indécidables exprimables en termes d'équations diophantiennes, c'est-à-dire de la forme $P(x_1, \dots, x_n) = 0$ avec P polynôme à coefficients entiers, ou en termes d'équations diophantiennes exponentielles, c'est-à-dire de la forme $P(x_1, \dots, x_n) = 0$ où P est une fonction n'utilisant que des entiers, des additions, des multiplications et des exponentiations.

A la suite de la résolution par Matijasevic (1970) du dixième problème de Hilbert qui conclut à l'indécidabilité de la résolution des équations diophantiennes (c'est-à-dire à l'inexistence d'un algorithme traitant de la résolubilité de toutes les équations diophantiennes) il est devenu possible de formuler des versions «purifiées» du théorème de Gödel dont nous donnons ici un exemple tiré de Davis Matijasevic Robinson (1976):

Soit A un système d'axiomes dans un langage comportant les symboles mathématiques $0, s, +, \cdot, <$ et satisfaisant:

- (a) A est consistant;
- (b) A est récursivement énumérable;
- (c) A est suffisamment puissant pour prouver tous les énoncés vrais de la forme:
 $a+b=c \quad a.b=c \quad a < b$ où a, b, c sont l'une des séquences de symboles 0, s0, ss0, sss0, ...

Alors il est possible de construire une équation diophantienne $F(x_1, \dots, x_n) = 0$ associée à A telle que $F=0$ n'ait pas de solutions en nombres entiers, mais telle que l'énoncé:
 non $\exists x_1 \dots \exists x_n F(x_1, \dots, x_n) = 0$ ne soit pas démontrable à partir de A.

2.5.2. Un système d'équations universel

A la suite des travaux sur le dixième problème de Hilbert on a aussi cherché à expliciter des équations insolubles (c'est-à-dire dont aucun système formel ne peut traiter tous les cas). Voici l'un des plus beaux exemples d'un tel système dû à Jones (1982):

Pour tout système formel correct S il existe des valeurs entières de x z u y telles que le système suivant n'a pas de solution et tel que S ne peut pas le démontrer:

$$\begin{aligned}
 &elg^2+A=(b-xy)q^2n \quad q=b^{560} \quad L+q^4=1+Lb^5 \\
 &T+2z=b^5 \quad l=u+tT \quad e=y+mT \quad b=2^w \quad n=q^{16} \\
 &r=[g+eq^3+lq^5+(2(e-zL)(1+xb^5+g)^4+Lb^5+Lb^5q^4)q^4][n^2-n] \\
 &+[q^3-bl+l+TLq^3+(b^5-2)q^5] [n^2-1] \quad Nn^2=(2r)!/(r!)^2
 \end{aligned}$$

Ce système d'équations est insoluble parce que:

$$x \in W_{\langle zu y \rangle} \Leftrightarrow \text{le système a des solutions.}$$

($\langle zu y \rangle$ désigne un entier qui dépend bijectivement et récursivement de z, u, y, et $W_{\langle zu y \rangle}$ désigne l'ensemble récursivement énumérable de numéro $\langle zu y \rangle$).

Tout problème du type «F résulte-t-il du système formel S» peut être traduit dans ce système d'équations en calculant le numéro de F et le numéro de l'ensemble r.é. qui correspond à l'ensemble des théorèmes de S. On en déduit d'ailleurs que tout théorème peut être «prouvé» grâce à ce système d'équations avec quelques multiplications et additions (en fait à l'aide d'un autre système du même genre Jones montre que toute démonstration est équivalente à au plus 100 additions et multiplications).

2.5.3. Une équation dont tous les cas sauf un nombre fini sont indécidables

La théorie algorithmique de l'information elle aussi a voulu formuler ses résultats sous forme spectaculaire et purifiée. Cela donne le résultat de Chaitin:

Il existe (et on peut effectivement l'écrire) une équation diophantienne exponentielle $P(m, x_1, \dots, x_n) = 0$ telle que tout système formel correct S, ne peut résoudre qu'un nombre fini (de l'ordre de la complexité algorithmique de S) de cas du problème: (m fixé, $P(m, x_1, \dots, x_n) = 0$ a-t-il un nombre fini de solutions)? Voir Chaitin 1987b, 1987c; Delahaye 1989; van Lambalgen 1989. Tous les cas du problème de l'équation de Chaitin sauf un nombre fini d'entre eux sont donc des indécidables de ZF + AC (ou de n'importe quel système formel).

Monsieur Importance: Il me semble qu'aucun doute n'est plus possible sur la simplicité des énoncés indécidables qu'on peut rencontrer en mathématiques, ni sur leur fréquence.

Monsieur Insignifiance: Sauf que des systèmes d'équations comme celui écrit plus haut, aucun mathématicien n'en a jamais étudié un seul. Ces systèmes sont toujours écrits en transformant un problème métamathématique sous forme arithmétique: ils sont faits exprès! ce ne sont que des problèmes métamathématiques déguisés! Quant à l'équation de Chaitin elle présente quelques inconvénients: d'abord on ne s'intéresse pas à l'existence d'une solution mais à l'existence d'une infinité de solutions (ce qui est déjà moins simple et fait

quitter le fini dont justement l'arithmétique est le langage); mais de plus un système formel S étant donné, il est impossible de connaître d'une façon effective ne serait-ce qu'un seul cas de l'équation qui échappe à S : on sait que tous sauf un nombre fini échappent à S mais on ne peut en connaître précisément aucun; enfin l'équation de Chaitin une fois écrite comporte 12000 variables et occupe 200 pages: si elle n'était la traduction arithmétique d'un autre problème, il est certain, que jamais aucun mathématicien ne s'y serait intéressé.

2.6. Les indécidables de la théorie des ensembles

«CH is thus an example of an "interesting" statement about sets which is independent of the axioms of ZFC. (...). These results (about abelian groups), and others too numerous to mention, show that many interesting Mathematical questions cannot be settled on the basis of the Zermelo-Fraenkel axioms for set theory» (Mac Lane 1986: 385).

«Ce système (ZF) correspond exactement aux besoins de tous les mathématiciens, excepté, bien sûr, les logiciens et aussi ceux que leur attitude philosophique empêche d'accepter les prémisses d'un tel système, c'est-à-dire les mathématiciens dits intuitionnistes ou constructivistes» (Dieudonné 1982: 17).

On sait que le système le plus simple donnant satisfaction aux mathématiciens est $ZF + AC$, et qu'il est bien sûr sujet aux théorèmes d'incomplétude de Gödel. Mais, et c'est plus intéressant, il est facile de produire directement des énoncés indécidables de $ZF + AC$, et cela sans avoir recours à un codage, c'est-à-dire sans avoir à utiliser la méthode donnée dans la démonstration des théorèmes d'incomplétude de Gödel.

L'hypothèse du continu HC par exemple est un tel énoncé. Mais même $ZF + AC + HC$ peut facilement être dépassé, par exemple avec un axiome de grand cardinal, comme l'axiome de l'existence de cardinaux inaccessibles.

Depuis la mise au point par Cohen de la technique du «forçage» (1966; Kunen 1980) de nombreux énoncés ont été

produits et démontrés indépendants de $ZF + AC$. En fait des hiérarchies infinies de systèmes plus forts que $ZF + AC$ ont été définies par les théoriciens des ensembles. Ces résultats confirment en un certain sens que les indécidables sont bien là, et que pour les produire il n'est pas nécessaire de recourir à du codage.

Parmi ces axiomes indépendants il faut distinguer ceux qui ont des effets arithmétiques de ceux qui n'en ont pas. Ceux qui n'ont pas d'effets arithmétiques produisent des systèmes formels qui ne sont pas plus forts puisqu'ils ne permettront pas de montrer $0=1$ si cela n'était pas possible sans eux.

Les axiomes AC , HC , $V=L$ par exemple n'ont pas d'effets arithmétiques (aucun nouvel énoncé d'arithmétique ne peut être produit par eux qui ne soit déjà produit par ZF , les ajouter ne donne pas un système formel plus fort). Par contre les axiomes d'existence de grands cardinaux eux, impliquent la consistance de ZF et donc ont des conséquences arithmétiques et même des conséquences «diophantiennes».

Ainsi que nous le verrons aussi pour les sous-systèmes de l'arithmétique du second ordre, il y a deux manières de compléter un système formel en lui ajoutant des indécidables: soit on n'en modifie pas la force arithmétique (extension conservative) soit on la modifie. Les extensions qui ne modifient pas la force arithmétique sont bénignes, les autres prennent des risques.

La force d'une extension peut être mesurée par ses effets arithmétiques: ce qui établirait définitivement que «l'infini compte pour le fini» serait la démonstration d'un énoncé arithmétique intéressant (et ne résultant pas simplement d'un codage) grâce à des axiomes «risqués» comme ceux d'existence de grands cardinaux. Ce serait la preuve que réellement même lorsqu'il ne s'agit que de l'arithmétique les indécidables ont du sens. Ce serait la preuve que lorsqu'il s'agit des entiers il faut réellement accepter d'utiliser plus que PA .

Ce pas, Friedman (1987) pense l'avoir franchi en montrant par l'utilisation d'un axiome d'existence de grands cardinaux (cardinaux de Mahlo), un énoncé combinatoire pur qui n'est pas la simple codification arithmétique d'un énoncé métamathématique. L'énoncé en question n'est pas suffisamment simple pour pouvoir être reproduit ici (nous verrons des énoncés combina-

toires analogues mais plus simples au paragraphe suivant) et Friedman reconnaît d'ailleurs qu'il faut sans doute encore simplifier ses énoncés combinatoires.

Feferman (1987) lui est sceptique, et se demande s'il n'y aurait pas une forme de cercle vicieux dans l'utilisation faite de ces énoncés par ceux qui y voient une preuve de l'importance des mathématiques de l'infini pour les mathématiques du fini: on prétend que des principes ensemblistes infinitistes forts peuvent produire des énoncés arithmétiques intéressants et vrais, mais en fait la seule raison qu'on a de croire que ces énoncés arithmétiques sont vrais est qu'ils sont équivalents aux énoncés ensemblistes qui permettent de les produire. On se convainc donc de l'importance des axiomes de grands cardinaux pour le fini en les supposant consistants, alors que justement ils ne sont importants que s'ils sont consistants.

Rien ne nous interdit de croire que les énoncés combinatoires que Friedman produit ne sont que le codage déguisé d'énoncés de consistance, comme les équations arithmétiques du paragraphe précédent l'étaient. Cela n'apparaissait pas au premier coup d'œil (voir le système de Jones), le progrès de Friedman serait simplement d'avoir su mieux déguiser le codage.

Certains spécialistes de la théorie des ensembles recherchent des axiomes qui pourraient entraîner HC ou non HC. Et bien que des avancées aient été réalisées ces dernières années (Maddy 1988) pour l'instant le problème de trouver des axiomes qui aient des conséquences sur HC (ce qui établirait que certains indécidables ont un rôle à jouer) reste irrésolu. A ce propos il faut remarquer que si vraiment les indécidables sont importants alors la recherche de nouveaux axiomes est dans l'essence des mathématiques mêmes, et donc le fait que très peu de mathématiciens s'adonnent à cette recherche confirme bien que dans sa pratique le mathématicien ordinaire ne croit pas à la véritable portée des indécidables.

Monsieur Importance: La multitude d'énoncés indécidables qui a été produite en théorie des ensembles montre bien que pour les systèmes les plus puissants le phénomène d'indécidabilité se produit encore, et cela sans avoir recours à

aucun codage. Dès qu'on cherche un peu on rencontre les indécidables de Gödel. C'est faire un drôle de pari que croire qu'ils ne peuvent jamais intervenir dans un vrai problème.

Monsieur Insignifiance: Les énoncés indécidables de $ZF + AC$ n'ont jamais servi jusqu'à présent en mathématiques à résoudre de vrais problèmes, seuls les logiciens les utilisent et s'y intéressent. Et d'ailleurs les mathématiciens (à l'exception de quelques logiciens en théorie des ensembles) ne recherchent pas de nouveaux axiomes susceptibles de compléter $ZF + AC$. Par leur pratique ils montrent donc que pour eux l'indécidabilité n'a jamais à être envisagée: sans doute attendent-ils pour changer d'attitude qu'on leur montre de vrais problèmes où leurs bons vieux axiomes sont insuffisants! La très grande énergie dépensée (depuis plus de vingt ans) à utiliser les méthodes de forçage n'a en définitive rien produit de très intéressant pour le mathématicien non logicien.

Monsieur Importance: Il faut laisser aux mathématiques le temps de se faire: les derniers résultats de Friedman s'ils n'établissent pas définitivement la preuve de l'importance mathématique des indécidables de $ZF + AC$, approchent le but de très près. La recherche de nouveaux axiomes ayant des conséquences sur les sous-ensembles de $\mathbb{R}$ et en particulier sur HC progresse, les succès même partiels de ces recherches prouvent indirectement et d'une autre façon que les indécidables ont du sens.

2.7. Les indécidables de Paris-Harrington et Friedman

«A new landmark in the expanding evolution of theory started by Gödel's 1931 is the discovery by Paris and Harrington in 1977 of a mathematically simple and interesting proposition, not depending on a numerical coding of notions from logic, which is undecidable» (Kleene 1986: 140).

«The instructions for constructing G are quite indirect; if they were carried out explicitly, the result would be a very long formal sentence,

involving many intercalated references to prime factorizations. Considered just as a sentence of arithmetic, G is not very interesting. Recently J. Paris has constructed a **true and relevant sentence of Peano arithmetic** which is not provable in that arithmetic» (Mac Lane 1986: 383).

«Jusqu'à ces dernières années, les seules arithmétiques non classiques et consistantes (si l'on suppose consistants les axiomes de Peano) étaient artificiellement obtenues au moyen des formules indécidées construites par Gödel, **formules atrocement compliquées et dépourvues d'intérêt mathématique**, leur existence seule étant importante pour les logiciens. Depuis 1976 sont apparues des formules arithmétisant le théorème de Ramsey» (Fraïssé 1982: 51).

«Although this result of Gödel is astonishing and remarkable, the criticism has often been raised that Gödel's concrete examples of unprovable finite combinatorial theorems are mathematically rather artificial, since they involve coding of logical syntax. This objection is to be taken seriously, for it is just conceivable that all mathematically meaningful finite combinatorial theorems are provable. The situation has become much clearer as a result of the 1977 work of Paris and Harrington. Namely, this work furnished a **completely transparent theorem of finite combinatorics** which is not provable in PA » (Simpson 1985: 88).

«It seems to me that as one passes from statements of the sort provided by Paris-Harrington and Kirby-Paris for PA , through those provided by Friedman for ATR_0 and $(\Pi^1_1-CA)_0$, on to those provided by Buchholz for $(\Pi^1_1-CA)+BI$ the connection with practice becomes more and more tedious. Viewed as an observer from the side-lines, there is a nagging feeling that the statements still have a 'cooked-up' look» (Feferman 1987: 201)

Depuis 1977 des énoncés purement combinatoires et ne résultant pas du codage direct d'énoncés métamathématiques ont été produits et démontrés être des indécidables par exemple pour PA (l'arithmétique de Peano). Ces nouveaux résultats sont unanimement considérés comme un pas en avant dans la démonstration de l'importance (de la gravité?) de l'indécidabilité.

Nous allons en donner deux exemples, à chaque fois avant d'arriver à l'énoncé indécidable nous en présentons des formes intermédiaires.

2.7.1. Le résultat de Paris-Harrington

Théorème de Ramsey infini (1928): Pour tout k et toute partition $C_1, C_2, \dots, C_r$ de l'ensemble $[N]^k$ des parties à k éléments de l'ensemble des entiers N , il existe un ensemble infini M de N tel que $[M]^k$ est inclus dans l'un des C_i .

Cas particulier $k=r=2$: tout graphe infini complet dont les arcs sont rouges ou verts est tel qu'il existe un sous-graphe infini dont tous les arcs sont de la même couleur.

Théorème de Ramsey fini: Pour tout k, r, m il existe un n tel que si N est un ensemble de n entiers et que $C_1, C_2, \dots, C_r$ est une partition de $[N]^k$ alors il existe M inclus dans N , ayant plus de m éléments tel que $[M]^k$ est inclus dans l'un des C_i .

Cas particulier $k=r=2$: Pour tout m il existe un n tel que tout graphe complet de plus de n nœuds dont les arcs sont rouges ou verts, contient un sous-graphe complet unicolore de plus de n nœuds.

Exemple: tout ensemble de six personnes contient trois personnes qui, soit se connaissent deux à deux, soit sont étrangères deux à deux: pour $n=3, m=6$ convient. (voir Graham Spencer 1990).

Théorème de Ramsey fini de Paris et Harrington: Pour tout k, r, m il existe un n tel que si N est un ensemble de n entiers et que $C_1, C_2, \dots, C_r$ est une partition de $[N]^k$ alors il existe M inclus dans N , ayant plus de m éléments tel que $[M]^k$ est inclus dans l'un des C_i et tel que $\text{cardinal}(M) \geq \text{minimum}(M)$.

Ce résultat purement combinatoire ne faisant référence ni à la logique, ni à aucune notion d'algorithme a été montré indépendamment des axiomes de l'arithmétique du premier ordre PA. (Paris Harrington 1977).

On peut remarquer que c'est un énoncé de la forme «pour tout n il existe m : P » avec P primitif récursif. Les indécidables arithmétiques que donnent le théorème de Gödel sont de la forme

«pour tout n : P » P primitif récursif. Chaque instance est prouvable (k, r, m fixés) c'est la rapidité de croissance de la fonction qui échappe à l'arithmétique de Peano (Smorynski 1982; Gallier 1991).

L'énoncé de Paris-Harrington équivaut à un énoncé de consistance: celui de la théorie obtenu en ajoutant consistance(PA) à PA , puis consistance(PA + consistance(PA)), etc jusqu'à l'ordinal ϵ_0 (Smorynski 1982).

2.7.2. Le résultat de Friedman sur le théorème de Kruskal

Théorème de Kruskal: si $T_1, T_2, \dots$ est une suite infinie d'arbres alors il existe i et j , $i < j$ tels que T_i est homéomorphe à une partie de T_j .

T_i est homéomorphe à une partie de T_j signifie qu'en enlevant des morceaux de T_j on peut obtenir un arbre exactement semblable à T_i .

Forme finie du Théorème de Kruskal de Friedman (FFF): Pour tout k il existe un n tel que si $T_1, \dots, T_n$ est une suite d'arbres vérifiant $\text{cardinal}(T_i) \leq k.i$, alors il existe i et j , $i < j \leq n$ tels que T_i est homéomorphe à une partie de T_j .

Le théorème de Kruskal entraîne directement la forme finie. La forme finie est un énoncé d'arithmétique et Friedman a montré qu'elle était un indécidable de l'arithmétique de Peano, et même qu'elle était indécidable pour des systèmes plus forts. (Simpson 1985; Smorynski 1982; Feferman 1987; Gallier 1991).

Cette fois-ci l'énoncé est encore très simple mais sa force est bien plus grande. En fait il équivaut à l'énoncé de consistance d'une théorie non prédictive, et certains y ont vu la réfutation de la philosophie prédictive des mathématiques:

«FFF is a concrete assertion about finite objects instantly understandable to any predicativist (predicatician?); but any proof of it must appeal to impredicative principles. In short FFF would have been meaningful to Poincaré, but he would not have been able to prove it or disprove it or accept any proof of it given to him» (Smorynski 1982: 187).

La simplicité de ces énoncés est grande mais n'empêche pas de remarquer que:

«None of the statements P (of Paris-Harrington or FFF) thus shown to be independant were previously considered *per se* in ordinary combinatorial work. It is a matter of pure speculation whether they would naturally have arisen for consideration in the normal course of such work» (Feferman 1987: 202).

Monsieur Importance: Cette fois-ci il n'y a plus de codage les énoncés sont simples et ont un contenu mathématique immédiat.

Monsieur Insignifiance: Oui, c'est mieux mais il faut bien constater que ces énoncés-là sont toujours des produits de logiciens et non pas de mathématiciens. Je me demande, comme pour les équations diophantiennes et la théorie des ensembles, si ces prétendus énoncés simples indécidables ne sont pas uniquement un progrès de la codification.

3. Le sens des résultats récents de la théorie de la preuve

«The necessary use of higher set theory in the mathematics of the finite has yet to be established. (...). Higher set theory is dispensable in scientifically applicable mathematics» (Feferman 1987: 153).

«Despite Gödel's theorem, one can give a finistic reduction for a substantial portion of infinistic mathematics including many of the best-known nonconstructive theorems» (Simpson 1988: 349).

«I would estimate that at least 85% of existing mathematics can be formalized with in WKL_0 or WKL_0^+ or stronger systems which are conservative over PRA with respect to Π_2 sentences». (Simpson 1988: 361).

Deux types de résultats récents en théorie de la preuve conduisent à des conclusions intéressant notre propos. Pour plus de dé-

tails et des définitions précises on se reportera à Simpson (1985a, 1985b, 1986, 1988); Friedman (1980, 1986); Feferman (1987, 1988); Harrington/Morley/Scedrov/Simpson (1985).

3.1. Réductibilité et extensions conservatives

Il y a d'une part les résultats de réductibilité et d'extensions conservatives: certains systèmes formels entre Z_1 (arithmétique du premier ordre, aussi notée PA) et Z_2 (arithmétique du second ordre) sont suffisamment forts pour contenir toutes les mathématiques ordinaires et pourtant sont des extensions conservatives vis-à-vis des formules Π_2 de PA ou PRA (arithmétique primitive récursive). Ces systèmes ne sont donc pas logiquement plus puissants que PA (ou PRA) et dans ces systèmes la consistance de PA (ou PRA) reste indécidable. Par contre ils sont mathématiquement plus puissants: concernant l'infini ils donnent plus de résultats. Cela prouve qu'en un certain sens la consistance de PA (ou PRA) n'est pas un «vrai énoncé des mathématiques ordinaires». Ces systèmes, si on les identifie aux mathématiques «réelles», prouveraient que les indécidables (de type consistance(S)) ne concernent pas les mathématiques. Concernant PA et l'idée que c'est un système complet pour l'arithmétique on lira Isaacson (1987, 1991).

Ici il faut donner quelques précisions sur les mathématiques ordinaires.

«By ordinary mathematics we mean roughly speaking, mainstream or non-set-theoretic mathematics, i.e. mathematics as it was before the abstract set theorists got hold of it (or perhaps: as it would have been if the abstract set theorists had never gotten hold of it). Thus ordinary mathematics includes number theory, geometry, calculus, differential equations, real and complex analysis, combinatorics, countable algebra, separable Banach spaces, computability theory, and topology of complete separable metric spaces. It does not include abstract functional analysis, abstract set theory, universal algebra, or general topology» (Simpson 1985: 461).

S. Feferman parle aussi de «scientifically applicable 20th century mathematics». Son point de vue sur la possibilité de développer toutes les mathématiques ordinaires dans des systèmes formels conservatifs vis-à-vis de PA est d'ailleurs très clair:

«It seems then that all of applicable classical and modern analysis can be developed in this $[VT\mu]$ conservative extension of PA» (Feferman 1987: 206).

La vision des mathématiques données par les résultats sur les extensions conservatives de PA est assez nouvelle: partant d'un système intuitivement clair et évident comme PA (qu'il n'est pas question de chercher à fonder!) deux grandes familles d'extensions seraient possibles:

- La famille des extensions comme Z2, ZF, ZF+{axiomes grands cardinaux}, etc. qui étendraient les énoncés élémentaires Π^0_1 de l'arithmétique qu'on peut démontrer et qui en particulier décideraient (positivement bien sûr) de la consistance de PA. Avec de tels systèmes l'indécis diminue mais la consistance devient de plus en plus douteuse. C'est la voie des mathématiques dangereuses, des mathématiques qui prennent le double risque de l'inconsistance et de l'ontologie illusoire.
- La famille des extensions comme Π^1_0 -CA₀ ou $VT\mu$ qui s'élèvent aussi au-delà de PA et autorisent bien plus que PA, mais conservativement vis-à-vis de PA et donc sans craindre l'inconsistance, et sans prendre d'engagement ontologique supplémentaire. Ces extensions qui d'une certaine façon refusent de croire au sérieux des indécidables, on le sait ne redonnent pas toutes les mathématiques, mais, et c'est cela qui est nouveau et fascinant, redonnent une part tellement grande des mathématiques qu'on peut y faire toutes les mathématiques applicables. Ces réalisations partielles du programme de Hilbert (Feferman 1988; Simpson 1988; Sieg 1988) sont certainement l'une des nouveautés les plus importantes de ces dernières années en logique.

Devant elles la tentation est grande peut-être de tracer un trait entre les mathématiques de l'illusion qui se complaisent dans l'indécidable et lui attribuent du sens, et les mathématiques ordinaires qui elles s'en passent très bien. Qu'on puisse faire de l'analyse, de l'algèbre dans des extensions conservatives de PA (c'est-à-dire dans des systèmes dont la consistance est aussi assurée que celle de PA) indique entre autres choses que le continu n'oblige pas à l'engagement ontologique qu'on a cru (et auquel les physiciens n'ont jamais souscrit).

En un sens donc au moins, aujourd'hui encore, on peut croire que les indécidables de Gödel sont sans importance mathématique.

3.2. Le programme des «reverse mathematics»

Il y a d'autre part les résultats de ce que Simpson appelle les «reverse mathematics». Une étude soigneuse des sous-systèmes de l'arithmétique du second ordre Z_2 montre que la plupart des grands résultats de l'analyse classique sont équivalents à des énoncés existentiels. La hiérarchie simplifiée des sous-systèmes de Z_2 mise en évidence par ce programme des «reverse mathematics» distingue quelques systèmes chacun obtenu à partir des autres par ajout d'un ou de plusieurs axiomes existentiels (Simpson 1985; Harrington/Morley/Scedrov/Simpson 1985).

La vue donnée par ces résultats est que certains indécidables des systèmes faibles (qu'il faut justement rechercher) sont exactement les énoncés nécessaires pour passer au niveau suivant. La classification des résultats mathématiques en fonction des hypothèses existentielles qu'ils utilisent (et auxquels ils sont équivalents bien souvent) met en valeur certains indécidables qui se trouvent alors placés au centre même des mathématiques.

De la même façon qu'on considère qu'identifier précisément où sert l'axiome du choix dans les résultats qu'on démontre, et qu'on cherche en arithmétique à purifier les méthodes utilisées, on doit admettre que c'est une véritable activité mathématique que de repérer les bons systèmes et les bons indécidables qui permettent de s'élever sans brusquerie inutile ($ZF + AC$ est trop

brusque, et donne beaucoup plus qu'il n'est nécessaire). La recherche des systèmes assez forts mais pas trop forts, la mise en évidence des bons axiomes, c'est-à-dire des bons indécidables est certainement une authentique activité mathématique.

3.3. Deux sortes au moins d'indécidables de Gödel

Les deux vues précédentes semblent se contredire complètement: l'une dénie tout sens aux indécidables, l'autre les met au centre de l'activité mathématique. En fait elles ne se contredisent pas si on distingue soigneusement:

- D'une part les indécidables de consistance, ceux construits par codification de la syntaxe, ceux «construits exprès» comme ceux de Paris-Harrington par exemple et qui même s'ils ressemblent aux énoncés que les mathématiciens étudient d'eux-mêmes n'en sont jamais (on découvre en fait que ces indécidables sont toujours des indécidables de consistance, voir par exemple Smorynski 1982).

- Et d'autre part certains énoncés existentiels (comme le lemme de König, le théorème que toute fonction continue sur $[0,1]$ atteint sa borne supérieure, le théorème de complétude de Gödel, l'axiome du choix) qui produisent des extensions conservatives des systèmes auxquels on les ajoute et donc n'engagent pas plus que les systèmes auxquels on les ajoute.

Monsieur Insignifiance: Puisqu'il y a des systèmes formels qui permettent raisonnablement de faire toutes les mathématiques ordinaires et qui sont des extensions conservatives de l'arithmétique de Peano cela signifie que les indécidables produits par les théorèmes d'incomplétude de Gödel (comme celui qui affirme la consistance de Peano ou d'autres systèmes encore plus forts) sont sans importance mathématique. De même qu'il y a des nombres transcendants de peu d'importance (ceux de Liouville) et des nombres transcendants intéressants (comme Π , e) il y aurait des indécidables «faciles»

mais sans intérêt (ceux donnés par les théorèmes de Gödel), et des indécidables profonds, respectueux des noyaux de base des mathématiques.

Monsieur Importance: Mais ce sont d'autres indécidables repérés par le programme des «reverse mathematics» qui aujourd'hui disent réellement ce qu'il en est de la pratique mathématique et donc même s'il ne s'agit pas d'indécidables de consistance (ou de diagonalisation-codification) ce sont quand même les indécidables de Gödel qui sont au centre des mathématiques.

Monsieur Insignifiance: Les indécidables de Gödel (au sens de la définition donnée dans l'introduction) OUI, mais pas les indécidables produits par les théorèmes d'incomplétude de Gödel.

Monsieur Importance: Nous sommes d'accord, il faut distinguer deux types d'indécidables. Je reste persuadé quant à moi que les deux types sont importants ...

Monsieur Insignifiance: ... et moi que les indécidables de consistance ne le sont pas, et que les autres sont des énoncés mathématiques utiles à identifier mais qu'on ne doit pas les considérer comme des indécidables authentiques puisqu'ils ne sont pas indécidables dans les extensions conservatives les plus intéressantes de PA.

4. Conclusion

Les progrès en logique mathématique ces dernières années ont eu deux conséquences remarquables.

(1) Ils ont conduit à proposer beaucoup de nouveaux indécidables. Mais malgré tout aucun énoncé mathématique formulé pour son propre intérêt et indépendant de la logique n'a été montré indécidable, dans un système fort.

(2) Ils ont conduit à des réalisations partielles ou modifiées du programme de Hilbert qui permettent de croire sérieusement qu'on peut faire toutes les mathématiques ordinaires dans des extensions conservatives de systèmes faibles.

Donc aujourd'hui encore en philosophie des mathématiques:

- On peut être **réductionniste** et croire que les indécidables sont sans importance pour les mathématiques applicables (ou ordinaires). Sur ce terrain se rencontrent des mathématiciens convaincus comme Dieudonné que ZF et quelques variantes de ZF constituent un bon cadre pour toutes les mathématiques, et des logiciens qui refusant la philosophie platonicienne considérée comme «moyenâgeuse» (Feferman 1987: 209) tentent d'élaborer des fondements faibles des mathématiques ou au moins des mathématiques ordinaires. Le «**réductionnisme de tranquillité**» du mathématicien «qui travaille», s'accorde avec le «**réductionnisme de perplexité**» du logicien qu'effraient les ontologies débordantes et incroyables. Le premier attend toujours qu'on lui démontre que le Grand Théorème de Fermat ou un «vrai» problème mathématique est indécidable vis à vis de Peano ou de ZF + AC. Le second qui vient de mener à bien des réalisations partielles du programme de Hilbert considère qu'il prouve par là que les indécidables de Gödel d'une certaine façon ne concernent pas les mathématiques ordinaires.
- On peut aussi voir dans les théorèmes d'incomplétude et toute nouvelle occurrence d'indécidables (et particulièrement celles produites par Paris-Harrington et Friedman) la **preuve de la faillite du réductionnisme**. Sur ce terrain le platonicien et l'intuitionniste se rejoignent. Le platonicien considère que les indécidables de l'arithmétique ne sont que la manifestation d'entités de plus hauts niveaux dont ils constituent la preuve indirecte d'existence, l'intuitionniste y voit simplement la manifestation de l'extensibilité indéfinie du concept d'entier et du concept de démonstration.

Aujourd'hui il est clair que c'est cet anti-réductionnisme qui prédomine largement en philosophie des mathématiques. Je pense qu'il a à répondre à bien des questions et qu'une brèche est ouverte dans son assurance trop facile qui se réjouit des indécidables («cooked-up» comme dit Feferman) de Paris-Harrington et Friedman. Cette brèche c'est la mise à jour d'extensions conservatives puissantes de PA qui permettent tout ce dont on a besoin pour la plus grande partie des mathématiques. Et cela sans prendre les engagements sur le «plus loin» et le «plus grand» que nous invitent à souscrire les anti-réductionnistes partisans du monde Cantorien, mais dont on pressent qu'ils défendent un univers trop beau pour être vrai. Et cela aussi sans avoir à se lier pieds et mains comme les autres anti-réductionnistes nous le proposent au nom de mathématiques de l'intuition et de la construction dont ils ont abusivement réduit l'univers.

Note: On remarquera que le «mathématicien qui travaille», par son scepticisme vis-à-vis des indécidables n'est pas du même côté que le platonicien: c'est que le «mathématicien qui travaille» n'a pas besoin vraiment de croire aux objets qu'il manipule: le platonisme déclaré de certains mathématiciens n'est jamais sérieux, seul l'est celui des logiciens en théorie des ensembles.

Monsieur Importance en cœur avec Monsieur Anti-réductionniste: Les indécidables sont partout y compris en arithmétique et en physique. Ils sont la preuve de l'irréductibilité du monde mathématique.

Monsieur Insignifiance en cœur avec Monsieur Réductionniste: Les recherches en théorie de la preuve ont permis de repérer des systèmes formels faibles mais suffisants pour faire toutes les mathématiques ordinaires applicables. Ces systèmes indiquent qu'il n'y a pas d'indécidables intéressants (pour les mathématiques ordinaires), et c'est sans doute ce qui explique l'échec des logiciens, qui pour l'instant n'ont jamais réussi à montrer qu'un énoncé mathématique authentique auquel les mathématiciens s'intéressent pour lui-même, est indécidable vis-à-vis de la théorie de base qui en permet l'expression.

Annexe: Quelques systèmes formels remarquables

Pour plus de détails voir Feferman (1977, 1987, 1988), Friedman (1980, 1986), Gallier (1991), Harrington/Morley/Scedrov/Simpson (1985), Kirby/Paris (1977), Simpson (1985a, 1985b, 1986, 1988), Sieg (1988).

Parmi les systèmes formels mis en évidence par le programme des «reverse mathematics» et de la recherche d'extensions conservatives de PA les systèmes suivants sont les plus importants:

(i) RCA_0 (Recursive Comprehension Axiom) ($RCA_0 = \{ \text{Axiomes de semi-anneau pour } N \} + \Delta^0_1\text{-CA} + \Sigma^0_1\text{-IA} \}$).

C'est un sous-système de l'arithmétique du second ordre très faible mais juste assez fort pour prouver l'existence de tous les ensembles récursifs. Il permet de développer les éléments de la théorie des fonctions continues de variables réelles. Dans RCA_0 on peut établir le théorème des intervalles emboîtés pour $[0,1]$ ainsi que le théorème des valeurs intermédiaires. C'est le système qui sert de base à la construction des autres (Simpson 1985).

(ii) WKL_0 est une extension de RCA_0 obtenue en ajoutant le Weak König Lemma WKL_0 qui affirme que tout sous-arbre infini de $2^{<\omega}$ comporte une branche infinie. Kirby et Paris (1977) ont montré que ce système était une extension conservative de PRA vis-à-vis des Π^0_2 . C'est aussi une extension conservative de RCA_0 vis-à-vis des Π^1_1 .

WKL_0 est un système qui permet de mener un peu plus loin la théorie des fonctions continues de variables réelles.

Dans RCA_0 on a l'équivalence des propriétés suivantes (Simpson 1985):

- (i) WKL_0
- (ii) Toute fonction continue sur $[0,1]$ est uniformément continue.
- (iii) Toute fonction continue sur $[0,1]$ atteint sa borne supérieure.
- (iv) Existence locale de solutions pour les équations différentielles ordinaires.

- (v) Tout anneau commutatif dénombrable possède un idéal premier.
- (vi) Théorème de complétude du calcul des prédicats du premier ordre de Gödel.

On voit donc qu'une part importante des mathématiques «ordinaires» peut être menée dans cette extension conservative de PRA, (c'est-à-dire dans cette extension dont la consistance est équivalente à celle de PRA).

(iii) ACA_0 (Arithmetical Comprehension Axiom) obtenu en ajoutant à RCA_0 l'axiome de compréhension pour toutes les formules Π^1_0 . Dans RCA_0 on a l'équivalence des propriétés suivantes:

- (i) Σ^0_1 -CA.
- (ii) Toute suite croissante de nombres réels possède une sous-suite convergente.
- (iii) Toute suite de Cauchy de nombres réels est convergente.
- (iv) Toute suite croissante de nombres réels possède une borne supérieure.
- (v) Toute suite monotone croissante de nombres réels est convergente.
- (vi) Tout espace vectoriel dénombrable possède une base.
- (vii) Tout anneau commutatif dénombrable possède un idéal maximal.

Ce système est une extension conservative de PA, et toute l'analyse prédictive (de Weyl) peut y être menée. Celle-ci se trouve donc finiment fondée.

(iv) ATR_0 (Arithmetical Transfinite Recursion) obtenue en ajoutant au système précédent un axiome qui équivaut à «deux ordinaux dénombrables sont toujours comparables».

La forme finie du Kruskal FFF n'y est pas démontrable, et en fait est équivalente à la 1-consistance de ATR_0 .

C'est un système considéré comme non prédicativiste. La forme finie du Théorème de Kruskal FFF serait donc hors d'atteinte des mathématiques prédicatives (Smorynski 1982).

(v) Π_1 -CA₀ (Π_1 -Comprehension Axiom).

FFF est prouvable dans ce système.

Mais Friedman a proposé une forme étendue de FFF qui n'est pas prouvable dans Π_1 -CA₀

Bien que fort, ce système est réductible à des systèmes intuitionnistes (et donc d'une certaine façon est finiment fondé) (Feferman 1987: 200).

(vi) ALPO contenu dans ZFC très riche conceptuellement et conservatif vis-à-vis de PA (Feferman 1987: 203).

(vii) VT μ de Feferman est aussi une extension conservative de PA (Feferman 1987).

*Laboratoire d'Informatique Fondamentale
Université des Sciences et Techniques de Lille
(U. A. CNRS 369 Bât M3)
F - 59655 VILLENEUVE D'ASCQ CEDEX*

Références bibliographiques

- BEESON, M.J. (1988). Computerizing mathematics: logic and computation. In: R. Herken (ed.), *The Universal Turing Machine: A Half-Century Survey*. Oxford: University Press, 191-225.
- BENNETT, C.H. (1979). *On Random and Hard-to-Describe Numbers*. IBM Research Report RC 7483 1-16-79. IBM Watson Research Center.

- BENNETT, C.H. (1982). The thermodynamics of computation - a review. *International Journal of Theoretical Physics*, Vol 21, n° 12, 905-940.
- BENNETT, C.H. (1988). Logical depth and physical complexity. In: R. Herken (ed.), *The Universal Turing Machine: A Half-Century Survey*. Oxford: Oxford University Press, 227-257.
- BERLEKAMP, E., CONWAY, J. & GUY, R. (1982). *Winning Ways for Your Mathematical Plays*. Orlando: Academic Press.
- BOOLOS, G.S. & JEFFREY, R.C. (1980). *Computability and Logic*. Cambridge: Cambridge University Press (second edition).
- BOURBAKI, N. (1969). *Eléments d'histoire des mathématiques*. Paris: Hermann.
- CHAITIN, G.J. (1987). Incompleteness theorems for random reals. *Advances in Applied Mathematics*, 8, 119-146. (Repris dans Chaitin 1987a).
- CHAITIN, G.J. (1987a.). *Information, Randomness and Incompleteness: Papers on Algorithmic Information Theory*. Singapore: World Scientific.
- CHAITIN, G.J. (1987b). *Algorithmic Information Theory. Cambridge Tracts in Theoretical Computer Science 1*. Cambridge: Cambridge University Press.
- COHEN, P.J. (1966). *Set Theory and the Continuum Hypothesis*. Amsterdam: Benjamin.
- DAVIS, M. (1977). Unsolvability problems. In: J. Barwise (ed.), *Handbook of Mathematical Logic*. Amsterdam: North-Holland Pub., 567-592.
- DAVIS, M. (1978). What is a computation. In: L. A. Steen (ed.), *Mathematics Today: Twelve Informal Essays*. Berlin: Springer Verlag, 241-267.
- DAVIS, M., MATIJASEVIC, Y. & ROBINSON, J. (1976). Hilbert's tenth problem. Diophantine equations: positive aspects of a negative solution. *Symposia in Pure Mathematics*. Vol 28. American Mathematical Society, 323-378.
- DELAHAYE, J.-P. (1989). Chaitin's equation: An extension of Gödel's theorem. *Notices of The American Mathematical Society*, 984-987.

- DELAHAYE, J.-P. (1991). Thermodynamique et informatique théorique: une nouvelle définition de l'entropie physique. *Pour la Science*, 17-20.
- DIEUDONNE, J. (1982). Mathématiques vides et significatives. In: *Penser les mathématiques. Séminaire de Philosophie et Mathématiques de l'Ecole Normale Supérieure*. Paris: Seuil.
- DIEUDONNE, J. (1987). *Pour l'honneur de l'esprit humain: Les mathématiques aujourd'hui*. Paris: Hachette.
- DRAKE, F.R. (1985). How recent work in mathematical logic relates to the foundation of mathematics. *The Mathematical Intelligencer*, Vol 7, n° 4, 27-35.
- FEFERMAN, S. (1977). Theories of finite types related to mathematical practice. In: J. Barwise (ed.), *Handbook of Mathematical Logic*. Amsterdam: North-Holland Pub., 913-971.
- FEFERMAN, S. (1987). Infinity in mathematics: is Cantor necessary? In: *Infinity in Science*. Roma: Istituto della Enciclopedia Italiana, 151-209.
- FEFERMAN, S. (1988). Hilbert's program relativized: proof-theoretical and foundational reduction. *The Journal of Symbolic Logic*, 53, 2, 364-384.
- FRAISSE, R. (1982). Les axiomatiques ne sont-elles qu'un jeu? In: *Penser les mathématiques. Séminaire de Philosophie et Mathématiques de l'Ecole Normale Supérieure*. Paris: Seuil.
- FRIEDMAN, H. (1980). A strong conservative extension of Peano arithmetic. In: J. Barwise et. al. (eds), *The Kleene Symposium*. Amsterdam: North-Holland Pub., 113-122.
- FRIEDMAN, H. (1986). Necessary uses of abstract set-theory in finite mathematics. *Advances in Mathematics*, LX, 92-122.
- GALLIER, J.H. (1991). What's so special about Kruskal's theorem and the ordinal Γ_0 . A survey of some results in proof theory. *Annals of Pure and Applied Logic*, 53, 199-260.
- GIRARD, J.-Y. (1987). *Proof Theory and Logical Complexity*. Napoli: Bibliopolis.
- GÖDEL, K. (1931). Über formal unentscheidbare Sätze des Principia Mathematica und verwandter Systeme I, *Monatshefte für Mathematik und Physik*, 38, 173-198. (Traduction anglaise: On formally undecidable proposition of

- Principia Mathematica and related systems I. *Kurt Gödel Collected Works, Vol. I (Publications 1929-1936)*. Oxford: Oxford University Press, 1986. (Traduction française: Sur les propositions formellement indécidables des Principia Mathematica et des systèmes apparentés. In: E. Nagel *et al.* (eds). *Le théorème de Gödel*. Paris: Seuil, 1989).
- GRAHAM, R. & SPENCER, J. (1990). La théorie de Ramsey. *Pour la Science*, 58-63.
- GUREVICH, Y. (1989). The challenger-solver game: variations on the theme of $P=?NP$. *Bulletin E.A.T.C.S.* (European Association for Theoretical Computer Science), n° 39, 112-121.
- HARRINGTON, L., MORLEY, M., SCEDROV, A. & SIMPSON, S.G. (1985). *Harvey Friedman's Research in the Foundations of Mathematics*. Amsterdam: North-Holland Pub.
- HARTMANIS, J. & HOPCROFT, J.E. (1976). *Independence Results in Computer Science*. Department of Computer Science Cornell. TR 76-296
- ISAACSON, D. (1987). Arithmetical truth and hidden higher-order concepts. *Logic Colloquium '85*. Edited by the Paris Logic Group. Amsterdam: Elsevier Science Pub., 147-169.
- ISAACSON, D. (1991). Some considerations on arithmetical truth and the ω -rule. In: M. Detlefsen (ed.), *Proof, Logic and Formalization*. London: Routledge.
- JONES, J.P. (1982). Universal diophantine equation. *The Journal of Symbolic Logic*, 47, 549-571.
- KIRBY, L. & PARIS, J. (1977). Initial segments of models of Peano's axioms. In: *Set-theory and Hierarchy Theory. Notes in Math*. Berlin: Springer Verlag, 211-226.
- KLEENE, S.C. (1986). Introductory note to 1930b, 1931, and 1932 (of Gödel). In: *Kurt Gödel Collected Works, Vol. I (Publications 1929-1936)*. Oxford: Oxford University Press.
- KOLATA, G. (1985). Does Gödel's theorem matter to mathematics? In: L. Harrington, M. Morley, A. Scedrov, S.G. Simpson (eds), *Harvey Friedman's Research in the Foundations of Mathematics*. Amsterdam: North-Holland Pub., 399-406.

- KOLMOGOROV, A. N. (1965). Three approaches for defining the concept of information quantity. *Information Transmission*. Vol. 1, 3-11.
- KOLMOGOROV, A. N. & USPENSKII, V. A. (1987). Algorithms and randomness. *SIAM Theory Probab. Appl.* Vol. 32, 389-412.
- KUNEN, K. (1980). *Set Theory: an Introduction to Independence Proofs*. Amsterdam: North-Holland Pub.
- LI, M. & VITANYI, P. M. B. (1990). Kolmogorov complexity and its applications. In: J. van Leeuwen (ed.), *Handbook of Theoretical Computer Science*. Amsterdam: Elsevier Science Pub., 187-254.
- MAC LANE, S. (1986). *Mathematics, Form and Function*. Berlin: Springer Verlag.
- MADDY, P. (1988). Believing the axioms (I and II). *The Journal of Symbolic Logic*, vol. LIII, 481-511.
- MADDY, P. (1992). Physicalistic platonism. In: A. Irvine (ed.), *Physicalism in Mathematics* (to appear).
- MATIJASEVIC, J. (1970). Enumerable sets are diophantine. *Soviet Math Doklady*, 11, 354-357.
- PARIS, J. & HARRINGTON, L. (1977). A mathematical incompleteness in Peano arithmetic. In: J. Barwise (ed.), *Handbook of Mathematical Logic*. Amsterdam: North-Holland, 1133-1142.
- ROGERS, H. (1967). *Theory of Recursive Function and Effective Computability*. McGraw-Hill, Serie in Higher Mathematics.
- ROSSER, J. B. (1936). Extensions of some theorems of Gödel and Church. *The Journal of Symbolic Logic*, 1, 87-91.
- SIEG, W. (1988). Hilbert's program sixty years later. *The Journal of Symbolic Logic*, 53, 2, 338-348.
- SIMPSON, S. G. (1985a). Reverse mathematics. In: A. Nerode & R. Shore (eds), *Recursion Theory; «Proc. Symp. Pure Maths.»*, XLII. Providence (R. I.): Amer. Mat. Soc., 461-471.

- SIMPSON, S.G. (1985b). Nonprovability of certain combinatorial properties of finite trees. In: L. Harrington, M. Morley, A. Scedrov, S.G. Simpson (eds), *Harvey Friedman's Research in the Foundations of Mathematics*. Amsterdam: North-Holland Pub.
- SIMPSON, S.G. (1986). Unprovable theorems and fast-growing functions. In: S.G. Simpson (ed.), *Logic and Combinatorics*. Providence (R. I.): Amer. Math. Soc.
- SIMPSON, S.G. (1988). Partial realizations of Hilbert's program. *The Journal of Symbolic Logic*, 53, 2., 349-363.
- SMORYNSKI, C. (1977). The incompleteness theorems. In: J. Barwise (ed.), *Handbook of Mathematical Logic*. Amsterdam: North-Holland Pub.
- SMORYNSKI, C. (1982). The varieties of aboreal experience. *Mathematical Intelligencer*, 4, 182-189.
- TYMOCZKO, T. (1986). *New Directions in the Philosophy of Mathematics*. Boston: Birkäuser.
- VAN LAMBALGEN, M. (1989). Algorithmic information theory. *The Journal of Symbolic Logic*, 54, 1389-1400.
- ZUREK, W.H. (1989). Algorithmic randomness and the physical entropy. *Physical Rev. A. Vol. 40*, 4731-4751.
- ZVONKIN, A.K. & LEVIN, L.A. (1970). The complexity of finite object and the development of the concepts of information and randomness by means of the theory of algorithms. *Russ. Math. Survey*, 25, 6, 83-124.

Travaux de logique

Liste des numéros parus

1. Denis Miéville: Introduction à la théorie des systèmes formels. Première partie. Septembre 1985 (épuisé).
2. Denis Miéville: Introduction à la théorie des systèmes formels. Deuxième partie. Janvier 1987 (épuisé).
3. James Gasser: La syllogistique d'Aristote à nos jours. Juin 1987.
4. Denis Miéville: Introduction à la théorie des systèmes formels. Première partie. Avril 1991 (réédition du n° 1).
5. Denis Miéville: Introduction à la théorie des systèmes formels. Deuxième partie. Avril 1991 (réédition du n° 2).
6. Denis Miéville: La négation, une étude logique. Mai 1991.

Ces publications peuvent être obtenues auprès du Centre de Recherches Sémiologiques au prix de Fr.s. 10.-.

Travaux du Centre de Recherches Sémiologiques

Liste des numéros parus

- *1. G. Vignaux: La nouvelle rhétorique. Revue critique et perspectives d'application. 1969-70.
- *2. G. Vignaux: L'argumentation antique: Aristote. Janvier 1970.
- *3. M.-J. Borel: Pour définir l'argumentation. 1969-70.
- *4. F. Bugniet: Remarques sur les notions d'assertion linguistiques et de proposition logique. Septembre 1970.
- *5. M.-J. Borel/G. Vignaux: L'étude de l'argumentation. Séminaire 1969-70.
- *6. G. Vignaux: L'argumentation: bibliographie sélective. Janvier 1971.
- *7. J.-B. Grize: Logique de l'argumentation et discours argumentatif. Mais 1971.
- *8. J.-L. Galay: La rhétorique du discours de philosophie systématique. Essais d'analyse. Mars 1971.
- *9. C. Morier: Charles Sanders Peirce et la sémiotique. Mars 1971.
- *10. G. Vignaux: L'argumentation et le résumé. Mars 1971.
- *11. C. Gillieron/C. Bonnet: Peut-on définir l'argumentation? Avril 1971.
- *12. J.-B. Grize: Notes sur l'ontologie et la méréologie de S. Lesniewski. Mars 1972.
- *13. M. Hirsbrunner/P. Fiala: Les limites d'une théorie saussurienne du discours et leurs effets dans la recherche sur l'argumentation. Avril 1972.
- *14. C. Gillieron/A.-M. Badonnel/J.-P. Iacazzi: Les recherches psychologiques et psycholinguistiques sur la négation et les relations d'opposition. Mai 1972.
- *15. J.-L. Galay: Esquisse pour une théorie figurale du discours. Septembre 1972.
- *16. Y. Oppel: Sémiotique littéraire, à propos de la coordination, répétition et opposition dans un texte littéraire. Mai 1973.
- *17. P. Fiala/C. Ridoux: Essai de pratique sémiotique. Juin 1973.
- *18. M. Hirsbrunner: Pour une critique de la sémiotique de Roland Barthes. Juillet 1973.
- *19. Y. Oppel: Colloque sur l'analyse du discours «Divergences et convergences». Février 1974.
- *20. (Collectif): Logique, argumentation, discours (LAD). Recherche. Septembre 1974.
- *21. (Collectif): Logique, argumentation, discours (LAD). Recherche. Septembre 1974.
- *22. A.-F. Schmid: Philosophie et sciences chez Henri Poincaré: lecture philosophique. Octobre 1974.

- *23. M.-J. Borel: Schématisation discursive et énonciation. Arguments théoriques et approche descriptive (LAD I). Octobre 1975.
- *24. A. Licitra: Les relations interpropositionnelles. Huit types d'après R. Longacre (LAD I). Octobre 1975.
- *25. (Collectif): Discours et structures sociales. Janvier 1977.
- *26. M. Ebel: Langage, histoire, action: les recherches de Jean Pierre Faye. Septembre 1975.
- *27. M. Ebel/P. Fiala: Recherches sur les discours xénophobes I. Juillet 1977.
- *28. M. Ebel/P. Fiala: Recherches sur les discours xénophobes II. Juillet 1977.
- *29. J.-B. Grize: Matériaux pour une logique naturelle (LAD I). Mai 1976.
- *30. D. Miéville/M.-J. Borel/A. Licitra: Discours et analogie (LAD II). Mai 1977.
- *31. J. Moeschler: Contribution linguistique à une sémiotique du cinéma. Mai 1977.
- *32. A. Lecomte: Paraphrase et thématization. Essais d'analyse logique. Décembre 1978.
- *33. (Collectif): Langue et discours I. Colloque Besançon-Neuchâtel, 2-4 octobre 1978. Mars 1978.
- *34. (Collectif): Langue et discours II. Colloque Besançon-Neuchâtel, 2-4 octobre 1978. Mars 1978.
- *35. P. Baldi/J. Moeschler: Comment contrôler le discours: interaction et réfutation dans le débat Giscard-Mitterrand (1974). Juillet 1979.
- *36. (Collectif): Quelques réflexions sur l'explication. Février 1980.
- *37. M. Sanchez-Mazas: Traduction arithmétique des graphes et des relations binaires et applications logiques et informatiques. Juin 1981.
- *38. (Collectif): Le discours explicatif I. Septembre 1981.
- *39. (Collectif): Le discours explicatif II. Septembre 1981.
- 40. C. Wülser: Actes de langage explicatifs. Février 1982.
- *41. (Collectif): Logique naturelle du raisonnement. Avril 1982.
- 42. (Collectif): Linguistique et sémiologie I. Colloque Besançon-Neuchâtel, 5-6 octobre 1981. Juillet 1982.
- 43. (Collectif): Linguistique et sémiologie II. Colloque Besançon-Neuchâtel, 5-6 octobre 1981. Juillet 1982.
- 44. (Collectif): Raisonnements et raisons. Avril 1983.
- 45. F. Albera: Problèmes de l'énonciation au cinéma. Février 1984.
- 46. (Collectif): Construction et transformations des objets du discours I. Colloque Besançon-Neuchâtel, 3-4 octobre 1983. Mars 1984.
- 47. (Collectif): Construction et transformations des objets du discours II. Colloque Besançon-Neuchâtel, 3-4 octobre 1983. Mars 1984.

- *48. (Collectif): Analyse de texte assistée par ordinateur. Utilisation du logiciel DEREDEC. Janvier 1985.
- *49. (Collectif): Problèmes et méthodes d'une analyse de texte articulant organisation cognitive, argumentation et représentations sociales. Juin 1985
- 50. (Collectif): Actes du colloque «Dialogisme et Polyphonie», 27/28 septembre 1985. Avril 1986.
- *51. (Collectif): Le discours descriptif. Du texte aux objets de connaissance I. Juillet 1986.
- *52. (Collectif): Le discours descriptif. Du texte aux objets de connaissance II. Juillet 1986.
- 53. (Collectif): La référence. Points de vue linguistique et logique. Mars 1987.
- 54. D. Apothéloz/J.-B. Grize: Langage, processus cognitif et genèse de la communication. Septembre 1987.
- 55. (Collectif): La schématisation descriptive. Types textuels, formes et fonctions discursives. Janvier 1988.
- 56. D. Miéville/R. Martin/A. Culioli/G.G. Granger/C. Gilliéron/G. Seel/J. Molino/L. Frey/J.-B. Grize: La négation sous divers aspects. Actes du colloque, Neuchâtel 22-23 octobre 1987. Septembre 1988.
- *57. D. Miéville/D. Apothéloz/P.-Y. Brandt/ G. Quiroz/J.-B. Grize: La négation. Contre-argumentation et contradiction. Septembre 1989.
- 58. M. Charolles: De l'art de nager et des différentes manières d'en parler. Septembre 1990.
- 59. D. Miéville/M.-J. Borel/J.-P. Desclés/J. Gasser/P.-Y. Brandt; D. Apothéloz/J. Moeschler/J. Jayez/M.F. Blès: La négation. Le rôle de la négation dans l'argumentation et le raisonnement. Actes du colloque, Neuchâtel 11-12 octobre 1990. Septembre 1991.
- 60. D. Miéville/D. Apothéloz/P.Y. Brandt: Les organisations raisonnées. Analyse de l'articulation de séquences discursives. Juin 1992.

Les titres précédés d'un astérisque sont épuisés.

Les publications disponibles peuvent être obtenues auprès du Centre de Recherches Sémiologiques au prix de Fr.s 10.-. Dès le n° 59 Fr.s. 15.-.

Couverture: Atelier Seth, Peseux

Impression: Zentralstelle der Studentenschaft der Universität Zürich

