Turing et Wittgenstein, Cambridge 1939

Patrick Goutefangea

Turing et Wittgenstein, Cambridge 1939

Patrick Goutefangea

Wittgenstein commence, en avril 1939, à Cambridge, un cours de philosophie sur « Les fondements des mathématiques ». On sait que son intention n'est pas de discuter les travaux menés sur la consistance des mathématiques, notamment dans le contexte, ouvert par Gödel, des « théorèmes d'incomplétude », mais, bien plutôt, de développer l'idée radicale que les mathématiques n'ont pas à être « fondées ».

Turing, rentré depuis quelques mois de Princeton où il a passé deux ans, est d'ores et déjà considéré comme l'un des principaux acteurs du débat sur les fondements, après la publication de son mémoire de 1936, *Théorie des nombres calculables, suivie d'une application au problème de la décision* (Girard, 1995), dans lequel il a mis au point la notion de « machine de Turing ». Il donne lui-même un cours sur la question des fondements, où il initie ses étudiants au formalisme de Hilbert et aux principaux résultats obtenus jusque là. L'un de ses amis, Alister Watson, qui vient de publier un article sur le même sujet, dans lequel il évoque la « machine de Turing », et qui l'a présenté formellement un peu plus tôt à Wittgenstein (Hodges, 1988, 122), le convainc de l'accompagner aux leçons proposées par ce dernier.

En réalité, les deux hommes se connaissent déjà. Wittgenstein est, depuis son premier séjour à Cambridge, un pilier du *Moral Science Club* où Turing donne une conférence dès décembre 1933. Turing ne manquera pas, du reste, en janvier 1937, d'envoyer à Wittgenstein un tiré à part de son article sur les « nombres calculables ».

Ce ne peut pas être un intérêt proprement mathématique qui pousse Turing à suivre le cours de Wittgenstein, dont, techniquement parlant, il n'a rien à apprendre ; on peut supposer, en revanche, qu'il y a été amené par une curiosité d'ordre philosophique. S'il a, en effet, démontré qu'une machine peut calculer exactement comme le fait un calculateur humain, que la « machine de Turing » a la même puissance logique qu'un homme qui calcule, il a également établi que cette machine ne peut pas remplacer le mathématicien. Plus généralement, la machine de Turing, qui fonctionne pourtant selon les « lois de la pensée », ne *pense* pas. Il y avait là quelque chose de suffisamment troublant pour que Turing s'intéresse à ce que Wittgenstein avait, éventuellement, à en dire, fût-ce de manière indirecte.

Il sera le seul mathématicien professionnel de l'assistance, et pas n'importe lequel, aussi va-t-il y incarner cela même à quoi s'attaque Wittgenstein, à savoir le « sens commun » des mathématiciens, le socle de convictions, en partie implicites, que tous les mathématiciens « en exercice » partagent peu ou prou. Ce « sens commun », aux yeux de Wittgenstein, a trouvé son expression la plus claire dans un article de G.H. Hardy, « Mathematical Proof » (Hardy, 1929) paru dans *Mind* en 1929. Turing lui-même connaît bien Hardy, qu'il a entendu à Cambridge et avec qui il a entretenu de bonnes relations professionnelles à Princeton.

Turing va donc devoir subir le harcèlement que Wittgenstein aurait réservé à Hardy si celui-ci avait assisté à son cours, harcèlement impitoyable, sans être inamical, puisque ce n'est pas au travail mathématique des mathématiciens que Wittgenstein entend s'attaquer, mais aux principes, aux idées vagues, imprécises ou implicites, qui constituent, à ses yeux, ce qu'on pourrait appeler, pour reprendre une formule ancienne d'Althusser, la « philosophie spontanée » des mathématiciens. Du reste, le travail de Turing lui-même vient de confirmer une conviction que Wittgenstein partage avec Hardy, à savoir qu'une machine qui calcule ne fait pas de mathématiques.

Les remarques de Wittgenstein vont souvent agacer Turing, ou le laisser, pour le moins, perplexe, mais elles lui fourniront aussi un éclairage pour s'orienter face au paradoxe apparent de cette machine qui ne « pense » pas alors même qu'elle manipule les « lois de la pensée » :

une machine logique, soutient Wittgenstein, *n'applique* pas de règles, contrairement à un homme. Qui plus est, Wittgenstein ouvrira à Turing une perspective pour aller plus loin dans la question même du rapport de la machine à la pensée : les règles s'établissent et s'appliquent collectivement, comme usages, et par l'apprentissage. C'est cette piste que Turing explorera dans l'article célèbre qu'il publiera en 1950 dans *Mind,* « Les ordinateurs et l'intelligence » (Girard, 1995) : une machine, imagine-t-il, peut « apprendre », elle peut avoir avec des humains – qui pensent - un *échange de paroles,* et s'insérer dans un système collectif d'usages. Cette inspiration wittgensteinienne de la dernière réflexion de Turing sur les machines est essentielle dans la perspective d'une clarification du débat qui a accompagné la naissance et le premier développement de l'intelligence artificielle, celui mené à propos de la thèse dite de « l'IA forte ».

Nous verrons tout d'abord en quoi consiste le « sens commun mathématicien » que Wittgenstein prend pour cible lors de ses leçons de 1939 et quelle curiosité peut pousser Turing à assister à celles-ci. Nous examinerons ensuite quelques aspects de la discussion qui eut lieu à cette occasion entre lui et Wittgenstein, avant d'essayer de dégager ce que l'inspiration de l'article de 1950, « Les ordinateurs et l'intelligence », a de wittgensteinien.

1 Le « sens commun » des mathématiciens, la « machine de Turing » et l'interrogation philosophique de Turing

1.1 G. H. Hardy, « Mathematical Proof »

Dès sa première leçon, Wittgenstein fait référence à G.H. Hardy et à l'article que celui-ci a publié dans *Mind* dix ans auparavant. La démarche de Hardy, lequel quitte, le temps de cet article, sa chaire de pur mathématicien pour s'intéresser à la philosophie des mathématiques, justifie, en un sens, celle de Wittgenstein, « amateur » qui prétend faire un cours sur les « fondements » des mathématiques : « Comment puis-je – moi ou qui que ce soit d'autre qui n'est pas mathématicien – traiter pareil sujet ? Quel droit un philosophe a-t-il de parler de mathématiques ? » (Wittgenstein, 1995, 1). A quoi il ne répond pas que ce droit est le même que celui qu'a pris Hardy en s'aventurant sur un terrain qui n'était pas le sien : Wittgenstein n'entend pas faire le mathématicien et discuter de la question proprement mathématique des fondements - « je n'y connais rien », dit-il - mais il n'entend pas non plus se situer exactement sur le terrain des philosophes spécialistes des mathématiques. Ce qui l'intéresse, ce sont les « confusions qui ont leur source dans les mots de notre langue ordinaire de tous les jours, comme "preuve", "nombre", "série", "ordre", etc. », et plus particulièrement « l'expression "les fondements des mathématiques" » en tant que telle (Wittgenstein, 1995, 2). Or, la « langue ordinaire de tous les jours », tel était aussi le lieu d'où Hardy, dans son article de 1929, choisissait explicitement de parler.

Hardy s'exprime, en effet, à la première personne : il expose son sentiment personnel de mathématicien en exercice. Son point de vue est celui que peut avoir sur ce qu'il fait l'homme de métier, qui pratique les mathématiques tous les jours, comme un menuisier fabrique des meubles, sans souci excessif des théories, éventuellement savantes, élaborées et défendues, ailleurs, à propos de cette pratique. Il explique ainsi qu'il y a, pour lui, des philosophies des mathématiques « sympathiques », au sens anglais du terme, « favorables » ou « bienveillantes », et d'autres qui ne le sont pas ; des philosophies, autrement dit, qui rencontrent sa pratique de mathématicien, qui énoncent les idées que lui-même peut avoir sur ce qu'il fait, et d'autres qu'il ne comprend pas ou qu'il ne peut accepter parce qu'elles lui paraissent aller à l'encontre des exigences de son activité professionnelle de mathématicien. On ajoutera qu'il ne suffit pas, bien entendu, pour qu'une philosophie des mathématiques l'intéresse, qu'elle lui soit « sympathique », il faut encore qu'elle soit « défendable », qu'elle soit bien argumentée, qu'elle soit « *tenable* » (Hardy, 1929, 4).

Hardy va donc s'efforcer de « décrire ce qui lui semble être la vision du sens commun mathématique » et de « rendre explicites quelques uns des préjugés instinctifs du

“mathématicien de la rue“ » (Hardy, 1929, 5)[1].

Deux points sont essentiels pour qu’une philosophie lui soit « sympathique ». Tout d’abord, elle ne doit pas imposer au mathématicien de limites à son investigation. C’est pourquoi, parmi les « écoles » entre lesquelles se partagent les philosophes des mathématiques, Hardy trouve particulièrement peu sympathique celle de Brouwer, l’intuitionnisme, lequel « rejette […] toute tentative de pousser l’analyse des mathématiques au-delà d’un certain point... » (Hardy, 1929, 10)[2], ce qu’il ne peut admettre, convaincu qu’il est que « le chemin le plus attrayant » ne peut pas être « celui qui passe le long de la haie, loin du taureau » (Hardy, 1929, 10)[3]. Le mathématicien « de la rue » n’aime pas les prescriptions qu’on cherche à lui imposer pour des raisons qui ne sont pas strictement mathématiques. Hardy rappelle, à ce propos, la réflexion de Hilbert affirmant que les difficultés logiques liées aux constructions mettant en jeu l’infini ne feront jamais renoncer aucun mathématicien au « paradis de Cantor » (Hardy, 1929, 6).

Ensuite, et surtout, pour être « sympathique », une philosophie des mathématiques doit admettre « d’une manière ou d’une autre, la validité immuable et inconditionnelle de la vérité mathématique » ; une philosophie sympathique est celle pour laquelle la « vérité mathématique » fait partie de la réalité objective (Hardy, 1929, 5)[4] et, comme telle, est indépendante de la connaissance que nous en avons. Connaître un théorème mathématique, c’est connaître « quelque chose », un « objet ». Hardy prend l’exemple du théorème de Goldbach, qui n’a pas encore été prouvé, mais que lui-même tend à croire vrai, et parle de son « sentiment invincible » que, si des mathématiciens « croient » au théorème de Goldbach, alors il y a bien quelque chose qui lui correspond, un « quelque chose » identique pour tous, qui sera encore le même après leur mort, et non seulement quand d’autres mathématiciens, éventuellement plus doués, auront démontré que le théorème est vrai, mais aussi dans le cas où ils auront démontré qu’il est faux – ces mathématiciens auront, certes, fait apparaître que le quelque chose en question n’est pas bien représenté par le théorème, mais il restera que ce quelque chose était « déjà-là » du temps de ceux qui croyaient à la validité du théorème et que la démonstration ne l’a pas changé. Cette « envie » (*craving*) d’une vérité mathématique indépendante est, pour Hardy, profondément liée à la pratique mathématique comme telle, elle est la mathématique « vécue » par le mathématicien : c’est, pour lui, une sorte de subjectif universel, qu’une philosophie des mathématiques qui lui serait véritablement sympathique - pour autant qu’une telle philosophie soit indispensable - devrait être à même d’expliquer et de légitimer.

Moyennant quoi, parmi les deux autres « écoles » de philosophie des mathématiques que reconnaît Hardy, hors l’intuitionnisme de Brouwer, à savoir celle des « logiciens », représentés par Russell, Whitehead et le Wittgenstein du *Tractatus*, et celle des formalistes, autrement dit l’école de Hilbert, c’est cette dernière qui lui apparaît la plus « sympathique ». Le formalisme, chez Hilbert, se donne en effet un appui sensible : les signes concrets utilisés par la démonstration, les formes dessinées sur la feuille de papier par le mathématicien, « objet concret, extra-logique, immédiatement présent à l’intuition et perçu indépendamment de toute pensée » (Hardy, 1929, 12)[5]. Cette réalité sensible du signe renvoie à une « réalité mathématique » : une démonstration de Hilbert se présente sous la forme d’une page de signes dessinés sur une feuille de papier, avec un crayon ; Or, note Hardy, si nous recopions cette démonstration, c’est-à-dire cette page de signes, sur une autre feuille, avec un autre

1 « I will therefore try to sketch what seems to be the view of mathematical common sense to make explicit a few of the instinctive prejudices of the “mathematician in the street“ ». C’est nous qui traduisons.

2 « Finitism rejects, first, all attempts to push the analysis of mathematics beyond a certain point »

3 « the prettiest path is that which passes on the side of the hedge away from the bull »

4 « It seems to me that no philosophy can possibly be sympathetic to a mathematician which does not admit, in one manner or another, the immutable and unconditional validity of mathematical truth. Mathematical theorems are true or false ; their truth or falsity is absolute and independent of our knowledge of them. In *some* sense, mathematical truth is part of the objective reality. »

5 « some concrete, extra-logical object, immediately present to intuition and perceived independently of all thought... »

crayon, nous serons bien toujours devant les *mêmes* mathématiques. Ce qui satisfait son désir que « la vérité mathématique » existe indépendamment de sa saisie par le mathématicien.

Par ailleurs, le formalisme de Hilbert, le recours à la « métamathématique » et le projet hilbertien de théorie de la démonstration, clarifient la question de la consistance des mathématiques en montrant qu'elle suppose un système d'axiomes et de règles en nombre fini permettant de décider si une formule quelconque est démontrable ou non. C'est, du reste, pour Hardy, une évidence qu'un tel système n'existe pas, en l'occurrence une évidence du sens commun du mathématicien : si un tel système existait pour les mathématiques, dit-il, une machine serait en mesure de résoudre tous les problèmes de mathématiques et il ne serait plus besoin de mathématiciens. En revanche, il est permis de penser que l'impossibilité qu'existe un théorème de consistance soit, quant à elle, démontrable, et c'est du reste ce que fera, en effet, après Gödel, un étudiant que Hardy croisera dans les couloirs de Trinity ou de King's College : Alan Turing.

C'est précisément cette certitude intuitive qui conduit Hardy à noter que la *preuve,* dans la sympathique philosophie de Hilbert, renvoie à deux sens différents : il y a la preuve telle qu'elle est établie formellement, à partir des axiomes et des règles du système, et qui correspond à la démonstration mathématique en tant que telle ; et la preuve telle qu'elle peut être établie, non plus dans les mathématiques, mais au niveau « métamathématique », là où les règles du jeu sont définies, et qui ne peut, par là-même, être une preuve formelle. « La structure de ces preuves n'est pas dictée par nos règles formelles ; en les construisant, nous sommes guidés, comme dans la vie ordinaire, par l'"intuition" et le sens commun. » (Hardy, 1929, 18)[6].

Hardy constate, en réalité, l'impossibilité, dans l'absolu, de la preuve mathématique telle que l'entend le sens commun. Cette preuve mathématique « parfaite », indiscutable, existe, certes, c'est la « démonstration » selon Hilbert : l'enchaînement réglé de certaines suites de signes en fonction d'autres suites de signes qui sont des axiomes et d'autres encore qui sont des règles. Dans ce cas, ce qui échappe au strictement formel est limité à la capacité de reproduire un signe à la suite d'un autre sur un support matériel quelconque. Cependant, la « démonstration » est décrite ainsi dans le cadre d'une théorie où elle ne peut pas être appliquée telle qu'elle est décrite, comme une procédure purement formelle, puisque c'est cette théorie qui définit ce qu'est le formel. Qu'en est-il donc de la preuve dans ce contexte « métamathématique » ? Hardy l'explique à l'aide d'une image, celle de « l'observateur » d'un paysage de montagne situé dans le lointain : « J'ai moi-même toujours pensé à un mathématicien comme, en premier lieu, à un *observateur*, un homme qui regarde une ligne de montagnes au loin et qui note ses observations. Son objet est simplement de distinguer clairement autant de sommets différents qu'il peut et de les indiquer aux autres. » (Hardy, 1929, 19)[7].

L'observateur distingue certains sommets très clairement, mais il en est d'autres qu'il ne fait qu'entrapercevoir ou deviner. En regardant le paysage avec attention, à partir de tel sommet A qu'il voit clairement, il finit par distinguer une ligne de crête, qui, partant de A le mène jusqu'à un sommet B dont il n'avait jusque là qu'une perception floue. Il peut désormais indiquer ce sommet aux personnes qui l'accompagnent en leur montrant le sommet A de départ, puis en leur faisant suivre à leur tour la ligne de crête jusqu'au sommet B. Ce sommet B étant bien repéré, l'observateur peut continuer et « découvrir » d'autres sommets. Dans certains cas, pourtant, la ligne de crête finit par s'évanouir dans le lointain et l'observateur en est réduit à faire des hypothèses : le parcours le mène à un sommet perdu dans les nuages ou qui se

6 « Secondly there are the proofs of the theorem of the metamathematics […] These are informal, unofficial, signficant proofs, in which we reflect on the meaning of every step.The structure of these proofs is not dictated by our formal rules ; in making them we are guided, as in ordinary life, by "intuition" and common sense »

7 « I have myself always thought of a mathematician as in the first instance an *observer*, a man who gazes at a distant range of mountains and notes down his observations. His object is simply to distinguish clearly and notify to others as many different peaks as he can. »

trouve derrière l'horizon.

Hardy reconnaît que cette comparaison est assez sommaire, mais il lui semble qu'elle fait bien ressortir que la « preuve », en définitive, consiste à *montrer*, à désigner un quelque chose qui se trouve déjà là et qui y sera encore lorsque l'observateur sera rentré chez lui[8]. Par ailleurs, elle illustre également le fait que la preuve n'est pas toujours immédiate, directe ; certaines preuves sont comme extirpées au moyen d'intuitions, de conjectures, d'hypothèses. Enfin, l'image souligne qu'il faut trouver les moyens de faire apercevoir aux autres les « lignes de crête », c'est-à-dire de repérer, parmi les intuitions, celles qui sont justes et peuvent porter plus loin. Bref, il faut *convaincre.*

Wittgenstein, dans son cours de 1939, s'arrêtera longuement sur cette distinction entre une preuve strictement formelle et une autre reposant sur l'intuition, sur une preuve « directe » et une autre qui doit « convaincre ».

Ainsi, le formalisme hilbertien, en réduisant aux signes sur la feuille blanche le rapport de la démonstration mathématique au réel, et en indiquant par là, comme en creux, l'existence d'une réalité mathématique - ce qui ne change pas d'une copie à une autre de la page de signes – est bien une philosophie « sympathique » au « sens commun » du mathématicien, au « mathématicien de la rue » que veut être Hardy. Cependant, le formalisme interdit de caractériser cette réalité mathématique. Il ne fait, en effet, aucune place à une « représentation », à une *correspondance* entre la forme et un « contenu » de celle-ci qui renverrait à une « réalité » mathématique autonome ; c'est là ce qui ne peut satisfaire Hardy dans l'école hilbertienne, puisqu'il n'y trouve pas d'appui pour sa croyance indéfectible à une proposition, comme il dit, « réelle ».

C'est, au contraire, à cette réalité mathématique que renvoie l'idée de Russell et du Wittgenstein du *Tractatus* selon laquelle une proposition est l'image d'un « fait » (Hardy, 1929, 23)[9]. Pourtant la théorie russellienne, la théorie de l'école des « logiciens », ne peut pas, elle non plus, satisfaire Hardy, lequel prend, pour la critiquer, l'exemple de la proposition : « George est le père d'Edward ». Cette proposition est l'image d'éléments objectifs : le locuteur qui prononce la proposition, George, Edward, le fait d'être père, celui d'être fils. L'image que constitue la proposition est celle de la relation entre ces éléments ; elle partage avec le « fait » une « forme ». Hardy remarque que cette forme est également partagée par toutes les occurrences possibles de la proposition : en toutes les langues, selon tous les types d'écriture, selon toutes les prononciations, on retrouve une forme qui est celle de la relation unissant George à Edward (Hardy, 1929, 22). Or, l'oeuvre de référence de l'école des logiciens, les *Principia Mathematica* de Russell et Whitehead, ne permet pas de rendre pleinement compte de cette forme, c'est-à-dire de la réalité postulée par Hardy pour les mathématiques. La proposition « Edward est le père de George » a, en effet, la même forme que la proposition « George est le père d'Edward » (Hardy, 1929, 23), ce qui indique que la forme partagée par toutes les occurrences de la proposition et par le fait doit être *plus* que celle qui est décrite par la structure de la proposition. Du point de vue du mathématicien Hardy, la forme de la proposition dans la démarche de Russell ne rend pas compte de la vérité ou de la fausseté de la proposition. Hardy espérait trouver cette critique développée chez Wittgenstein, lequel reproche à la proposition russellienne de ne pas expliquer pourquoi « il devrait être impossible de juger un non-sens » (Hardy, 1929, 23)[10] : dans la théorie de Russell, la manière dont une

8 « The analogy is a rough one, but I am sure that it is not altogether misleading. If we were to push it to its extreme we should be led to a rather paradoxical conclusion : that there is, strictly, no such thing as mathematical proof ; that we can, in the last analysis, do nothing but *point*... » (Hardy, 1929, 19)

9 « ...it is undisputed that there is something objective, what Russell and Wittgenstein call the "proposition as fact", which enters into any judgment. When we judge, we form a picture of the reality about which we are judging, a form of words, a set of marks or noises, which we suppose, rightly or wrongly, to afford an image of the facts. This is the "proposition as fact"... »

10 « It was therefore with great relief that I found that Wittgenstein rejects Russel's theory, for a variety of reasons of which the most convincing seems to be that Russell's theory leaves it entirely unexplained why it shoud be impossible to judge a *nonsense.* »

proposition rend compte du fait de la paternité de George par rapport à Edward, est telle qu'elle pourrait aussi bien rendre compte d'une paternité de George par rapport à n'importe quoi : la proposition « George est le père du bleu » serait, par exemple, tout à fait valide (Hardy, 1929, 24)[11].

Wittgenstein, poursuit Hardy, analyse le rapport entre la proposition et le fait dont cette proposition est l'image, en montrant que ce que partagent les deux niveaux, la forme qui leur est commune, renvoie aux « constantes logiques », c'est-à-dire aux opérateurs tels que « et », « ou », « non », etc. Cependant, ces éléments ne *représentent* pas et ils ne sont pas eux-mêmes représentés : ils sont, certes, présents dans le fait et dans la proposition, mais leur présence dans celle-ci n'est pas une *représentation* de leur présence dans celui-là ; ils sont communs aux deux niveaux et présents simultanément en eux. Telle est la forme proprement *logique* et Hardy note avec dépit qu'elle n'établit nullement une *correspondance* entre fait et proposition.

Ni le formalisme hilbertien, aussi « sympathique » soit-il, ni le logicisme de Russell et du Wittgenstein du *Tractatus*, ne satisfont donc à l'exigence du « mathématicien de la rue » d'une correspondance entre la proposition mathématique et une réalité mathématique autonome. « Pourtant, écrit Hardy en conclusion de son article, ma dernière remarque doit être que je suis toujours convaincu que c'est vrai. » (Hardy, 1929, 26)[12] ; Hardy n'en démord pas : une telle correspondance existe.

Il ne fait aucun doute que c'est en suivant l'inspiration de Hardy que Turing s'est lui-même initié au formalisme hilbertien, dont la théorie du signe – une figure tracée sur une feuille de papier – lui sert à formaliser sa machine ; sa propre démonstration, dans son mémoire de 1936 sur les « nombres calculables », est parfaitement conforme aux remarques de Hardy sur le signe, de même que sa réponse négative au « problème de la décision » conforte la conviction de Hardy à propos de la « machine mathématique » ; sa thèse américaine de 1938, enfin, sur les « logiques ordinales », se termine elle-même par une réflexion sur « l'intuition » et « l'ingéniosité » qui fait écho, sur un plan très général, à celle de Hardy dans son article de 1929.

1.2 La simulation du calcul humain par la « machine de Turing »

La « machine de Turing » décrite dans le mémoire de 1936 reproduit, en effet, ce qu'on peut appeler les conditions intuitives du calcul pour un individu humain, telles que ces conditions avaient été énoncées par Hilbert, à savoir sous la forme de la perception du signe mathématique comme figure matérielle.

« La condition préalable à l'application des raisonnements logiques et à la mise en œuvre des opérations logiques, expliquait Hilbert, c'est que quelque chose soit déjà donné à la représentation : à savoir certains objets concrets, extra-logiques, qui sont présents dans l'intuition en tant que données vécues immédiatement, préalablement à toute activité de pensée. Pour que le raisonnement logique soit doué de solidité, il faut que l'on puisse embrasser ces objets du regard de façon complète dans toutes leurs parties et que l'on puisse reconnaître par intuition immédiate, en même temps que ces objets eux-mêmes, comme des données qui ne se laissent plus réduire à quelque chose d'autre ou qui en tout cas n'ont pas besoin d'une telle réduction, comment ils se présentent, comment ils se distinguent les uns des autres, comment ils se suivent ou comment ils sont rangés les uns à côté des autres [...] En mathématiques en particulier, l'objet de notre examen ce sont les signes concrets eux-mêmes.... » (Hilbert, 1926).

Turing, reprenant la description hilbertienne, notait qu'un humain qui calcule va utiliser, par exemple, un « cahier divisé en cases comme un cahier d'écolier » (Girard, 1995, 77) et dessiner

11 « It would seem, on Russell's theory, that if you can judge that Edward is the father of George, you should be equally capable of judging that Edward id the father of blue. ».

12 « Yet my last remark must be that I am still convinced that it is true »

sur les cases des symboles, lesquels, compte tenu des limites de la perception humaine, seront nécessairement en nombre fini. C'est là ce qui définit ce qu'il appelle, reprenant les notions utilisées par Hilbert, la « reconnaissabilité immédiate » (*immediate recognisability*) (Girard, 1995, 79)[13].

Le comportement d'un être humain qui calcule, poursuit Turing, est par ailleurs déterminé à tout moment, non seulement par le ou les symboles qu'il observe, mais aussi par son « état mental » (*state of mind*) (Girard, 1995, 78) à ce moment. Pour faire comprendre ce qu'il entend par là, Turing évoque un homme en train d'effectuer un calcul et contraint de s'arrêter alors qu'il n'a pas terminé. Avant d'abandonner son travail, cet homme va rédiger une note qui décrit le stade auquel il est parvenu et qui lui permettra de reprendre le calcul là où il l'avait laissé ; ces notes sont naturellement, pour les mêmes raisons que les symboles, en nombre fini.

Un moment quelconque de la démarche humaine de calcul peut donc toujours être défini comme une « configuration », constituée des symboles observés à ce moment sur la partie du papier appréhendée par celui qui calcule - c'est-à-dire dans l'ensemble des cases observées simultanément - et de son « état mental » à ce moment. A partir de cette analyse de ce qu'est un « signe concret » et de ce qu'il implique, la démarche de celui qui calcule peut être décrite comme une méthode définie, une « procédure effective », une suite ordonnée d'« opérations simples » - celles auxquelles toutes les autres peuvent être ramenées. Pour qu'un homme puisse calculer, il doit pouvoir changer d'état mental, de symbole lu, et de cases observées ; en ce sens, toute opération implique la possibilité, lorsqu'on change d'état mental, de changer de symbole lu et, lorsqu'on passe d'un ensemble de cases observées à un autre, de changer d'état mental.

Bref, une « machine de Turing » simule les conditions intuitives et le comportement nécessaires à tout calcul, et tels qu'ils peuvent être définis pour un calculateur humain, à partir de la perception par celui-ci du signe mathématique comme figure matérielle.

1.3 L' « Entscheidungsproblem » et les « logiques ordinales »

Turing avait élaboré son concept de machine pour traiter l' « Entscheidungsproblem » de Hilbert, le fameux « problème de la décision », auquel il donnait une réponse négative : il n'existe pas d'algorithme permettant de décider si, dans un système formalisant l'arithmétique, un énoncé est démontrable ou pas. La démonstration de Turing reprenait le procédé de la diagonale de Cantor et établissait que toute « machine de Turing » peut être décrite par une formule exprimée dans son propre langage, pour laquelle elle ne dispose d'aucune procédure de dérivation[14].

Cependant, la machine de Turing est « universelle » : la formule qu'elle-même ne peut dériver peut être calculée par une autre machine – une machine de Turing peut simuler une autre machine de Turing. Il est donc permis d'imaginer qu'une machine parvenue à un état de « blocage » - c'est-à-dire, en réalité, à un état dans lequel elle « boucle » indéfiniment car elle ne dispose pas d'une règle lui permettant de « décider » - puisse être calculée par une autre machine qui dispose, elle, de la règle nécessaire ; cette même machine rencontrera à son tour un état de blocage, mais elle pourra être calculée par une autre machine disposant de la règle nécessaire, et ainsi de suite.

Dans sa thèse américaine, soutenue en 1938 à Princeton sous la direction d'Alonzo Church, *Logic based on Ordinals* (Turing, 2001), Turing formalisait l'enchaînement de machines traitant la formule décrivant une machine précédente à l'aide de la suite des nombres ordinaux

13 Hilbert évoque « l'intuition immédiate » et le fait pour les signes de devoir être « reconnaissables » (Hilbert, 1926).

14 Pour une approche rigoureuse et technique de la démarche de Turing, voir le texte de Jean-Yves Girard, *La machine de Turing : de la calculabilité à la complexité* (Girard, J.-Y., 1995).

(« premier », « deuxième », « troisième », etc.). Il imaginait une machine qui, lorsqu'elle est parvenue à un état de « blocage », dispose d'une instruction lui permettant de faire appel à un « oracle », lequel a pour fonction de lui fournir la règle supplémentaire nécessaire pour sortir de cet état. Par là, la machine devenait une autre machine d'un niveau supérieur à la précédente.

Turing démontrait ensuite que cette « machine à oracle » - intuitivement, la suite des machines « s'imbriquant » les unes dans les autres – est elle-même une « machine universelle de Turing », vouée à rencontrer une formule, celle qui la représente, qu'elle ne peut dériver ; la « machine à oracle » est une suite de machines qui, fût-elle représentée sous la forme du transfini, est sans fin. La « machine à oracle », en réalité ne sait faire qu'une chose : passer en revue toutes les règles dont elle dispose jusqu'à trouver celle qui la conduit à aller chercher auprès d'un « oracle » la règle qui lui manque, et recommencer.

Turing concluait sa thèse par une réflexion générale sur l'activité du mathématicien : celui-ci a recours, dans son travail, a deux « facultés » : « l'intuition » et « l'ingéniosité ». L'intuition – un « jugement spontané qui n'est pas le résultat d'un raisonnement conscient » (Turing, 2001, 215)[15] - est une sorte de pari sur un résultat possible, c'est-à-dire sur une certaine proposition ; l'ingéniosité permet au mathématicien d'établir la vérité ou la fausseté de cette proposition en imaginant des combinaisons de propositions aboutissant à celle recherchée. Intuition et ingéniosité sont largement aidées par l'introduction des logiques formelles, lesquelles établissent des règles de construction des combinaisons possibles, c'est-à-dire des règles d'inférence : l'ingéniosité consiste alors à choisir parmi les possibilités ordonnées par le système de règles.

Avant Gödel et l'établissement de ses théorèmes d'incomplétude, il n'était pas interdit d'espérer pouvoir remplacer l'intuition et l'ingéniosité par la « patience », c'est-à-dire par l'inventaire systématique des combinaisons possibles. A supposer que celles-ci soient en nombre fini, un tel inventaire complet est précisément ce qu'une machine serait en mesure de faire. Or, Turing, par la formalisation de la machine et la réponse négative qu'il apporte au problème de la décision, par sa démonstration, enfin, concernant les « machines à oracle », établit que la « patience » mécanique ne peut pas remplacer l'ingéniosité. Bref, comme le savait intuitivement déjà Hardy, la machine ne peut pas prendre la place du mathématicien.

C'est celui-ci, en effet, qui fournit à la machine l'oracle qu'elle va consulter ; c'est lui qui invente et construit la machine et ses oracles. De sorte que, sous un angle qui n'est plus celui de la démonstration formelle, tout se passe comme si, alors qu'on venait d'établir, tout en même temps, que la capacité de calcul de la machine était aussi étendue que celle de l'homme et que celui-ci se heurtait, lorsqu'il calcule, à la même limite que la machine, on devait pourtant constater que la machine n'égalait pas le calculateur humain et que ce dernier échappait, d'une manière mystérieuse, à la limite même qu'il partageait avec elle. La machine de Turing met en jeu des « lois » logiques, des lois que le sens commun, y compris celui des logiciens et des mathématiciens, tend à considérer, à sa manière de sens commun, comme des « lois de la pensée », mais elle ne *pense* pas. Bref, la machine de Turing n'épuise pas les possibilités de son modèle, l'homme qui calcule, lequel, en l'occurrence sous les traits de Turing lui-même, est renvoyé, par là, à l'espèce de trouble, de *sentiment* d'incomplétude qui accompagne la réflexion sur les « fondements » des mathématiques, au *désir* de fondement qu'éprouve le mathématicien dès lors que la consistance des mathématiques paraît menacée. C'est la curiosité, plus philosophique que mathématique, ainsi alimentée, qui va pousser Turing à se rendre, au début de 1939, en compagnie d'Alister Watson, au cours dispensé par Wittgenstein.

Celui-ci propose, de fait, une réponse aux interrogations de Turing : la machine, affirme-t-il,

15 « The activity of the intuition consists in making spontaneous judgments which are not the result of conscious trains of reasoning… »

contrairement à son constructeur humain, *n'applique* pas de règles ; elle n'est rien de plus qu'une expression matérielle des règles inventées par l'homme qui calcule. Voilà pourquoi une « machine universelle de Turing », qui « sait » manipuler les mêmes suites de symboles que les humains, qui « sait » calculer les mêmes classes de fonctions que des mathématiciens tels que G.H. Hardy ou Turing eux-mêmes, ne « pense » pas, contrairement à ceux-ci, et contrairement à tous les humains manipulateurs de symboles.

2 La discussion entre Wittgenstein et Turing

Turing est, rappelons-le, parmi la petite quinzaine de personnes qui assiste au cours de Wittgenstein, le seul mathématicien « professionnel » ; par ailleurs, il n'est pas n'importe quel mathématicien puisqu'il travaille sur ce qui constitue l'objet même de la réflexion critique de Wittgenstein ; enfin, il vient d'établir formellement l'équivalence, jusque là intuitive, d'une procédure effective de calcul et d'une procédure mécanique. Aux yeux de Wittgenstein, Turing représente donc tout à la fois le « sens commun » du mathématicien et la démarche spécifique consistant à rechercher, pour les mathématiques, des « fondements » dont il soutient qu'elles se passent très bien. C'est pourquoi Turing sera pour lui un interlocuteur privilégié, celui qui devra, autrement dit, supporter son pilonnage incessant et déroutant.

Pour le « mathématicien de la rue », comme, du reste, pour « l'homme de la rue » en général, les mathématiques sont le paradigme de la nécessité ; le lieu où celle-ci se donne à voir comme telle. Et la métaphore de la machine est, ici, immédiatement présente ; le calcul, selon le « sens commun », se définit comme une suite d'étapes enchaînées automatiquement, mécaniquement, les unes après les autres, dans un ordre contraint, sans que l'agent qui mène le calcul dispose d'aucune liberté, comme si le processus s'imposait à lui sous la forme d'un « donné ». Ce sont là, aux yeux de Wittgenstein, des illusions qu'il entend dissiper : il va s'efforcer de démontrer que la nécessité mathématique, repose, en vérité, sur un acte de *décision*.

2.1 La question de l'infini

Wittgenstein sollicite Turing pour un premier échange sur l'infini. Combien de nombres Turing a-t-il appris à écrire ? Étrange question, à laquelle l'intéressé répond : $\aleph_0$. Wittgenstein prend acte, puis développe. On pouvait s'attendre à plusieurs types de réponses : par exemple, Turing aurait pu indiquer combien de nombres il avait écrit jusque là, ou encore combien de nombres il pensait écrire pendant toute son existence. Dans tous les cas, ce nombre, même colossal, serait fini. La question, cependant, n'est pas : « combien avez-vous écrit de nombres ? », ni « combien écrirez-vous de nombres dans toute votre vie ? », mais « combien de nombres avez-vous *appris* à écrire ? ». Et, en disant $\aleph_0$, Turing a répondu correctement : $\aleph_0$, le cardinal de l'ensemble des entiers naturels, est, en effet, quelque chose comme le plus grand nombre possible de nombres ; une manière de dire que Turing sait écrire tous les nombres qui peuvent être écrits, même s'il est clair qu'il ne les écrira jamais tous[16].

Turing ne s'est donc pas trompé. Pourtant, quelle différence y a-t-il entre savoir écrire une suite de nombres qui s'arrête au moment présent, ou qui s'arrête à notre mort, et savoir écrire une suite qui ne s'arrête jamais ? Quelle différence y a-t-il entre une suite finie de nombres et une suite infinie ? Dans le cas des entiers naturels, par exemple, en quoi la suite qui va jusqu'à 100 000 se différencie-t-elle de celle qui va « jusqu'à » $\aleph_0$? Ne sont-elles pas construites de la même façon, à l'aide de la « fonction successeur », laquelle renvoie à la suite infinie des

16 « Je n'ai pas demandé "combien de nombres y a-t-il ?". Cela est de la plus haute importance. J'ai posé une question portant sur un être humain, à savoir : "combien de nombres avez-vous appris à écrire ? Turing a répondu "$\aleph_0$", ce que j'ai accordé. En l'accordant, je voulais dire que c'est ainsi que l'on emploie le nombre $\aleph_0$. » (Wittgenstein, 1995, 21)

entiers naturels ? Cette construction d'une suite de nombres, montre Wittgenstein, repose sur l'application d'une règle consistant, non pas seulement à « ajouter 1 », mais à admettre qu'il est toujours possible, à chaque pas d' « ajouter 1 ». La règle de construction de la suite des entiers naturels établit la possibilité de refaire à chaque pas ce qui a été fait au pas précédent.

Cependant, que cela soit possible ne signifie pas que nous allons *nécessairement* le faire. Qu'est-ce qui nous empêche de nous arrêter après 100 000 ? Qu'est-ce qui nous empêche d'appliquer, après 100 000, une autre règle, une autre manière de faire, « ajouter 2 », par exemple ? Si nous continuons, ce sera, en vérité, parce que nous l'aurons *décidé*. Pour continuer, pour refaire ce que nous venons de faire, il faut, en tout état de cause, à chaque pas, le *décider*. C'est de cette décision que dépend la construction de la suite, et non de telle ou telle propriété du nombre.

2.2 La question de la preuve

Wittgenstein va poursuivre la discussion directement avec Turing sur une autre question tout à fait décisive, celle de la preuve. Qu'est-ce qu'une preuve ?

Sur un point, il s'accorde avec le « sens commun » du mathématicien : ce qui fait d'une proposition mathématique une preuve, c'est qu'elle montre « certaines connexions », une relation interne au système mathématique considéré (Wittgenstein, 1995, 132). Cependant, pour le sens commun du mathématicien, ce qui est ainsi montré par la preuve, est un « quelque chose », toujours déjà-là, qui existe indépendamment du mathématicien lui-même, et on se souvient de l'image du découvreur de paysage utilisée par Hardy pour expliquer sa conception de la position du mathématicien. Celui-ci peut être regardé comme un observateur qui s'efforce de distinguer chaque sommet du paysage de montagne qu'il a sous les yeux. Certains de ces sommets sont visibles directement, d'autres ne le sont qu'en suivant telle ou telle ligne de crête, d'autres, enfin, ne peuvent être que devinés. Le mathématicien « découvre » donc un paysage qui lui préexiste et qui continuera à exister après qu'il s'en sera détourné. Il s'efforce, en outre, de faire partager sa découverte aux autres hommes. Montrer à ceux-ci les sommets visibles, ainsi que les chemins par lesquels ils pourront, eux aussi, voir ceux qui ne le sont pas, c'est à peu près, soutenait Hardy, ce que fait le mathématicien quand il établit une preuve, laquelle est vraie parce qu'elle l'a toujours été et qu'elle continuera à l'être après la disparition du mathématicien qui l'établit ou l'utilise. Ce que Wittgenstein résume en disant que, pour le sens commun du mathématicien, la preuve est « intemporelle » : elle n'est pas liée à telle ou telle expérience sensible, à telle démonstration faite par tel mathématicien, tel jour à telle heure et à tel endroit (Wittgenstein, 1995, 24, 31).

Wittgenstein, quant à lui, insiste sur le fait qu'au contraire, une preuve est toujours liée à un certain emploi, à une application. Une preuve, en mathématique, consiste à établir la véracité d'une proposition, c'est-à-dire d'un résultat déjà connu, « découvert » dirait Hardy. Wittgenstein prend l'exemple de la multiplication « 36 x 21 = 756 ». Il y a plusieurs manières d'effectuer le calcul correspondant : selon la méthode classique apprise à l'école, ou encore en construisant un rectangle d'une longueur de 36 carrés et d'une largeur de 21 carrés. En comptant les carrés compris dans ce rectangle, on aboutit à 756. On dira que cette seconde manière de calculer « prouve » le résultat de la multiplication classique « 36 x 21 = 756 », c'est-à-dire établit la véracité de la proposition, de la suite de signes « 36 x 21 = 756 ». Le signe « = » ne peut pas être enlevé entre les deux membres de l'équation de même qu'énoncer, par exemple « 36 x 21 = 1000 » n'aurait simplement pas de sens. On voit bien, cependant, que la méthode du comptage des cases du rectangle n'est pas une preuve *en soi* ; elle ne l'est que *pour* la multiplication « 36 x 21 = 756 ». Ce qui importe, ici, est le résultat « 756 », c'est lui qui détermine la qualité de « preuve » de la méthode de comptage par rapport à la méthode classique, du fait qu'il est le résultat obtenu avec celle-ci.

Wittgenstein insistera, tout au long du cours, sur la diversité des manières de mener un calcul : c'est là qu'apparaît, en effet, le lien fondamental de la preuve à un emploi, à un usage. Ainsi,

lors de la discussion qu'il aura avec ses auditeurs, pendant les dernières leçons, à propos de la thèse de Russell et Whitehead selon laquelle les mathématiques sont fondées sur la logique. Aux deux méthodes de calcul évoquées à propos de la multiplication, s'en ajoute, ici, une autre : celle consistant à effectuer le calcul en adoptant le symbolisme russellien. On peut, par exemple, calculer une addition telle que 31 478 + 97 de trois manières : selon le symbolisme de Russell, selon une méthode apparentée à celle du rectangle de la multiplication « 36 x 21 = 756 », à savoir en comptant un à un les éléments de l'opération – en les mettant en relation avec la suite des entiers naturels – enfin en suivant la technique apprise à l'école – en « posant l'opération ». Rien ne garantit, cependant, et c'est là le point important, que nous obtenions le *même* résultat. La première méthode, utilisant le symbolisme russellien, est horriblement compliquée, et la seconde, c'est-à-dire le comptage, extrêmement laborieuse, de sorte que nous avons toutes les chances de nous tromper. Supposons que, par comptage, nous arrivions à un résultat de 31 576 au lieu de 31 575, pourra-t-on dire que le comptage « prouve » la proposition issue de l'addition classique « 31 478 + 97 = 31 575 » ? Quant à la méthode russellienne, n'implique-t-elle pas elle-même, pour établir qu'une proposition est une tautologie, le comptage des variables situées de chaque côté d'une équation ? Par quelle méthode ce comptage sera-t-il effectué ? Toujours est-il qu'aucune des méthodes mises en jeu, aucune des différentes manières de mener un calcul, ne peut être considérée, dit Wittgenstein, comme plus « fondamentale » que les autres (Wittgenstein, 1995, 280). Dès lors, si nous sommes dans une situation où nous éprouvons le besoin de vérifier un résultat et que nous souhaitons garantir une méthode à l'aide d'une autre, nous devons *décider* de celle qui permet de faire la vérification.

Cette question de la *décision* impliquée par la preuve rejoint la discussion sur l'infini que Wittgenstein avait eu avec Turing lors des premières leçons. L'exemple qu'il prend est celui de la division de 1 par 7.

1/7 = 0,142857142857142857142857142857…

On constate la récurrence de la séquence « 142857 » et on en fait une règle. Nous avons, par exemple, effectué la division selon la méthode classique, apprise à l'école, jusqu'à la centième position, nous avons constaté la récurrence des dividendes dans le reste, et cela nous a conduit à faire de l'autre récurrence, celle de la séquence « 142857 », une règle. Mais quelle récurrence « prouve » l'autre ? Nous pouvons aussi effectuer la division jusqu'à la trentième position et constater que nous obtenons le même résultat en copiant quatre fois la séquence des six premières positions. C'est, dans ce cas, l'effectuation de la division selon la méthode « classique » qui constituera une « preuve » pour la méthode consistant à copier x fois la séquence des six premières positions. Mais il faut pour cela qu'on ait *décidé* qu'il s'agit d'une preuve, c'est-à-dire qu'on ait décidé que le fait d'avoir fait la division au moins une fois de manière « normale » a prouvé quelque chose. On pourrait le décider après avoir fait cinq fois la division de manière classique, ou après l'avoir fait faire à vingt-cinq personnes différentes, mais en tout état de cause ce qui nous conduit à décider qu'il y a preuve, c'est le fait que le résultat soit toujours le même.

La *décision* prise a consisté à dire qu'on a le droit de poursuivre la division en copiant la séquence des six premières positions et qu'on a le droit de le faire à chaque étape, c'est-à-dire de répéter à chaque étape ce qu'on a fait lors de la précédente. Ce n'est donc pas parce que « quelque chose » nous contraint à poursuivre toujours de la même façon que nous nous arrêtons après x séquences – après tout, rien ne nous empêchera jamais de changer notre manière de faire au-delà, par exemple, de la trentième position, en *décidant* de continuer, non pas avec le reste « 1 », mais avec « 2 » ou « 80 » (nous pouvons toujours, dans ce cas, continuer de « calculer ») - mais parce que nous *décidons* que la suite sera toujours celle-ci, décision qui conduit à faire de la formule « 1/7 = 0,142857… » une proposition mathématique, à laquelle aucune autre ne peut être substituée. Comme le dit Hans-Johann Glock, « La preuve ne nous saisit pas à la gorge pour nous mener à la conclusion une fois que nous avons admis

les axiomes et les règles d'inférence. A chaque moment, nous pouvons faire et dire ce que nous voulons (dans les limites des lois physiques). Simplement, nous *n'appellerions* pas, par exemple, "1,500 x 169 = 18" une multiplication. La nécessité logique à l'intérieur d'un système se ramène à l'applicabilité de certaines expressions. »[17]. Aucun emploi n'existe pour une expression telle que « 1,500 x 169 = 18 », aussi bien ne fera-t-elle tout simplement pas partie de ce que nous appelons un calcul. Et il en serait de même si nous décidions, après la trentième position de la division de 1 par 7, de continuer en remplaçant le reste « 1 » par « 2 » ou « 80 ».

La preuve repose donc sur une décision et si une preuve est considérée par le sens commun comme toujours vraie, ce n'est pas parce qu'elle renvoie à un monde d'objets mathématiques toujours déjà-là, mais parce que la décision qui la fonde a consisté à la détacher de l'expérience à laquelle elle est liée pour en faire une règle *intemporelle*. Dans le cas de calculs simples, il est clair que la méthode du comptage, de la mise en corrélation des éléments sur lesquels porte le calcul avec la suite « 1, 2, 3, 4, 5, 6, 7, … », est facile à mettre en œuvre ; il n'en va plus de même dès lors que les calculs se compliquent. On peut considérer le comptage comme une preuve de la validité de la proposition « 3 + 4 = 7 », résultat de l'addition classique, mais ce n'est plus vraiment possible dans le cas de la proposition « 31 478 + 97 = 31 575 », et ça ne l'est plus du tout dans celui de calculs impliquant l'infini. L'établissement de *règles* devient alors nécessaire, règles qui, pour être telles, doivent être considérées comme toujours vraies, pour tout mathématicien ; ces règles, qui pourront servir de « preuves », c'est-à-dire entraîner l'adhésion de tous, reposent toujours sur une « décision » collective et sur le *pari* que la règle donne bien le résultat attendu. Un certain usage, une certaine manière de faire, sont admis par tous comme « règle », et, par là, sont détachés de l'expérience. Cette règle peut être *apprise* ; c'est par l'apprentissage qu'elle se diffuse, ou qu'elle manifeste son caractère d'universalité, à savoir le fait qu'elle ait été détachée de l'expérience singulière de tels ou tels mathématiciens.

Aussi bien le mathématicien, selon Wittgenstein, n'est-il pas un « découvreur de paysages », comme le voulait Hardy ; le mathématicien ne *découvre* pas certains chemins qui mènent de sommets en sommets dans le paysage de montagne qu'il a sous les yeux, il *invente,* plutôt*,* des chemins qui jusque là n'existaient pas, inventions qui sont partagées par la communauté des hommes. Le mathématicien est un constructeur de routes plutôt qu'un découvreur de paysages. Il n'est pas, comme le lui fait croire le sens commun qu'il partage avec les autres « mathématiciens de la rue », un « physicien des entités mathématiques » (Wittgenstein, 1995, 136), qui mettrait au jour les « lois », toujours déjà-là, du monde des entités mathématiques. Les mathématiciens appartiennent plutôt à une sorte de communauté de bricoleurs qui *inventent* des manières de faire dont certaines finissent par s'imposer à tous car elles « fonctionnent » (Wittgenstein, 1995, 10).

Les auditeurs de Wittgenstein, et au premier rang d'entre eux, Turing, sont perplexes. Un point, en particulier, les préoccupent dans l'argumentation de Wittgenstein : si la route est à « construire » et non à « découvrir », qu'est-ce qui empêche le mathématicien, lorsqu'il construit cette route, autrement dit lorsqu'il « invente » une manière de faire, de tomber dans la contradiction ?

2.3 La question de la contradiction

Pour le sens commun du mathématicien, si une contradiction apparaît dans un système mathématique ou logique, c'est qu'elle y était déjà, mais « cachée ». Son apparition, inévitable à un moment ou un autre, ruine le système. Les mathématiques et la logique ne peuvent pas

17 « The proof does not grab us by the throat and carry us to the conclusion once we have granted the axioms and rules of inference. At any point, we can do and say whatever we want (within the limits of physical law). It is just that we would not call, for example, `1,500 x 169 = 18´ a multiplication. The logical necessity within the system boils down to the applicability of certain expressions. », (Glock, 1996, 229). C'est nous qui traduisons.

renfermer de contradictions sous peine de perdre leur statut de science fondatrice pour les autres disciplines scientifiques et de ne plus pouvoir être ce « royaume » qu'elles sont pour les mathématiciens. Il est essentiel de les préserver de la contradiction, soit en élaborant un système dont il sera prouvé qu'il ne comporte aucune contradiction « cachée », soit, à défaut, en montrant de manière certaine, en *démontrant,* en *prouvant*, qu'un tel système ne peut pas être construit et en élaborant une méthode de recherche des contradictions.

Seule une démarche de cette nature préservera l'idée que se fait le mathématicien « de la rue » de la nécessité mathématique ou logique : celle d'une contrainte inexorable. La nécessité mathématique s'accomplit comme le mouvement d'une mécanique, transmis sans perte ni déviation par les éléments, les « pièces », parfaitement « rigides », qui la composent (Wittgenstein, 1995, 200-204).

Aussi bien le sens commun du mathématicien conduit-il à insister sur la portée de la question de la contradiction : à travers celle-ci, les mathématiques sortent de leur « royaume » et étendent leur empire, pour ainsi dire, sur la vie et l'expérience commune, car une contradiction en mathématiques peut entraîner des catastrophes, telles que, par exemple, l'écroulement d'un pont qui aura été construit sur la base de calculs recelant une contradiction cachée. La question des « fondements » des mathématiques devient ainsi une urgence non seulement pour le « mathématicien de la rue », mais pour « l'homme de la rue ».

Wittgenstein va s'efforcer de montrer qu'en réalité, la contradiction n'a pas, en elle-même, cette importance éventuellement dramatique qu'on lui accorde.

Prenons l'exemple d'un ordre contradictoire tel que : « Quittez cette pièce et ne la quittez pas ». Un ordre fonctionne parce qu'il renvoie à une règle qui indique ce qu'il faut faire et que cela doit être fait ; si l'on nous dit « Quittez cette pièce », nous devons sortir de la pièce. Si l'ordre est contradictoire - « Quittez la pièce et ne la quittez pas » - il est clair qu'il ne pourra pas être exécuté. La règle sera inapplicable. Mais on peut la modifier ; on pourrait imaginer, par exemple, que, devant un ordre tel que « Quittez la pièce et ne la quittez pas », la « règle » a suivre soit : « quittez la pièce avec hésitation ». C'est là, sans aucun doute, l'un des comportements qu'un individu humain dans une telle situation pourrait adopter : obéir à l'une ou l'autre partie de l'ordre contradictoire en ajoutant la modalité « hésitation » (quelque chose comme deux pas en avant, un pas en arrière). De toutes manières, en cas de blocage total, un autre événement finira par transformer la situation. En tout état de cause, la vie continuera… Confronté à un ordre contradictoire, un individu humain quelconque peut se contenter de rester sur place sans rien faire - et ce sera alors la solution qu'il aura « choisi » de donner au problème - ou bien interpréter la situation en sortant de manière hésitante (deux pas en avant, un pas en arrière), mais, dans un cas comme dans l'autre, il *fera* bien quelque chose et la situation continuera à évoluer (Wittgenstein, 1995, 211-212).

Ce que Wittgenstein veut faire ressortir, c'est que la contradiction est toujours liée à un emploi. Elle n'est pas un « quelque chose » qui existerait en soi, quelque part, comme un abîme non encore découvert. Si Wittgenstein dit : « Asseyez-vous et ne vous asseyez pas », à Lewy, à Malcolm ou à Turing, ceux-ci ne sauront certes pas comment exécuter cet ordre contradictoire, mais ils ne seront pas pour autant « bloqués » comme le serait une machine. Ils agiront, ils « décideront », fusse en ne faisant rien et, si la situation exige qu'ils fassent quelque chose, ce sera à leurs risques et périls ; si en revanche la situation n'exige rien, s'il n'y a pas de risque particulier, alors la contradiction n'aura aucune importance. Ainsi du paradoxe du menteur. Wittgenstein remarque qu'« il est en un sens très bizarre que ce cas ait pu mettre qui que ce soit dans l'embarras […] si un homme dit "je mens", nous disons qu'il en découle qu'il ne ment pas, d'où il découle qu'il ment, etc. Eh bien quoi ? Vous pouvez poursuivre ainsi jusqu'à en devenir noirs de fureur. Pourquoi pas ? C'est sans importance. » (Wittgenstein, 1995, 212).

Le paradoxe du menteur est un jeu de langage, mais sans emploi et, dès lors, on peut, sans doute, s'amuser, ou pas, avec ce jeu, mais cela n'a aucune importance et, de fait, la majeure partie des humains l'ignorent sans que cela change quoi que ce soit à leur existence. Ce qui

nous dérange dans ce paradoxe, dit Wittgenstein, c'est « l'absence de système. Vous voulez savoir pourquoi une contradiction apparaît avec "Je mens" et pas avec "Je mange" » (Wittgenstein, 1995, 213). Nous sommes en présence d'un usage linguistique inédit ; jusque là, avec les autres « verbes », nous n'avions rien rencontré de tel. Autrement dit, ce n'est pas la contradiction elle-même qui nous trouble, mais sa présence dans le langage. La contradiction, en l'occurrence, ne « bloque » rien : au bout d'un moment, nous cesserons de dire : « donc tu ne mens pas, donc tu mens, etc. » et, certainement avant d'être noirs de fureur, nous passerons à autre chose. Le mystère demeurera, mais nous l'oublierons et il ne nous gênera en rien.

Qu'en est-il, cependant, des conséquences pratiques qui peuvent découler d'une mathématique qui ne serait pas préservée de la possibilité de la contradiction ? Turing et Prince, pendant le cours, imaginent ainsi que les ingénieurs chargés de la construction d'un pont utilisent deux méthodes de calcul donnant des résultats différents : « … supposez, dit Turing, que l'on ait deux systèmes, dont l'un a toujours été dans le passé employé avec succès pour construire des ponts. Puis on emploie le second système et le pont s'effondre. En comparant ensuite les deux systèmes, on se rend compte que les résultats ne concordent pas. » (Wittgenstein, 1995, 219). Il y avait une contradiction « cachée ».

Wittgenstein objecte que, si une contradiction logique au sein d'un système logique, ou entre deux systèmes logiques, ne peut pas être quelque chose de contingent, le fait que le pont s'écroule, lui, l'est : il se peut que le pont ne s'écroule jamais. Qu'en est-il alors de la contradiction « cachée » ? Une telle notion est étrange, fait-il remarquer : existe-t-elle tant qu'elle est cachée ? Tant que que ce que nous tenons pour ses conséquences ne s'est pas produit, ou tant que la contradiction en tant que telle n'est pas apparue, tout va bien et il n'y a aucune raison de considérer que, jusque là, nous nous soyons trompés. Par ailleurs, si la contradiction apparaît, et qu'il s'en suit une catastrophe, cela ne signifiera pas pour autant que le problème soit d'ordre *mathématique*. Il pourra être du à une erreur dans l'usage, ou bien à une mauvaise interprétation du phénomène naturel (autrement dit à l'ignorance).

Si le pont, une fois construit, ne s'écroule pas, mais qu'on s'aperçoive, en refaisant les calculs, qu'il y a, non pas une erreur de calcul, mais bien une contradiction dans le système utilisé pour faire ces calculs, cette contradiction n'aura, en tant que telle, causé aucun dommage, même si on ne peut exclure un écroulement du pont dans l'avenir. L'embarras dans lequel nous plonge l'écroulement du pont est causé par l'écroulement du pont, et non par la contradiction logique en elle-même. Dans le cas précis de tel pont qui s'écroule à tel moment et à tel endroit, il sera tout à fait exceptionnel qu'on puisse rattacher sans discussion, sans aucun *doute* possible, l'accident à une contradiction logique dans le système général de calcul utilisé. Si cela, du reste, pouvait être établi, la contradiction *mathématique* ne serait certainement pas la cause de notre embarras, ou alors celui-ci devrait être le même – amusé ou horrifié - devant le paradoxe du menteur et devant le pont écroulé.

Les ingénieurs, lorsqu'il font les calculs nécessaires à la fabrication d'un pont, suivent certaines règles, mais ce ne sont pas ces règles en tant que telles qui importent, c'est le pont ; il s'agit de faire en sorte qu'il ne s'écroule pas. On vérifiera donc les calculs par divers moyens, qui ne sont pas nécessairement mathématiques - maquettes, tests expérimentaux, etc -. Le calcul, en l'occurrence, est un outil et s'il s'avère que cet outil, tel qu'il est à notre disposition, fait défaut, il s'agira d'en trouver un autre, c'est-à-dire de formuler d'autres règles.

A l'objection que d'un système contradictoire n'importe quelle conclusion peut être tirée, Wittgenstein répond en disant : « Eh bien, ne tirez donc aucune conclusion d'une contradiction. » (Wittgenstein, 1995, 215), ce qui laisse Turing à tout le moins perplexe – on l'imagine haussant les épaules : « Vous semblez dire que si l'on faisait preuve d'un peu de bon sens, on ne tomberait pas dans l'embarras » (Wittgenstein, 1995, 227). Wittgenstein proteste : il n'a jamais prôné un tel recours au bon sens le plus plat. Ce qu'il veut mettre en évidence, c'est le fait qu'il n'existe pas quelque chose du genre d'un parapet infranchissable qui

constituerait un obstacle *objectif*, quasi matériel, à la contradiction. Le « bon sens » n'est pas davantage un obstacle de ce genre : il y a des contradictions que nous n'avons pas prévues – les fameuses « contradictions cachées » - dans lesquelles nous tomberons malgré tout notre bon sens. Toujours est-il que les contradictions n'ont pas forcément l'importance démesurée que leur donnent les mathématiciens ou les logiciens.

S'agissant de la contradiction « cachée », en réalité, c'est-à-dire *en pratique*, tant qu'elle n'a pas été « découverte », le calcul n'est pas erroné et il est conservé tel quel (du moins dans le cas où nous ne disposons pas d'une méthode pour chercher les contradictions) ; une fois que la contradiction a été découverte, les règles sont réaménagées, c'est-à-dire qu'on passe à un *autre* calcul : « Prenons le cas d'un calcul dont vous découvrez plus tard qu'il comporte une contradiction. Nous pourrions également dire qu'aussitôt que vous avez découvert la contradiction, ce n'est plus le même calcul. » (Wittgenstein, 1995, 233).

Du reste, est-ce que la possibilité de la contradiction empêche les mathématiciens de faire des mathématiques ? Évidemment non. Le maître et collègue de Turing, Hardy, en est le meilleur exemple : ne revendique-t-il pas de pouvoir continuer à faire des mathématiques, en utilisant des notions philosophiquement mal établies, telles que l'infini par exemple, quand bien même le débat sur les fondements n'est pas tranché ? Se montrer « ingénieux », établir des preuves indirectes, destinées à « convaincre », des preuves qui ne peuvent pas être démontrées formellement, mais qui font appel aux moyens de « la vie ordinaire », à savoir le « bon sens » et « l'intuition », tout cela, Hardy en est convaincu, est essentiel à l'activité du mathématicien. Aussi bien, et c'est là une des croyances du mathématicien praticien, comme Hardy le mentionnait dans son article de 1929, si les machines peuvent calculer, elles ne font pas de mathématiques. Turing l'avait établi, peu avant sa discussion avec Wittgenstein, en formalisant la notion de machine et en montrant qu'un procédé mécanique ne peut pas prendre la place de « l'ingéniosité ». Cette croyance, cette conviction du mathématicien « de terrain » (ou « de la rue »), Wittgenstein ne la conteste pas ; il s'efforce au contraire d'expliquer en quoi elle est fondée.

Quand on dit que la contradiction « bloque », c'est, en effet, à la machine qu'on pense : ce n'est que dans le cas où elle concerne une machine que l'image du blocage prend sens. Il est de fait que la machine qui se trouve face à des instructions contradictoires « s'arrête », est empêchée, « bloquée », dans son processus, mais c'est parce qu'elle ne sait pas *appliquer une règle*. La machine incarne les règles, elle les exprime, les formalise dans un symbolisme particulier, mais elle ne les *applique* pas.

Wittgenstein prend ainsi l'exemple du réveil. Un horloger nous montre que l'aiguille des minutes tourne d'un quart de tour lorsque nous faisons tourner celle des heures de trois tours ; nous pouvons théoriquement, à partir de là, anticiper la situation suivante et prévoir ce que fera l'aiguille des heures lorsque nous faisons tourner celle des minutes. Mais nous pouvons tout aussi bien ne pas réussir à prévoir correctement le mouvement des aiguilles, nous pouvons ne pas bien comprendre ce qui nous a été montré, ou encore, comprendre qu'on a voulu nous montrer qu'il y a un certain rapport, mesurable mathématiquement, entre le mouvement des deux aiguilles, mais ne pas être convaincus pour autant que ce mouvement sera toujours déterminé par le rapport mathématique entre elles que l'horloger nous a indiqué. Devant notre scepticisme, l'horloger démontera le cache du réveil et nous fera voir le mécanisme ; nous constaterons alors que le mouvement des aiguilles résulte de l'enchaînement des mouvements d'un système de roues dentées et de ressorts et il ne fait guère de doute que nous nous convaincrons, non seulement qu'il y a une certaine nécessité dans le mouvement des aiguilles, mais que cette nécessité a son principe dans le mécanisme, dans le système de roues dentées et de ressorts. Pourtant, le problème aura été simplement déplacé : qu'est-ce que l'horloger, en nous révélant le mécanisme, nous aura montré d'autre que lorsqu'il nous désignait le mouvement des aiguilles ? Pourquoi l'examen du mécanisme devrait-il faire disparaître notre scepticisme à l'égard de la nécessité qui gouverne le

mouvement des aiguilles ? Ce qui nous a été montré, dans le cas du mouvement des aiguilles seules, puis dans celui du mécanisme dans son ensemble, c'est une certaine régularité, laquelle, dans le second cas, lorsqu'elle est celle du mouvement des roues dentées et des ressorts, n'est pas d'une autre nature que dans le premier, où elle est celle du mouvement des aiguilles. Le mécanisme pris dans son ensemble n'est qu'une autre figure de cette régularité. La « règle », en l'occurrence, c'est-à-dire le rapport constant entre les mouvements des deux aiguilles, n'est certainement pas davantage « appliquée » par le mécanisme, par le système de roues dentées, qu'elle ne l'est par le déplacement des deux aiguilles, elle est simplement figurée par l'un et par l'autre, elle prend la forme de l'un et de l'autre.

Bref, la machine ne fait pas de mathématiques, car celles-ci sont d'abord une affaire d'usages, d'emplois, d'applications ; elles mettent en jeu des processus qui sont de l'ordre de la « décision », en l'occurrence d'une décision collective. La nécessité dont elles sont l'expression renvoie à de tels processus et non à un monde « d'entités mathématiques » dont le mathématicien « découvrirait » les lois comme le physicien découvre les lois de la physique.

On comprend qu'une telle philosophie ne pouvait être « sympathique » aux mathématiciens « de la rue », auxquels elle refusait la satisfaction d'un de leurs désirs les plus profonds, celui d'une *correspondance* entre la proposition mathématique et une réalité mathématique autonome. Ce n'est pas un bouleversement « à la marge » de leurs convictions implicites les mieux ancrées que Wittgenstein leur demandait, mais une remise en question radicale. La perplexité des auditeurs de Wittgenstein est manifeste et on perçoit l'agacement de Turing devant les « supposez que... », « imaginez que.. », etc. qui rythment l'argumentation de Wittgenstein, et qui, dans certains cas, lui paraissent tout à fait arbitraires. Turing, institué le représentant quasi officiel du « sens commun » mathématicien systématiquement pris à contre-pied par Wittgenstein, ne peut manquer d'être troublé. Du reste, il n'assistera pas aux six derniers cours[18].

La réflexion de Wittgenstein rejoignait pourtant la conviction profonde qui était déjà celle de Hardy, et que Turing avait confirmée : l'ingéniosité ne peut être remplacée par la patience, en l'occurrence une patience mécanique, de sorte que Turing, troublé, agacé sans doute, se voyait aussi proposer une réponse à l'interrogation soulevée par la constatation qu'une machine manipulant les « lois de la pensée » exactement comme un humain, ne « pense » pas elle-même. Wittgenstein était le seul à fournir une piste qui ne renvoie pas d'une manière ou d'une autre au « mystère » de la pensée : penser, c'est former collectivement des règles et les appliquer, ce que ne fait pas la machine qui calcule. Certes, les calculs effectués par une « machine universelle » de Turing, par une « machine à oracle », ont la même extension que ceux effectués par un humain ; la machine de Turing manipule des symboles comme le fait un calculateur humain qui écrit des signes sur du papier ; cependant elle n'invente pas de règles, et n'applique pas de règles. Son propre rapport à la règle est le même que celui qu'a à cette dernière la séquence de signes ; la machine n'est rien d'autre qu'une suite de signes.

Nul ne sait combien de temps a duré le trouble de Turing, qui fut rapidement pris, dès l'été 1939, dans le tourbillon de la guerre. On ne sait pas non plus si les deux hommes ont eu de nouveaux entretiens[19]. Cependant, publiant en 1950, dans *Mind,* « Les ordinateurs et l'intelligence », Turing laisse entrevoir en quel sens la discussion de 1939 a relancé sa propre réflexion. Les remarques de Wittgenstein lui ouvrait une perspective : une machine de Turing n'applique pas de règles, mais ne peut-on la rendre capable de le faire ?

La règle correspond à un usage, elle apparaît comme telle lorsqu'il est collectivement admis qu'une certaine manière de faire peut être détachée de l'expérience et posée comme « intemporelle ». La règle s'installe alors par l'apprentissage. Une machine de Turing ne peut-

18 Rien ne dit, cependant, que ce ne soit pas simplement pour des raisons d'emploi du temps ; le cours se termine quelques semaines avant que la guerre n'éclate, et alors que Turing travaille déjà sur « L'Enigma » à la *Government Code and Cypher School*.

19 Ils se sont certainement rencontrés en 1947-1948, pendant les quelques mois où ils ont été présents tous les deux à Cambridge.

elle apprendre comme le fait un humain qui applique des règles ? Question qui met en jeu la capacité, non seulement à manipuler des signes, comme le fait la machine, mais à communiquer à travers cette manipulation. C'est à quoi Turing va réfléchir après la guerre. Son article « Les ordinateurs et l'intelligence » est le fruit de cette réflexion.

3 La machine et l'application de règles ; « Les machines peuvent-elles penser ? »

Dans « Les ordinateurs et l'intelligence », Turing entend « considérer la question : les machines peuvent-elles penser ? » (Girard, 1995, 135). Il précisera plus loin que, telle quelle, la question « a trop peu de sens pour mériter une discussion » (Girard, 1995, 148) ; c'est pourquoi il propose de la reformuler sous l'angle d'un « test », le fameux « jeu de l'imitation ». Il examine ensuite une série d'objections à l'hypothèse qu'une « machine de Turing » puisse réussir le test, et termine enfin son argumentation en opposant à ces objections l'idée que les machines peuvent « apprendre » et en décrivant ce que pourraient être les bases du processus « d'éducation » d'une machine.

3.1 Le jeu de l'imitation

Turing avait exposé le principe du jeu de l'imitation, à la fin des années quarante, dans un texte intitulé *Intelligent Machinery* (Turing, 1993), rédigé à l'intention du National Physical Laboratory, l'organisme de recherche auquel il était alors rattaché. Imaginons, disait-il, un joueur d'échec, de niveau plutôt moyen, affrontant deux adversaires de force à peu près égale à la sienne, et dont l'un serait une « machine de Turing » ; ce joueur, remarquait-il, aurait certainement les plus grandes difficultés à décider lequel de ses adversaires est la machine (Turing, 1993, 127)[20]. Cet exemple servait de modèle au jeu décrit peu après dans « Les ordinateurs et l'intelligence ».

Dans une première configuration, le jeu oppose un homme A, une femme B et un examinateur C. Les joueurs ne peuvent ni se voir ni s'entendre, ils communiquent à l'aide d'un téléscripteur. C, en posant des questions à A et B doit décider qui est l'homme et qui est la femme. A s'efforce de tromper C en se faisant passer pour la femme B ; celle-ci au contraire aide C. Les questions posées par C peuvent porter sur n'importe quel sujet.

Les trois joueurs sont des individus humains quelconques, ils sont donc, par principe, à peu près de la même force, de sorte que A, sur une série significatives de sessions de jeu, aura *grosso modo*, 30 % de chances de l'emporter.

Que se passe-t-il, demande Turing, si A est remplacé par une machine ? C se trompera-t-il aussi souvent que lorsque le jeu se déroule entre un homme et une femme ?

L'hypothèse de Turing est que, dans cette nouvelle configuration du jeu, les chances des joueurs seront réparties de la même manière que dans la version précédente : « un interrogateur moyen n'aura pas plus de 70 % de chances de procéder à l'identification exacte après cinq minutes d'interrogation. » (Girard, 1995, 136). Bref, selon Turing, une machine peut avoir autant de chances de faire bonne figure au jeu de l'imitation qu'un individu humain quelconque ; elle peut se montrer *l'égale* d'un être humain placé dans les conditions du jeu - réduit à la parole - c'est-à-dire l'égale d'un être dont les observateurs du jeu admettent qu'il pense.

Il est important de souligner que le test de Turing ne vise pas à établir une quelconque supériorité du mécanique sur l'humain, ou l'inverse. Le test doit permettre de constater simplement qu'une entité mécanique et un individu humain *peuvent se mesurer* à un jeu tel que celui de l'imitation. Le point décisif est bien que la machine y ait les mêmes chances que

20 Turing s'était lui-même essayé à concevoir une machine « de papier » jouant, de manière élémentaire, aux échecs (Hodges, 1988, 186-190).

des joueurs humains quelconques. Le test, en somme, sera réussi si une machine s'avère simplement capable de tenir sa place au jeu, comme le ferait n'importe quel joueur humain.

La condition d'une telle participation est évidemment que la machine puisse avoir avec ses adversaires humains un *échange de paroles*.

Or, Turing avait établi, dans les années 1930, que, s'il est possible de *simuler,* à l'aide d'une machine, la manipulation de symboles propre aux humains, cette simulation ne peut pas être confondue avec la manipulation réelle, telle qu'elle s'effectue à travers le langage. De ce point de vue, la machine de Turing donnait, au mieux, l'exemple d'un discours à la troisième personne, de l'ordre du « il y a » ou du « ça », le prototype de tels énoncés étant la proposition mathématique.

Il n'est certes pas impossible d'imaginer qu'une machine de Turing puisse participer au jeu de l'imitation : on lui fera puiser dans une base de données des réponses toutes préparées aux questions posées pendant le jeu. C'est là ce que savent faire, avec un degré de sophistication plus ou moins grand, les programmes d'intelligence artificielle depuis quasiment les années 1950. Aucun programme de ce type n'a, pourtant, jusque là satisfait au test de Turing. John Searle a montré leurs limites à l'aide de sa célèbre expérience de pensée dite de « la chambre chinoise ». Imaginons qu'un être humain qui ne comprend pas un mot de chinois soit enfermé dans une pièce où il dispose de paniers dans lesquels se trouvent des symboles de la langue chinoise et d'un manuel, écrit dans sa propre langue, contenant des règles purement syntaxiques de manipulation de ces symboles. D'autres symboles chinois sont introduits dans la pièce et le manuel fournit des règles indiquant à notre opérateur que certains symboles doivent être sortis de la pièce dans un certain ordre. Supposons, enfin, que, sans que l'opérateur le sache, les chaînes de symboles introduites soient des « questions » et celles qu'il sort des « réponses » à ces questions. Si les règles ont été correctement rédigées, et si l'opérateur ne fait pas d'erreur en les suivant, tout se passera exactement comme si, à des questions posées par un chinois, des réponses étaient données par un chinois. Pourtant, remarque Searle, « dans une telle situation », c'est-à-dire dans celle de l'opérateur humain placé dans la « chambre chinoise », « je vous défie d'apprendre un mot de chinois... » (Searle, 1985, 43). On voit ainsi que le succès éventuel d'un programme classique d'intelligence artificielle au test de Turing ne suffirait pas à prouver qu'il aurait eu un véritable échange de paroles avec ses interlocuteurs humains. Quand bien même l'examinateur C du jeu croirait avoir affaire, non à une machine, mais à un humain, l'interaction entre lui et la machine ne pourrait être considérée comme un « dialogue ». C ne dialoguerait avec personne – en tout cas pas avec une *personne.* Bien plus, on peut douter qu'une machine ainsi programmée puisse réussir autre chose qu'une version très affaiblie du test. Rappelons qu'il s'agit pour la machine de participer au jeu de l'imitation en ayant les *mêmes* chances que les deux autres joueurs humains. C'est du reste ce qui « sauve » le test de Turing, lequel n'est valide que si a lieu entre les interlocuteurs un échange de paroles au plein sens du terme, impliquant de la part de chaque protagoniste un engagement existentiel, à la *première* personne ; la participation au jeu de l'imitation suppose, non seulement la construction de suites de signes, mais un *acte d'énonciation.* L'acte d'énonciation, c'est là, en effet, ce que partagent avant tout les interlocuteurs d'un dialogue véritable. Comment Turing imaginait-il rendre une machine capable d'un tel acte ?

3.2 Les « machines inorganisées »

Wittgenstein lui avait montré que ce qui différenciait les humains d'une « machine de Turing », était le fait que cette dernière *n'appliquait* pas de règles, lesquelles sont liées à des usages, c'est-à-dire à l'élaboration collective de certaines manières de faire, et à leur diffusion à travers l'apprentissage.

Wittgenstein n'avait jamais envisagé qu'une machine de Turing puisse participer à un tel processus et la question, pour lui, n'aurait pas eu de sens. Elle en prend un pour Turing dans la

mesure où elle lui permet de se dégager de réflexions renvoyant au « mystère » de la pensée. Turing va s'efforcer d'imaginer de quelle manière, théoriquement plausible, une machine pourrait *acquérir* la « faculté » de manipuler des signes et de communiquer avec des interlocuteurs humains.

Il avait réfléchi à la question dans le rapport de 1948 déjà cité, *Intelligent Machinery,* en s'appuyant sur la possibilité, notée dans son mémoire de 1936 sur les « nombres calculables », de concevoir une machine de Turing non déterministe : « Pour certains objectifs, nous pouvons utiliser des machines (machines à choix ou *c*-machine), dont le mouvement n'est que partiellement déterminé par sa configuration [...]. Lorsqu'une telle machine atteint l'une de ces configurations ambiguës, elle ne peut continuer tant qu'un choix arbitraire n'a pas été fait par un opérateur extérieur. » (Girard, 1995, 52)[21].

Dans *Intelligent Machinery*, Turing décrivait une telle machine « aléatoire » (*random machine*) : à certaines étapes de son parcours, deux directions, au moins, sont possibles, et la machine est conçue de manière à ce que le choix de l'une ou l'autre soit déterminé par une procédure faisant intervenir le hasard - un équivalent électronique du coup de dé. Turing appelait « machine inorganisée » (*unorganized machine*) une machine de ce type.

Il imaginait qu'une machine inorganisée pouvait être soumise à des « interférences », c'est-à-dire à certaines conditions imposées de l'extérieur et entraînant une modification de son comportement. Turing soutenait, en particulier, qu'il serait possible d'obtenir ces modifications en simulant le système « punitions-récompenses » qui, peu ou prou, est utilisé dans tout processus d'éducation d'un enfant (Turing, 1993, 121). Sous cet angle, l'action d'une machine sera déterminée par trois éléments : son programme, le signal d'entrée qui lui est fourni, autrement dit, le contact avec l'extérieur qui peut donner lieu à interférence, et, enfin, un « stimulus de douleur » ou de « plaisir », correspondant à une « punition » ou à une « récompense ». Dans le cas où l'action sera considérée comme une mauvaise réponse au signal d'entrée, un « stimulus de douleur » - un signal « punition » - sera envoyé à la machine et entraînera un changement aléatoire des instructions correspondantes du programme de celle-ci ; dans le cas où, au contraire, l'action sera considérée comme une réponse juste, c'est un « stimulus de plaisir » - un signal « récompense » - qui sera envoyé, entraînant la fixation des instructions correspondantes du programme. Le comportement que l'on visera à faire adopter par la machine pourra alors être défini sous la forme d'un ensemble de signaux d'entrée figurant différents cas susceptibles d'être rencontrés dans le contexte de la situation correspondant à ce comportement. Pour chacun de ces signaux d'entrée, la machine tendra statistiquement, par essais et erreurs sanctionnés par un signal de peine ou de plaisir, à atteindre une configuration fixe. Il est clair qu'il ne sera pas possible de définir *a priori* l'ensemble des signaux d'entrée constituant un comportement relativement complexe, même si cet ensemble est considéré comme fini ; à tout moment la machine pourra se trouver face à un signal d'entrée jusque là non essayé et déterminant une réponse qui n'a pas encore été évaluée ; pour autant, selon Turing, rien n'interdit de penser que, sur une durée suffisamment longue, la machine tendra à donner une réponse considérée comme « bonne » du point de vue de l'ensemble des signaux d'entrée définissant le comportement visé.

Une machine inorganisée pourra ainsi être transformée en une « machine organisée », conçue dans un but déterminé. Plus précisément, Turing estimait qu'une machine inorganisée pourrait être transformée en une machine universelle, susceptible de simuler le comportement d'autres machines.

Une machine inorganisée peut donc évoluer. Sa transformation même en machine universelle n'est pas directement inscrite dans sa structure ; elle pourrait être transformée en tout autre chose. Son devenir dépendra, en réalité, du système de punitions-récompenses qui est appliqué, ou, plus exactement, de « l'histoire » des punitions et récompenses effectivement

21 Notion que Turing utilisait dans sa thèse américaine, *Systems of Logic Based on Ordinals,* pour raisonner sur la « machine à oracle ».

« vécues » par elle. Or, le même principe peut être adopté à l'égard de la machine lorsqu'elle est devenue une machine universelle puisque celle-ci présente, on le sait, la particularité de pouvoir se modifier elle-même : elle peut être conçue de telle sorte que le résultat auquel elle parvient entraîne une modification de la configuration de la machine simulée. Il est ainsi permis d'imaginer une machine universelle qui soit pourvue d'un système initial modifiable à l'aide du processus de « punitions-récompenses » ; les modifications ne porteront pas sur la structure de la machine universelle en tant que telle, mais seulement sur ce système initial.

3.3 La « machine qui apprend »

Tel est le principe que Turing met en œuvre, dans « Les ordinateurs et l'intelligence », lorsqu'il réfléchit à l'apprentissage par une machine et à la voie à suivre pour construire une machine qui serait éventuellement capable de participer au jeu de l'imitation : « Au lieu de produire un programme qui simule l'esprit de l'adulte, pourquoi ne pas essayer plutôt d'en produire un qui simule celui de l'enfant ? » (Girard, 1995, 169). Il est, en effet, raisonnable de penser que le cerveau d'un enfant sera plus facile à simuler que celui d'un adulte, s'il est vrai qu'une grande partie des processus que celui-ci met en œuvre ne sont pas « innés », mais acquis au cours de l'éducation[22]. Turing distingue, en effet, trois composantes dans un « esprit humain adulte » : « (a) L'état initial de l'esprit, disons à la naissance. (b) L'éducation à laquelle il a été soumis. (c) D'autres expériences, que l'on ne peut pas décrire comme éducatives, auxquelles il a été soumis » (Girard, 1995, 169). C'est pourquoi le problème peut être divisé « en deux parties : le programme-enfant et le processus d'éducation » (Girard, 1995, 169), et la tâche de programmation proprement dite – au sens informatique classique - concernera uniquement la première partie.

Avant d'examiner celle-ci, Turing précise ce qu'il entend, en l'occurrence, par « processus d'éducation ». L'« éducation » de la machine sera conduite à l'aide d'une simulation du système « punitions-récompenses », de telle sorte que la probabilité associée aux « bonnes » réponses devienne dominante[23]. Cependant, un enseignement strictement fondé sur le système punitions-récompenses poserait un problème : « En gros, si le maître n'a pas d'autres moyens de communiquer avec l'élève [NB : que le système punitions-récompenses] la quantité d'information qui peut lui parvenir ne dépassera pas le nombre total de récompenses et punitions utilisées » (Girard, 1995, 170). Si la machine-enfant ne pouvait « apprendre » que par le biais d'un tel système, elle serait sans doute, le plus souvent, dans l'incapacité de décider si elle doit ou non faire telle chose, les réponses possibles n'ayant pas encore été significativement pondérées par le système punitions-récompenses. C'est pourquoi « il est [...] nécessaire d'avoir d'autres canaux de communication ' non-émotionnels ' » (Girard, 1995, 171)[24] : la machine-enfant devra être aidée. Lorsque cela sera possible, son constructeur devra lui fournir directement les connaissances dont elle peut avoir besoin dans telle ou telle circonstance ; il s'agira là d'un procédé « autoritaire » assez proche, selon Turing, de l'apprentissage « par coeur » pour les individus humains. Par là, enfin, remarque Turing, le processus d'éducation peut être rapproché de celui de l'évolution biologique :

« Il y a un lien évident entre ce processus et l'évolution, à travers les identités suivantes :

Structure de la machine-enfant = matériel héréditaire.

Changement dans la machine-enfant = mutations.

Sélection naturelle = jugement de l'expérimentateur. » (Girard, 1995, 169).

22 « Il est probable que le cerveau d'un enfant est une sorte de carnet acheté en papeterie : assez peu de mécanisme et beaucoup de feuilles blanches [...] Notre espoir est qu'il y ait si peu de mécanisme dans le cerveau d'un enfant qu'il soit très facile de le programmer. » (Girard, 1995, 169).

23 « Il faut construire la machine de manière que les événements qui précèdent immédiatement l'apparition d'un signal-punition aient peu de chances de se reproduire, alors qu'un signal-récompense doit accroître la probabilité de répétition de l'événement qui l'a provoqué. » (Girard, 1995, 169).

24 Les moyens « non-émotionnels » sont ceux qui ne reposent pas sur le système « punitions-récompenses ».

Bref, l'expérimentateur sélectionnera les changements intervenus dans la machine-enfant. Il en découle que le processus ne sera pas entièrement soumis au hasard puisque c'est en fonction d'un but déterminé que l'expérimentateur sélectionnera les « mutations ». Le modèle mis en œuvre est, en quelque sorte, celui d'une évolution à finalité explicite[25]. L'expérimentateur choisira, entre plusieurs « mutations », celle qu'il jugera la plus « utile » ; il pourra, du reste, introduire lui-même dans le programme une modification destinée à provoquer la « mutation » souhaitée - laquelle pourra entraîner d'autres mutations, quant à elles imprévues.

La question décisive est celle de savoir quel doit être « l'état initial » de la « machine-enfant ». Quel « système initial » devra être fourni « en entrée » à la machine universelle ? La réflexion de Turing à cet égard fait très clairement écho aux discussions qu'il avait eues avec Wittgenstein en 1939. Ainsi, par exemple, à propos de la « théorie des types » de Russell, discutée, on le sait, par Wittgenstein : « Les processus d'inférence utilisés par la machine [NB : la « machine-enfant »], note-t-il, n'ont pas besoin d'être de nature à satisfaire les logiciens les plus exigeants. Il se pourrait par exemple qu'il n'y ait pas de hiérarchie des types » (Girard, 1995, 172). Certes, le risque, alors, sera que la machine se trouve face à des paradoxes logiques[26], cependant, explique Turing, « ceci ne veut pas obligatoirement dire que les erreurs de type auront lieu, pas plus que nous ne sommes obligés de tomber du haut de falaises non protégées » (Girard, 1995, 172). Il ne s'agit pas de mettre la machine à l'abri des paradoxes logiques, mais de faire en sorte qu'elle parvienne, comme l'homme, soit à les éviter, soit à circonscrire leurs effets, dans l'exercice du raisonnement tel qu'il est mis en œuvre, non pas d'abord par le logicien confirmé, mais par un individu quelconque. Lorsque la machine, dans des circonstances déterminées, entrera en conflit avec elle-même, la hiérarchie nécessaire pour éviter ce conflit devra être construite dans le cadre de l'« éducation », non pas, autrement dit, par l'ajout au système d'inférence d'une règle *logique* supplémentaire, mais par ce que Turing appelle un « impératif » : « Des impératifs adéquats (exprimés *à l'intérieur* du système, ne faisant pas partie des règles *du* système), tels que : "n'utilisez pas une classe, à moins qu'elle ne soit une sous-classe de l'une de celles que le maître a mentionnées", peuvent avoir un effet similaire à : "ne t'approche pas trop près du bord" » (Girard, 1995, 172).

L'objectif poursuivi ne consistera pas, en effet, à construire une machine à calculer[27], mais une machine effectuant, dans des circonstances données, des « choix » qui, par le biais des « impératifs », soient de moins en moins arbitraires, et le soient le moins possible au moment du choix. Turing, en somme, ne suppose pas que les êtres humains soient eux-mêmes dotés, à leur naissance, c'est-à-dire dans leur « état initial », d'un système d'inférence logique susceptible de satisfaire les logiciens les plus exigeants. Il fait plutôt fond sur l'idée qu'un système d'inférence satisfaisant sur le plan logique ne peut, précisément, relever, chez l'homme, que de « l'acquis », non de « l'inné ». On voit, enfin, que ce que Turing appelle, ici, des « impératifs », ressemble fort aux « règles » analysées par Wittgenstein, c'est-à-dire à des « manières de faire » relevant d'un accord implicite de la part des acteurs, « apprises » et susceptibles d'être modifiées.

Telle est donc, selon Turing, la machine qui serait susceptible de participer au jeu de l'imitation, à l'égal d'un individu humain. On voit en quoi cette machine qui « apprend » se distingue de la machine à calculer proprement dite, dans le cas de laquelle, dit Turing, « l'objet est d'avoir une représentation mentale claire de la machine à tout moment du calcul » (Girard, 1995, 173). La

25 « S'il sait trouver la cause d'une faiblesse, il est probablement en mesure d'imaginer le type de mutation qui l'améliorera. » (Girard, 1995, 169).

26 Russell propose, pour éviter le paradoxe de l'ensemble de tous les ensembles qui ne s'appartiennent pas à eux-mêmes, qu'une classe ne puisse contenir que des éléments situés un niveau au-dessous d'elle.

27 On sait que la mise à l'écart progressive de Turing lors des travaux qui ont mené à la construction des premiers ordinateurs, a tenu à la différence d'approche entre les commanditaires de ces travaux – et les ingénieurs chargés de les réaliser – qui visaient la construction de super calculateurs, et Turing lui-même, qui entendait construire une machine « intelligente ».

« machine à calculer » peut toujours être ramenée à son « programme », c'est-à-dire à une suite bien définie d'instructions régissant analytiquement son action ; or on ne peut, pour « la machine qui apprend », parler de programme que pour l'état initial de la machine, et l'action de celle-ci est bien davantage fonction de son devenir que de cet état initial. L'hypothèse des machines qui apprennent, appuyée sur la description de la possibilité pour une machine « inorganisée » de s'organiser elle-même, est ainsi ce qui permet à Turing de passer de la machine à calculer, ou de la machine à jouer aux échecs de *Intelligent Machinery* **-** une machine qui ne sait que produire des suites de signes - à la machine qui peut participer au jeu de l'imitation proprement dit car un échange de paroles avec elle est possible - une machine qui énonce un discours *à la première personne*. Dans « Les ordinateurs et l'intelligence »*,* Turing fait donc l'hypothèse qu'une machine peut être conçue pour être insérée dans le procès collectif de manipulation de signes qui caractérise le langage humain et que cette machine s'inscrira ensuite elle-même dans ce procès car, comme les individus humains, elle *appliquera,* alors, des règles.

Conclusion

La réflexion de Turing, on le sait, est à l'origine de l'apparition de l'intelligence artificielle, née officiellement en 1956, lors d'un séminaire d'été du Dartmouth College en Angleterre, et portée sur les fonts baptismaux par John McCarthy, Marvin Minsky, Nathan Rochester et Claude Shannon[28].

C'est en se détachant des débats proprement philosophiques qui ont accompagné sa naissance et ce qu'on peut appeler son enfance, ou tout au moins en mettant ces débats entre parenthèses, que l'intelligence artificielle est devenue une discipline scientifique à part entière, dont le poids est désormais, et pour longtemps, déterminant dans le domaine des sciences appliquées.

La dimension proprement philosophique de la question a, en effet, assez rapidement pris la forme de la thèse dite de « L'IA forte » et de sa discussion. L'aspect assez académique de cette discussion ressort d'emblée du fait que la notion même d' « IA forte » a été proposée par l'un de ses principaux critiques, John Searle, et acceptée par tous, y compris par ses plus farouches défenseurs. Searle opposait une « IA faible », outil permettant d'étudier les comportements cognitifs, fournissant des modèles théoriques de compréhension des « comportements intelligents », à « l'IA forte », pour laquelle la machine en mesure de simuler un comportement intelligent *est* elle-même un être intelligent. « D'après l'IA faible, écrivait-il, la principale valeur de l'ordinateur dans l'étude de l'esprit, c'est qu'il est pour nous un outil très puissant. Ainsi il nous permet de formuler et de tester des hypothèses de façon plus rigoureuse et plus précise. D'après l'IA forte en revanche, l'ordinateur n'est pas simplement un outil d'étude de l'esprit ; l'ordinateur convenablement programmé est véritablement un esprit, en ce sens que des ordinateurs munis des bons programmes *comprennent* et ont d'autres états cognitifs. » (Searle, 1987, 354).

La thèse de l'IA forte s'inspire directement de l'article de Turing, mais elle met en jeu un ensemble de notions dont le caractère spéculatif n'est pas toujours bien circonscrit. Ainsi du concept de « simulation », certes solidement défini en informatique, théorique ou appliquée, mais qui reste très ouvert, voire lâche, du point de vue philosophique. En tant que « machine universelle », la « machine de Turing » est, par définition, une machine qui « simule » ; le jeu de l'imitation lui-même repose sur cette dimension de la machine, comme en atteste en particulier le fait que ses protagonistes, alors même que leurs échanges peuvent porter sur n'importe quel sujet, sont sans « corps », ou que, du moins, leur corporéité est réduite à la perception de signes écrits : c'est là le seul moyen par lequel les joueurs peuvent évoquer ce

28 Turing a rencontré Shannon à plusieurs reprises, en 1943 lors d'un séjour aux États-unis et en 1947 à l'occasion de la venue de Shannon à Cambridge.

qui relève du corps chez eux comme chez leurs adversaires. Dès le début des années 1960, Hubert Dreyfus montrait les difficultés liées à cette absence de corps véritable (Dreyfus, 1972). Qu'est-ce que peut être un corps *simulé* – et, en l'espèce, simulé sous la forme d'une suite de signes - ? L'échange entre locuteurs tel qu'il est sollicité par Turing dans son test met en jeu un vaste champ *d'implicite* qui renvoie précisément à la dimension corporelle de cet échange et Dreyfus avait beau jeu, dans les années 1960, de s'attaquer aux programmes d'intelligence artificielle qui, alors, présupposaient l'explicitation de toutes les actions de la machine ; sa critique allait, cependant, plus loin, puisqu'il refusait la possibilité même, défendue par Turing dans « Les ordinateurs et l'intelligence », de l'implicite pour une machine. Ces questions ont entraîné des débats très vifs et peuvent donner lieu à des vertiges spéculatifs dont seul, sans doute, le retour à une inspiration wittgensteinienne – celle de Turing lui-même - peut préserver.

C'est, en effet, parce qu'il avait bien compris les difficultés liées à la notion de simulation que Turing proposait un « test ». La seule réponse possible à la question « Est-ce que les machines peuvent penser ? » est la construction effective d'une machine « pensante », et non pas seulement d'une machine qui « ferait comme si » elle pensait ; l'idée d'une « simulation » de la pensée est étrange : qu'est-ce qui, hors de la pensée elle-même, pourrait simuler la pensée ? Le test met en jeu ce qui est, selon Turing, la condition pour qu'on puisse décider qu'une telle machine a bien été construite, à savoir qu'elle ait avec un homme une conversation, ce qui implique un « je », une « première personne », de la machine. Tant que cette machine n'aura pas été construite, affirmer sa possibilité restera un objet de spéculation ou une visée générale de la recherche du même ordre, toutes choses égales, que l'unification de la physique quantique et de la relativité générale pour les physiciens.

La thèse inverse, à savoir qu'une machine « pensante » n'est pas possible, a, quant à elle, un fort coût philosophique : elle revient à affirmer qu'une machine, parce qu'elle est une machine, ne peut pas « apprendre », comme le font les individus humains, qu'elle ne peut pas s'insérer dans un procès collectif d'élaboration et d'application de règles ; la thèse renvoie, par là, au « mystère de la pensée », c'est-à-dire à la mise en jeu de présupposés métaphysiques qui sont précisément ceux dont Wittgenstein a montré qu'ils étaient au principe du « sens commun des mathématiciens », ces présupposés mêmes dont il a fourni une critique si décisive.

Bibliographie

Dreyfus, H. (1972), *What Computers Can't Do : : a critique of artificial reason*, Harper & Row.

Girard, J. - Y. (1995), *La machine de Turing*, Paris, Seuil.

Glock, H. -J. (1996), « Mathematical Proof », dans *A Wittgenstein Dictionary*, Oxford, Blackwell Publishers.

Hardy G. H. (1929), « Mathematical Proof », dans *Mind*, Vol. 38, N° 149, Jan 1929. http://wwwprof.uniandes.edu.co/~amartin/cursos/filomat/bibliografia/Hardy.pdf

Hilbert, D. (1926), « Über das Unendliche », *Mathematischen Annalen*, 95, p. 170,171, trad. fr/all. Ladrière, J., dans Ladrière, J. (1957), *Les limitation internes des formalismes*, Gauthier-Villars.

Hodges, A. (1988), « Alan Turing ou l'énigme de l'intelligence », Paris, Payot.

Searle, J. (1985), *Du cerveau au savoir,* Paris, Hermann.

Searle, J. (1987), « Esprits, cerveaux et programmes », dans Hofstadter, D., Dennett, D. (eds), *Vues de l'esprit,* Paris, InterEditions.

Turing, A. (1937), « Théorie des nombres calculables, suivie d'une application au problème de la décision », trad. fr/angl. Basch, J., dans Girard, J. - Y., (1995), *La machine de Turing*, Paris, Seuil.

Turing, A. (1950), « Les ordinateurs et l'intelligence », trad. Fr/angl. Blanchard P., dans Girard, J. - Y., (1995), *La machine de Turing*, Paris, Seuil.

Turing, A. (2001), « Systems of Logic Based on Ordinals », dans *Collected Works of A. M. Turing, Mathematical Logic,* Amsterdam, North-Holland.

Turing, A. (1993), « Intelligent Machinery »*,* dans *Collected Works of A.M. Turing, Mechanical Intelligence,* Londres, North Holland, 1992.

Wittgenstein, L. (1995), *Cours sur les fondements des mathématiques. Cambridge, 1939*. Mauvezin, Editions T.E.R.