Kurt GÖDEL (1906-1978) est le plus grand logicien du XX[e] siècle. Sa principale période d'activité se situe entre 1930 et 1950, d'où se détachent la cohérence *relative* de l'axiome du choix et de l'hypothèse du continu (1938) et bien entendu le très (trop ?) médiatisé *Théorème d'Incomplétude* en deux volets (1931) qui nous occupera ici.

Jean-Yves Girard

Dictionnaire d'histoire et philosophie des sciences, éd, Dominique Lecourt, 2006

I-CONTENU TECHNIQUE : Sous des hypothèses très lâches, un système *cohérent* **T** (i.e., non-contradictoire, en franglais « consistant » : qui ne prouve pas n'importe quoi) :

(i) contient un énoncé G vrai, *mais non prouvable* ;

(ii) parmi les G possibles, $Coh_{\mathbf{T}}$, la cohérence de **T** : « **T** ne démontre pas sa propre cohérence ».

À l'origine **T** = **PM** (les *Principia Mathematica* de Russell & Whitehead, système très daté) n'utilise que la possibilité d'écrire un petit bout d'arithmétique dans T. En particulier, elle s'applique à *toutes* les formalisations des mathématiques, par exemple le théorie des ensembles **ZF**. Un système formel non sujet au théorème de Gödel serait donc, au choix

a) Inexpressif : incapable d'énoncer sa propre cohérence, il n'a pas d'état d'âme à ce sujet !

b) Fautif : si **T** peut démontrer sa propre cohérence, mais alors prouvera des contrevérités, e.g., 3 x 7 = 22, qui ne sont pas forcément des contradictions genre $0 \neq 0$; mais quelle foi attacher alors à ce que « dit » **T ?**

c) Non-déductif : on peut passer outre au premier théorème, en « complétant le système », i.e., en « rajoutant des axiomes ». Cela dit, il n'y a *aucune* façon de savoir lesquels ajouter – ne pas ajouter, c'est ce que nous dit... l'incomplétude. Et les « systèmes non-monotones » ainsi obtenus ne sont pas déductifs : faute d'avoir une notion effective d'axiomes, ils n'ont pas de notion de démonstration. Ils sont tellement mal faits que, quand ils peuvent exprimer leur propre cohérence, ils sont alors condamnés à laréfuter.

d) Farfelu : on peut bricoler des logiques « para-consistantes », où il est disons, interdit de conclure $0 \neq 0$, ce qui résout tous les problèmes : ça revient à faire baisser le crime en ne prenant plus les déclarations de vol à la tire.

II-DÉMONSTRATION : Fastidieuse, sans être vraiment difficile, elle repose sur les points suivants :

(a) Le codage : on associe à chaque expression A du langage de **T** (énoncé, démonstration formelle) un entier $\ulcorner A \urcorner$ [$\ulcorner A \urcorner$ est le nombre de Gödel de l'expression A ; le nombre entier $\lceil A \rceil$ est une représentation de l'expression A] en commençant par les primitives, 11 pour « (», 13 pour «) », etc. Les propriétés de la syntaxe formelle deviennent des énoncés arithmétiques, ainsi « A est démontrable dans T » devient-il une propriété arithmétique $Thm_{\mathbf{T}}(\lceil A \rceil)$ du *nombre de Gödel* $\lceil A \rceil$ de l'énoncé A, cette propriété faisant elle-même partie du langage formel **T**.

(b) La réflexion : si l'énoncé A est démontrable dans **T**, ce fait lui-même, i.e., l'énoncé $Thm_{\mathbf{T}}(\lceil A \rceil)$, est démontrable dans **T**.

(c) La diagonalisation : l'énoncé à un paramètre A [signe de classe – Frege] qui dit « A (x) dans lequel on a fait x = $\lceil A \rceil$, n'est pas prouvable dans T » est représenté dans **T** par un énoncé $B(\lceil A \rceil)$. $B(\lceil B \rceil)$ signifie donc littéralement « je ne suis pas prouvable ».

(d) Le premier théorème : si $B(\lceil B \rceil)$ était démontrable, sa démonstrabilité, i.e., la négation de $B(\lceil B \rceil)$ serait démontrable et **T**, prouvant un énoncé et sa négation, incohérente. Donc, si **T** est cohérente, $B(\lceil B \rceil)$ n'est pas démontrable, i.e., est vraie.

(e) Le second théorème : une seconde réflexion permet de formaliser la démonstration du premier théorème dans **T**, qui démontre donc l'énoncé $Coh_{\mathbf{T}} \Rightarrow B(\lceil B \rceil)$ (qui veut dire « si **T** est cohérente, alors $B(\lceil B \rceil)$ est vraie »). Comme $B(\lceil B \rceil)$ n'est pas démontrable dans **T**, $Coh_{\mathbf{T}}$ ne l'est pas non plus.

III-NOUVEAUTÉ :

(a) La diagonalisation : c'est un procédé récurrent depuis Cantor (~1885) et Russell (1902), on identifie deux paramètres et on modifie le résultat, ainsi $f_n(m)$ devient $f_n(n) + 1$ (Cantor), $x \in y$ devient $x \notin x$ (Russell), avec chaque fois des résultats paradoxaux, e.g., l'ensemble X des ensembles qui ne s'appartiennent pas est tel que $X \in X \Leftrightarrow X \notin X$. La variante obtenue par Gödel échappe à la contradiction de justesse (vrai ≠ prouvable).

(b) Le codage : nettement plus original, en particulier, l'utilisation du « lemme du chinois » pour coder des listes d'entiers par des entiers, en fait technique de pointe reprise de la théorie du *corps de classes*, alors toute récente.

(c) La réflexion : l'école formaliste (à la suite de Hilbert 1862-1943) proposait une justification mathématique des mathématiques au moyen de démonstrations de *cohérence* formelle. Pour éviter tout soupçon d'« auto-amnistie », ces démonstrations devaient se faire dans un système « à part », de *métamathématiques*, qui jouerait le rôle d'un arbitre neutre et impartial, une sorte de « conseil constitutionnel ». Gödel innove en insérant les métamathématiques à l'intérieur des mathématiques, elles ne sont plus « à côté », elles en deviennent un morceau $\mathbf{T}_0 \subset \mathbf{T}$. Le résultat est d'autant plus impressionnant qu'il ne dépend pas du choix de $\mathbf{T}_0$ (qu'on aurait pu prendre « trop faible »), Gödel fait $\mathbf{T}_0 = \mathbf{T}$, ce qui est bien le choix le plus libéral, celui de l'auto-amnistie pure et simple : même ce choix ne fonctionne pas !

IV-SIGNIFICATION :

(a) L'infini : Hilbert se proposait d'éliminer, d'évacuer, l'infini au moyen d'un tour de passe-passe linguistique, bureaucratique : au lieu d'étudier l'infini, on étudiera le langage (qui parle) de l'infini. Le théorème nous montre qu'on recule pour mieux sauter : l'infini a été transféré de l'infini des objets à l'infini du langage, ce qui montre les limites du *finitisme* Hilbertien.

(b) Le récessif : Il s'agit de propriétés qu'on peut vérifier à un degré arbitraire de précision. Par exemple, le fait que $p^3 + q^3 \neq r^3$ quand p, q, r sont des entiers non-nuls, est de cette forme : si la propriété était fausse, on trouverait un contre-exemple (ce qui a d'ailleurs lieu pour le cas voisin $p^2 + q^2 \neq r^2$: prendre p = 3, q = 4, r = 5), et on peut lire la propriété comme « pour tous les choix (p, q, r) effectués, on a observé l'inégalité », autrement dit « jusqu'ici ça va ». Pour Hilbert les seuls énoncés mathématiques signifiants sont récessifs, idée reprise par Popper (propriétés « falsifiables », franglais pour « réfutable »). *A contrario*, les propriétés *expansives*, négations de récessives, sont susceptibles d'être vérifiées, en exhibant un exemple, ainsi la remarque que $3^2 + 4^2 = 5^2$.

(c) Démonstrabilité : le travail de poliorcétique de Gödel établit que tout énoncé expansif vrai est démontrable dans T : pour cela il suffit de remarquer qu'une vérification (genre $3^2 + 4^2 = 5^2$) peut se formaliser, c'est pour ça que **T** doit avoir un pouvoir expressif minimum. Dualement, pourvu que **T** soit cohérente, tout énoncé récessif démontrable dans T est vrai.

(d) Réflexion : le codage du langage, fait apparaître le côté expansif de la démonstrabilité (si je « tombe » sur une démonstration de A, alors A est démontrable), et *a contrario*, le côté récessif de la cohérence, qui dit qu'un A donné (par exemple $0 \neq 0$) n'est pas démontrable. Le théorème de
Gödel, c'est avant tout la stricte séparation entre récessif et expansif : l'énoncé de Gödel, la cohérence de **T**, sont récessifs, mais en aucune façon expansifs (sinon, ils seraient prouvables).

(e) Un à-peu-près instructif : on peut être tenté par un « théorème de Blair », les armes de destruction massives (ADM) existent, sans qu'on puisse le prouver. Cette analogie est trompeuse : la recherche sur le terrain et la pratique de la Question auraient forcément localisé les ADM au cas où. En d'autres termes l'existence d'ADM n'est pas récessif, mais expansif : si elles existent, on peut le prouver en les trouvant ; plutôt que l'incomplétude, c'est la paradoxe du menteur « je mens ». L'analogie fonctionne mieux avec Saddam : qui niait les ADM, se plaçant donc dans une position récessive : il avait raison, mais c'était invérifiable du fait des obstacles qu'il opposait aux contrôles.

(f) **L'extinction du Poppérisme :** l'ontologie de Hilbert/Popper a pour elle une indéniable tenue interne (en particulier, la réduction à la cohérence, propriété « falsifiable », donc dotée de signification). Mais elle n'arrive pas à convaincre, en particulier, parce qu'elle dénie tout sens au théorème de Gödel, énoncé par nature non-falsifiable, puisqu'il sortirait renforcé d'une réfutation, voir *infra*.

V-IMPACT :

(a) **Sur l'école formaliste :** à peu près nul, les « démonstrations » de cohérence n'ayant jamais cessé, parfois avec des idées originales (Gentzen 1908-1945), mais handicapées *a priori* par l'obligation d'une fondation de **T** sur « plus que **T** ». Cela dit, après 1931, l'école formaliste devient une secte, les mathématiciens se désintéressant de cet « intégrisme ».

(b) **Sur l'informatique :** d'abord la numérisation du langage, qui, même si l'idée est ancienne (Leibniz) apparaît pour la première fois de façon aboutie chez Gödel ; une nouveauté bien éventée, les ordinateurs codant mots, sons, images, etc. au moyen de listes de *bits* 0 et 1, i.e., d'entiers en numération binaire. Plus généralement, l'idée que « preuve formelle = calcul » se retrouve dans la conception des langages informatiques. D'ailleurs dès 1936, Turing (1903-1953) devait donner sa version du théorème de Gödel, pour le calcul : dans un langage de programmation universel il y a un programme qui diverge (argument de diagonalisation à la Cantor, $f(n) + 1$ est de la forme $f_N(n)$, on fait $n = n_N$, et donc $f_N(N)$ n'est pas défini, le calcul diverge) ; en fait un système formel comme **T** peut être vu comme un langage de programmation universel, c'est à ça que sert le codage.

(c) **Postérité dans la postérité :** c'est surtout dans l'informatique théorique que le formalisme se survit, en proposant des programmes d'*intelligence artificielle* particulièrement... imbéciles, en particulier en imaginant des machines répondant à toute question, OUI quand c'est oui, NON quand c'est non, J'SAIS PAS autrement. Ces programmes s'accompagnent de métaphores laborieuses, comme celle des 25 cocus de Bagdad qui tuent leurs femmes infidèles, qu'on va prendre à rebrousse-poil : disons qu'il n'y a qu'un cocu, un peu simplet, appelons-le W ; il sait que les autres ne le sont pas et qu'il y en a au moins un, mais ça ne l'inspire pas pour autant, et il épargne sa femme ; chaque autre, ne voyant qu'un cocu -W-, supposé en avoir vu un autre, conclut que c'est lui-même : subséquemment, les 24 moitiés fidèles sont assassinées ! Le grain de sable dans cette mécanique, c'est la différence entre « ne pas savoir » et « savoir que pas », produite ici par l'imbécillité de W, plus généralement par l'incomplétude et ses satellites (Turing, etc.). Voilà pourquoi les « logiques épistémiques » sont condamnées (sous peine d'incohérence) à ne pouvoir rien exprimer : elles se cantonnent sagement aux devinettes de cour de « récré ». Cet exemple illustre la différence entre constatation et raisonnement, entre calcul à l'ancienne, au moyen d'une table de fonctions, et calcul sur machine : dans le second cas on ne sait pas prévoir le temps de réponse.

(d) Non-déterminisme : si « preuve = calcul », il faut insister sur le côté *non-déterministe* : on peut *vérifier* une démonstration sur machine, on ne peut en aucune façon la trouver : c'est comme un labyrinthe où on peut aller où on veut, on ne se trompe jamais, sauf qu'on peut très bien ne pas en sortir. Parmi les 1001 formes équivalentes du théorème de Gödel, la moins spectaculaire et la plus utile, c'est qu'en aucune façon cet algorithme ne peut être rendu déterministe, i.e., qu'il n'y a pas de notion « dirigée » de démonstration formelle. Ceci est d'ailleurs en accord avec la technologie informatique : alors que les logiciels de *vérification* ont fait des progrès spectaculaires, les logiciels de *recherche* de démonstration sont toujours à la case départ.

VI-MÉTAPHORES :

(a) Remugles numérologiques : les nombres de Gödel comme nouvelle Kabbale. Mais alors les 128 codes ASCII du langage habituel (majuscules, minuscules, chiffres, ponctuations, diacritiques, 40 pour « (» , 41 pour «) », etc.) utilisés par l'ordinateur seraient tout aussi magiques. C'est ce que semble dire une certaine paralittérature : dans « les 9 milliards de noms de Dieu » (1953), Arthur C. Clarke – repris dans l'inénarrable *Matin des Magiciens* – propose une synthèse entre Kabbale et informatique, on invoque Dieu par ordinateur ; l'humour très involontaire de cette nouvelle tient aux progrès technologiques : de nos jours, la recherche proposée par l'auteur (tous les mots de 7 lettres) ne prendrait que quelques secondes, c'est vache pour Dieu !

(b) Métaphysiques : c'est surtout le transfert de signification du préfixe grec μέτα, « à côté », « après », en quelque chose plus proche de l'allemand « ur » ou du japonais « moto ». Pour prouver la cohérence, il faut passer dans un méta-système. C'est comme l'histoire des tortues : la Terre repose sur une tortue, Lui repose sur une autre tortue, et il y en jusqu'en bas... « Turtles all the way down ».

(c) L'impossibilité de se penser soi-même : en premier lieu, une réduction douteuse de la pensée à un exercice formel, que reste-il de nos philosophes, écrivains, poètes etc. à ce compte ? Deuxièmement un contresens littéral : les conditions (très techniques) sur le système **T** assurent précisément que **T** exprime suffisamment de choses sur elle-même. En fait sa propre cohérence est à peu près la seule propriété qui échappe à **T**. Mais du fait qu'on ne puisse pas revisser ses lunettes en les gardant sur le nez, faut-il en tirer des conséquences dramatiques, invoquer des méta-lunettes ?

(d) **L'auto-référence :** le cadre dans le cadre, le tableau dans le tableau, la pièce dans la pièce, le peintre qui expose sa méta-peinture dans une méta-galerie (aux toilettes), et ce monument de vulgarité, *Gödel-Escher-Bach*, entre Lewis Carroll et le *Matin des Magiciens* : tout ça pue l'arbitraire.

(e) **L'anti-totalitarisme :** pas de système universel (Régis Debray). La lecture est douteuse (système = système formel), mais, bizarrement, assez juste. En effet, l'idéologie formaliste participe du scientisme du début XX^e^ siècle, scientisme qui ne s'exprimait pas que dans la science, mais aussi dans des systèmes politiques dé-fi-ni-tifs. D'ailleurs l'expression de Hilbert (par ailleurs anti-nazi convaincu) « la solution finale du problème de la cohérence » nous rappelle la filiation scientiste des totalitarismes du XX^e^ siècle. Qu'on le veuille ou non, la fondation absolue, irréfragable, des mathématiques aurait conforté le totalitarisme ambiant.

(f) **Attention, danger :** le théorème de Gödel n'est pas la panacée contre l'obscurantisme scientiste, un peu de bon sens épistémologique suffit souvent. Un cas d'école : le « fameux » H. Simon fait « retrouver » par son ordinateur la troisième loi de Kepler, une relation entre deux séries de nombres a et t (demi-grands axes, périodes). Ici, Gödel ne nous aide en rien, car il s'agit de la mécanisation d'une partie très limitée des mathématiques. Par contre, on remarquera que la science cherche avant tout des questions, et ainsi l'apport de Kepler n'est-il pas d'avoir trouvé $a^3 = k.t^2$, mais d'avoir pensé à une relation possible (au demeurant facile à déterminer) entre a et t, au lieu de chercher le futur dans les astres comme son métier l'y invitait. Pour rester dans l'astrologie, les ordinateurs ont su « prédire » la mort d'une célèbre princesse à partir de Nostradamus ; mais seulement *après coup*, ce qui montre que le choix de la question reste crucial !

(g) **Synthèse :** en termes modernes, le formalisme, c'est la découverte d'une strate bureaucratique, informatique, dans les mathématiques, un peu comme l'état-civil est une strate de la vie. Si l'intérêt de cette découverte ne fait plus de doute, cf. l'informatique, la réduction *formaliste* de l'activité mathématique à une bureaucratie formelle est très douteuse, ce que disait déjà... Poincaré (1854-1912) dans *Science et Méthode.* Par rapport aux remarques de Poincaré, l'apport de Gödel, c'est ce pied de nez, la réfutation *formelle* d'une idéologie de la réduction formelle : qui se sert de l'épée périra par l'épée ! Le théorème doit être vu comme la fin définitive des *exagérations* – seulement des exagérations – formalistes, et reste un argument de poids contre le mécanisme et les réducteurs de tête en tout genre (de nos jours, les intelligents artificieux).

VII-RÉFUTATIONS :

(a) **Sous les fleurs :** sous la plume d'auteurs n'en comprenant que le mot-à-mot, le théorème de Gödel est assez abscons. Plus on le regarde de près, moins on « pige », et c'est cette attitude que privilégient les formalistes attardés qui veulent à tout prix neutraliser le résultat. À les en croire, le théorème de Gödel ne serait qu'un exercice extrêmement brillant de manipulation formelle. C'est par ce biais qu'on arrive à créer une image subliminale numérologique, à l'exact opposé de la signification du théorème de Gödel, qui conclut à l'échec d'une brutale réduction numérique/linguistique.

(b) **Réfutations formelles :** en provenance du milieu de l'intelligence artificielle, elles arrivent régulièrement, avec des variations saisonnières (l'année 2000 fut très faste, on se demande bien pourquoi). L'erreur est toujours la même, l'auteur utilise une inférence du genre « vrai ≠ prouvable », et il n'a guère de mal à conclure. Si ces gens avaient un peu de culture, on pourrait leur rétorquer que :

(**i**) Le théorème a été démontré, et tellement bien que la démonstration a même été vérifiée sur machine.

(**ii**) Une réfutation produirait donc une contradiction dans le système **T** de mathématiques où tout cela se passe.

(**iii**) Mais alors, **T** étant contradictoire, le théorème serait vrai par défaut : ce qui me tue me renforce !

VIII-LECTURES MODERNES :

(a) **Lecture académique :** Tarski (1901-1983) a « défini » la vérité : ainsi « A & B est vrai quand A est vrai *et* B est vrai », tout le reste à l'avenant, autrement dit A est vrai quand A est vérifié ; pour éviter les critiques des Béotiens, on dit que « *et* » est le « méta » de « & » : tout préexisterait donc à l'état de méta, comme dans le film « 2001 » (encore Arthur C. Clarke), où la méta-intelligence préexiste, ce qui est commode. Dans cette lecture « Thomiste », l'incomplétude, c'est le fossé entre vérité et prouvabilité, qu'on peut réduire au moyen de systèmes formels de plus en plus « forts », sans pouvoir jamais le combler. Mais gare au fagot : dans la Trinité Père = Sémantique (vrai/ faux), Fils = Syntaxe (ou Verbe), Saint-Esprit = Méta, l'incomplétude apparaît comme non consubstantialité du Fils, une sorte de... Nestorianisme logique.

(b) **Un peu d'imagination :** d'abord, remarquons qu'un manque n'est pas forcément le manque de quelque chose, le vrai manque c'est celui qui ne se compense pas, c'est le puzzle inachevé car inachevable. De ce point de vue, la lecture académique « vrai ≠ prouvable » tend à réduire le sortilège en nommant la pièce manquante : un déficit de vérité. C'est oublier l'échec de Tarski à donner une définition convaincante de la vérité : la laborieuse paraphrase citée plus haut s'écrit mathématiquement, mais n'a aucun intérêt épistémologique, c'est une pure tautologie cuisinée au méta. Un résultat mathématique doit être interprété, sinon on en arrive (Bergier, dans le *Matin des magiciens*) à dire que paradoxe de Banach-Tarski (une bizarrerie liée à l'*axiome du choix :* un cube découpé en cinq morceaux, qui, recomposés, donnent un cube double) montre la possibilité de changer de taille à volonté, comme dans *Alice !* Il faut admettre l'évidence, à savoir que la notion de vérité *ex nihilo* n'a aucun sens, et même si cette évidence nous amène à des questions sans réponse, c'est mieux que l'impasse de l'interprétation académique dominante. On pourrait voir l'énoncé de Gödel, la cohérence formelle, comme représentatifs des limbes de la signification, des énoncés qu'on peut écrire, mais dont la signification est assez diluée « je ne veux rien dire », en quelque sorte. Le paradoxe de Richard (1905) « le plus petit entier non-descriptible en moins de 15 mots », quoique moins précis que le théorème de Gödel, est la préfiguration de ce *diabolus in logica* qu'est l'énoncé de Gödel. Tout cela nous invite à revenir sur le Thomisme qui imprègne nos conceptions, en particulier celle de l'infini. La vérité n'a de sens que si l'infini du langage est pris au sens *actuel*, alors que l'expérience, en particulier informatique, penche vers une lecture *potentielle*, i.e., dynamique. Il y a un siècle, les tentatives de « potentialisation » se sont fourvoyées, principalement, en réduisant le potentiel à la collection de tous les possibles (modèles de Kripke), i.e., à l'actuel, un contresens absolu, les possibles alignés comme des papillons, des trophées de chasse. Il devrait être possible de reprendre cette discussion avec des idées plus originales... qu'on pourrait chercher, disons, dans le monde de la physique quantique.

[zGodel] [zGirard] [b-Girard]