

La version finale de cet article est parue dans F. Nef et D. Vernant, dir., *Les années 1930 : réaffirmation du formalisme*, Vrin, 1998, pp. 1-41.

TURING : PENSÉE DU CALCUL, CALCUL DE LA PENSÉE

Turing est à l'origine de deux idées qui peuvent sembler contradictoires, puisque l'une conduit à un sévère confinement du formalisme, et l'autre à une considérable extension de son domaine. Ainsi formulée, cette opposition n'est qu'un procédé pour dramatiser une situation conceptuelle certes complexe mais libre de contradiction patente. Il serait néanmoins rapide de conclure à l'inverse que toute tension dans la pensée et dans la postérité de Turing n'est que le fruit d'une imagination philosophique friande de paradoxes : l'abondant corpus des études consacrées à Turing depuis une quinzaine d'années le montre, l'œuvre de Turing, dans son exceptionnelle originalité, porte au moins en germe des pensées virtuellement dissonantes.

Le texte que l'on va lire n'a aucunement l'ambition de proposer une synthèse résolutoire. Il se situe pour l'essentiel largement en deçà d'une tentative de conciliation définitive¹. Il se propose néanmoins de défendre l'idée d'une continuité essentielle, mais tente surtout de clarifier les enjeux et de préciser les contributions de Turing au projet formalistes². L'article suit un plan très simple : la première partie,

¹ A vrai dire, il n'est pas certain qu'il soit possible d'engager un tel travail sans entrer assez profondément dans la problématique logico-mathématique inaugurée par Hilbert, et ce n'est pas le lieu de le faire.

² Deux livres clés, dans un océan de publications et de manuels : Kleene 1952, Herken 1988. Pour les philosophes, Boolos & Jeffrey (1980) demeure une excellente présentation générale.

Turing : pensée du calcul, calcul de la pensée

destinée surtout au lecteur peu averti, présente la problématique dans ses grandes lignes ; les deuxième et troisième portent respectivement sur les articles de 1936 et de 1950, sur lesquels se fonde principalement l'attribution à Turing de thèses apparemment contradictoires sur le formalisme.

1. DE 1936 À 1950 : TENSIONS

Rappelons d'abord de la manière la plus ramassée, pour y revenir ensuite, de quoi il est question dans les deux articles les plus célèbres de Turing.

En 1936³, le jeune logicien fit paraître dans une revue mathématique un article dans lequel il proposait d'abord une « définition » mathématique de calcul « effectif »⁴. Cette définition était fondée sur un concept abstrait de machine, inventé par Turing. L'adéquation de la définition proposée reposait sur une thèse portant sur les processus psychologiques mis en jeu pendant l'activité mathématique. Les propriétés de la nouvelle machine étaient étudiées pour elles-mêmes. Enfin, ces notions étaient employées pour résoudre par la négative le problème de la décision ou *Entscheidungsproblem* de Hilbert : il n'existe pas de méthode générale pour déterminer si une formule du calcul de prédicats est prouvable ou non — partant, la même indécidabilité frappe les mathématiques dans leur ensemble.

En 1950, Turing publia dans une revue philosophique un article dans lequel il s'efforçait de montrer qu'il était concevable que l'on puisse dire un jour de certaines machines qu'elles sont capables de penser ou encore de faire

³ Il s'agit de Turing 1936-37. L'histoire de la publication de cet article est compliquée, chevauchant, pour différentes raisons, 1936 et 1937. Il n'est pas abusif ici de simplifier, comme on le fait d'ailleurs très souvent.

⁴ Les guillemets sont autant de précautions contre les risques d'objection prématurée de la part des *cognoscendi*. Les autres peuvent en faire abstraction.

preuve d'intelligence. Schématiquement, sa stratégie comportait deux volets, le premier conceptuel et négatif, consistant à priver de ses armes le bon sens qui veut qu'il soit contradictoire ou grotesque ou simplement faux d'attribuer de l'intelligence à une machine quelle qu'elle soit ; le second empirique et positif faisant valoir la plausibilité de l'hypothèse que les ordinateurs soient un jour capables d'un comportement intelligent.

L'un des effets de l'article de 1936 était de marquer les limitations du formalisme (tel que le concevait notamment Hilbert, tel qu'en gros nous le concevons aujourd'hui encore) et dans le même mouvement celles des machines. L'un des effets de l'article de 1950 était de montrer que contrairement à d'anciens préjugés, le domaine de la machine s'étendait potentiellement bien au-delà de l'exécution de tâches « mécaniques » — au sens de « stupides », n'impliquant ni pensée ni volonté —, et de manière concomitante que le formalisme concernait bien plus que la pure pensée mathématique ; Turing annonçait ainsi rien de moins que le programme de ce qui allait prendre quelques années plus tard le nom d'« intelligence artificielle ».

C'est donc ainsi que se précise quelque peu la contradiction apparente qui était annoncée. Une deuxième opposition vient d'une certaine manière l'encadrer. En amont : dans l'article de 1936, Turing veut justifier la réduction de la calculabilité « humaine » — la propriété, pour une fonction, d'être effectivement calculable par l'homme — à la calculabilité par machine ; pour cela il se fonde sur une analyse du processus psychologique du calculateur humain. En aval : dans l'article de 1950, renversant le rapport, Turing propose de voir dans le calcul mécanique un modèle de tout processus procédant de l'intelligence humaine. Ne fait-il pas avancer son bateau en soufflant dans la voile ? Ne fonde-t-il pas une conception du tout de l'intelligence à partir d'une conception d'une de ses parties, la machine ne servant que de transition, voire de travestissement, dans une induction téméraire ? A moins que

l'opération ne consiste au contraire à aller de la machine à calculer, « projetée » d'entrée de jeu sur le calculateur humain, à la machine intelligente, l'homme servant cette fois d'intermédiaire dans une induction portant sur les capacités des machines ?

Une dernière perplexité naît de la manière dont la matérialité semble tantôt absente, tantôt présente chez Turing. Ce qui chez l'homme peut selon Turing être imité par la machine est *conceptuellement* indépendant du corps humain dans sa spécificité — c'est ce « filtrage » du corps qui rend nécessaires la mise en scène et certaines règles du « jeu de l'imitation » dans l'article de 1950. Or tant la réduction proposée en 1936 du calculable humain au calculable mécanique, que l'ambition annoncée en 1950 d'une intelligence mécanique indiscernable de l'intelligence humaine, vont fortement dans le sens d'une théorie matérialiste de l'esprit.

Nous allons voir maintenant que le parcours de Turing, examiné d'un peu plus près, ne laisse place à aucune incohérence. Et que là où il peut sembler aujourd'hui s'être trompé, ce n'est pas conceptuellement mais empiriquement : sa mise en rapport de l'homme et de la machine, de la pensée et du calcul sont irréprochables logiquement ; là où nous sommes le plus fondés aujourd'hui à estimer que Turing s'est égaré, c'est dans l'évaluation de la complexité de l'apprentissage humain, et plus largement des rapports entre l'homme, penseur, organisme, avec l'environnement. Il était d'une époque dont les meilleures théories psychologiques étaient behavioristes et associationnistes, et il en fut sur ce sujet la victime. Cependant, il avait donné à ses hypothèses une forme suffisamment précise pour que la fausseté de certaines puisse apparaître clairement à la lumière du progrès scientifique et de l'expérience. Et il aurait sans doute été le premier, s'il avait vécu assez longtemps (il mourut à la veille de la naissance officielle des sciences cognitives, alors qu'il aurait pu les accompagner jusqu'à ce jour), à repérer ses erreurs et modifier sa conception aussi radicalement que

nécessaire. Quoi qu'il en soit, c'est la cohérence du cœur de sa doctrine qui en fait à la fois aujourd'hui un passage obligé de la réflexion sur la pensée, et un phare pour la route à venir, si différente qu'elle soit sans doute de ce qu'il anticipait il y a un demi-siècle.

2. RETOUR SUR L'ARTICLE DE 1936 : MACHINE ET CALCUL

C'est pour résoudre l'*Entscheidungsproblem* que Turing entreprit le travail rapporté dans l'article de 1936. En l'oubliant, comme on le fait souvent lorsqu'on pense à l'auteur comme inventeur de la machine de Turing, de l'ordinateur ou de l'intelligence artificielle, on occulte une continuité essentielle de sa pensée.

Tout problème de décidabilité a la forme suivante : soit (Q_i) une famille paramétrée de questions, l'indice i parcourant un ensemble I , en général infini ; existe-t-il une méthode algorithmique transformant la donnée d'une valeur i de l'indice en réponse à la question Q_i ? Une solution *positive* au problème est apportée par un algorithme particulier (ou par une réduction effective du problème à un autre problème de décidabilité pour lequel une solution positive est avérée). Or reconnaître qu'une méthode donnée est algorithmique (ou effective) ne pose pas de difficulté conceptuelle immédiate : c'est, pour le mathématicien, de l'ordre du constat. Il ne lui est pas nécessaire de pouvoir dire en vertu de quels *principes généraux* la méthode proposée peut être dite algorithmique pour savoir que le problème est résolu dans le sens positif. Au contraire, résoudre le problème par la négative, donc montrer qu'il n'existe pas de méthode algorithmique pour répondre à Q_i à partir de la donnée de i , consiste à prouver qu'une méthode⁵ ne peut à la

⁵ Ce terme étant entendu ici comme neutre, c'est-à-dire n'impliquant pas l'effectivité.

fois être algorithmique et conduire de i à Q_i , et pour cela la connaissance des principes de l'algorithmicité est indispensable.

C'est dans la recherche de ces principes que se sont engagés, simultanément et indépendamment, Church à Princeton et Turing à Cambridge. Or s'interroger sur les principes de l'effectivité, c'est aborder la question de la pensée d'une manière bien plus profonde et directe qu'on ne pouvait le concevoir avant Turing, et que bien des gens aujourd'hui encore, restés sur ce plan à l'ère pré-turingienne, ne le comprennent. Hâtons-nous de rappeler que Turing ne faisait aucunement fi de l'intuition, à laquelle il accordait au contraire une grande importance ; mais dans l'opération du mathématicien elle ne suffit pas à produire la démarche calculatoire. Et comme le montrera l'article de 1950, Turing avait parfaitement compris que le rôle du calcul, au sens où il l'a fixé en 1936, ne perd en rien de son importance capitale, quoiqu'il puisse en être de la réalité de fonctions non mécaniques ou non calculatoires de l'esprit, et de la possibilité de les reproduire dans une machine.

Mais nous anticipons. L'essentiel pour l'instant est que dès 1936 Turing est préoccupé du « fonctionnement » de l'esprit, et se livre, à des fins métamathématiques donc épistémologiques, à une investigation ressortissant de la psychologie théorique. Plus téméraire peut-être que des logiciens expérimentés tels que Gödel et Church⁶, Turing part à la recherche d'une caractérisation objective de procédure effective, et au lieu de passer par le détour de la mathématique, c'est-à-dire le produit, va au cœur de la question, qui touche à l'activité du mathématicien, c'est-à-dire la production, ou le procès. Quel est le *comportement* du mathématicien lorsqu'il calcule (le « *computeur* »⁷) ?, se

⁶ Gandy (1988) conjecture que l'innocence de Turing l'a servi : « Let us praise the uncluttered mind. » (p. 85)

⁷ Nous traduirons ainsi l'anglais « *computer* » pris dans son sens originel d'homme qui calcule. Nous éviterons ainsi à la fois l'emploi d'un mot anglais devenu lui-même ambigu dans ce contexte, et celui de son

demande Turing : « *What are the possible processes which can be carried out in computing a number ?* »⁸. En posant la question ainsi, il accomplit trois gestes d'un seul mouvement : *primo* il récuse l'interdit antipsychologiste prononcé par Frege et Husserl ; *secundo*, il adopte l'approche behavioriste de son époque ; *tertio*, il la dépasse en mêlant dans son analyse mouvements de la *pensée* et mouvements du *corps*, ou encore mouvements internes cachés et mouvements externes publics. Il va sans dire qu'il ne se soucie pas en la circonstance d'énoncer des principes (n'ayant sans doute guère conscience de prendre position dans des débats philosophiques), encore moins de les justifier ou de revendiquer pour eux une portée générale. Mais il les illustre de façon éclatante.

Poser la question d'une façon nouvelle n'était cependant pas *ipso facto* y répondre. Il fallait aussi progresser à partir de l'autre rive, pour ainsi dire, et se mettre en quête d'une conception suffisamment abstraite et générale de machine pour donner corps à l'idée classique, définitoire, de l'homme calculant (le *computeur*) comme agissant de façon « mécanique ». C'est ainsi que Turing aboutit à ses « *computing machines* ». Le chemin que sa propre pensée a suivi pour concevoir d'une part la machine, de l'autre le calculateur humain, fut sans doute fait d'allers-retours. Il n'importe, quoique puissent en penser ceux qui, comme Wittgenstein, accusent ses machines d'être trop humaines, ou ceux qui à l'inverse pensent qu'il a introduit subrepticement la machine dans l'homme. Ce qui compte, comme pour le pont une fois construit, c'est qu'en adossant l'une à l'autre ses deux conceptions, il leur a conféré à toutes deux une solidité et une permanence absolument inédites en

correspondant français « calculateur ». Gandy, pour résoudre le même problème, mais en anglais, a inventé le mot « *computer* ».

⁸ Turing 1936-7, § 8 : « Quels sont les processus que l'on peut exécuter en calculant un nombre ? ». La référence à un « nombre » s'explique par le fait que Turing a choisi de s'intéresser d'abord aux nombres réels calculables (plutôt qu'aux fonctions d'entiers).

philosophie. De grands commentateurs, dont Gödel lui-même, ont remarqué qu'il n'était jamais arrivé auparavant qu'une « notion épistémologique importante » reçoive une définition (mathématique) « absolue », c'est-à-dire « indépendante du choix du formalisme »⁹. La notion en question est celle d'effectivité — mais on peut en dire autant des notions de calculateur humain¹⁰ (une abstraction, certes intimement liée à celle de procédure effective, mais néanmoins distincte) et de machine à calculer — bref, de calculateur aux sens tant pré- que post-turingiens.

Le succès de Turing, quant aux objectifs qu'il s'est fixés en 1936, est complet et immédiatement reconnu par Gödel et par Church. L'équivalence entre humainement calculable et calculable par machine de Turing, que l'auteur propose en soulignant qu'il ne peut s'agir d'un théorème¹¹, emporte une adhésion sans réserve de la plupart des contemporains. Elle conduit non seulement à la stabilisation de la notion informelle de fonction calculable, mais aussi, comme y insiste Gödel, à celle de système formel — avec Turing l'essence d'un tel système devient manifeste : elle réside dans le caractère effectif de sa syntaxe. L'effectivité est désormais une notion objective et stable.

Il est particulièrement remarquable que la thèse de Church n'ait pas emporté une adhésion aussi complète, alors qu'elle est superficiellement très comparable (au point que

⁹ Gödel (1946).

¹⁰ Ce serait un grave contresens sur Gödel que de lui attribuer ici indirectement la pensée que le *mathématicien* (*a fortiori* le penseur en général) reçoivent ainsi par ricochet une définition mathématique absolue : c'est du *calculateur* humain qu'il s'agit, Gödel étant précisément de l'opinion que le mathématicien ne s'y réduit pas.

¹¹ Signalons au passage l'opinion (minoritaire) de Gandy (1988), qu'en ce qui concerne la rigueur démonstrative, la thèse de Turing ne le cède en rien aux théorèmes usuels ; seul, selon Gandy, le *sujet* de la thèse est inhabituel.

l'on parle couramment de « thèse de Turing-Church »¹²). On a esquissé à l'instant une explication très générale d'ordre épistémologique. W. Sieg, quant à lui, se place sur un plan plus technique et fournit une explication beaucoup plus approfondie, qui va dans le même sens : il fallait partir, selon Sieg, comme l'a fait Turing, des processus sous-jacents au calcul pour surmonter ce qu'il appelle l'« obstacle majeur » auquel se heurte la thèse de Church¹³, et qui s'apparenterait à une pétition de principe.

En voilà assez sur ce que l'on peut considérer comme les « effets directs » de l'article de 1936, à savoir la stabilisation canonique de la notion d'effectivité et concurrentement de celle de système formel, ainsi que la solution négative de l'*Entscheidungsproblem*. Venons-en aux effets indirects, ceux qui s'exerceront à plus long terme. Ils sont plus considérables encore.

Le premier n'est pas facile à décrire avec précision. Il consiste en un déplacement du centre de gravité de la logique, qui s'éloigne des mathématiques abstraites, platoniciennes pour se rapprocher des mathématiques constructives et de l'informatique, et qui en même temps s'éloigne de l'esprit, de l'homme comme idéal et parfait raisonneur, pour se rapprocher de la machine : la logique devient une sorte de « physique théorique » des machines — machines artificielles voire naturelles¹⁴.

Ce premier effet est indissociable du second, plus évident et mieux connu. Il s'agit d'une façon entièrement

¹² L'emploi de ce nom composé n'implique pas nécessairement l'oubli d'une distinction possible : certains (dont l'auteur) désignent ainsi la *conjonction* de deux thèses, l'une renvoyant à la pensée de Church, l'autre à celle de Turing.

¹³ J'ai pris connaissance de l'important travail de Sieg (1991) après avoir préparé mon intervention au colloque de Saint-Malo. Je dois donc me contenter ici de renvoyer le lecteur à cet article très riche.

¹⁴ Cf. Wagner (1994).

nouvelle de considérer les machines¹⁵. D'abord, une machine produit des inscriptions, ou encore ce que Newell et Simon appelleront bien plus tard¹⁶ des symboles physiques ; elle produit une matière dont la matérialité n'intervient que de façon abstraite, minimale ; et le mode de présentation de la machine n'en retient que les aspects pertinents à cet égard. Ce n'est évidemment pas la première fois que l'on décrit une machine en faisant abstraction de *certaines* de ses propriétés matérielles — la composition chimique, ou les dimensions, dans certaines limites évidemment, ou encore la couleur, voire la forme exacte, peuvent être absentes de la spécification d'une poulie. Mais il n'était pas arrivé que l'on fasse abstraction de *toutes* les propriétés matérielles d'une machine pour n'en préciser que des propriétés fonctionnelles. La chose est si surprenante que beaucoup ne voient encore aujourd'hui qu'un abus de langage dans l'emploi du mot « machine » pour désigner ces entités immatérielles : ce ne sont à leurs yeux que des idéalités mathématiques.

En deuxième lieu, Turing présente non pas une seule machine, mais une infinité. Ici encore, la nouveauté ne réside pas dans le simple fait de paramétrer un plan de machine : le levier est paramétré par ses dimensions, la roue dentée par le nombre de dents, la machine à vapeur par sa puissance, etc. C'est le paramétrage qui est inédit, et qui induit une partition en classes d'équivalence dans laquelle chaque classe C se caractérise à deux niveaux :

- (i) la table d'instructions \mathcal{T} (des machines appartenant à C) est un ensemble de propriétés fonctionnelles patentes ;

¹⁵ Ce n'est pas le lieu de discuter des précurseurs de Turing, en particulier de Babbage et de ses disciples. Rappelons seulement que, comme l'attestent Hodges (1983) et Gandy (*loc. cit.*), Turing ne découvre les idées de Babbage que dans les années 1940.

¹⁶ Newell & Simon (1975).

(ii) cette table est la source de la disposition (cachée, quoique parfois découvrable) à calculer une fonction donnée f .

L'appartenance d'une machine particulière \mathcal{M} à une classe d'équivalence donnée C est assurée par la relation d'instantiation : les propriétés physiques particulières de \mathcal{M} occupent par rapport à ses dispositions à réagir aux stimuli la place que les propriétés fonctionnelles du niveau (i) occupent par rapport à la disposition de niveau (ii) à calculer la fonction f . Cette égalité de rôles est assurée par l'architecture de la machine : la machine est ainsi faite que lorsqu'elle est mise en branle par un stimulus représentant (interprété comme) une valeur n de la variable, ses transitions se décrivent fonctionnellement comme les étapes du calcul de $f(n)$ selon la table \mathcal{T} . Ainsi, l'appartenance de \mathcal{M} à C , qui se décide au niveau (i), *explique* la conformité des comportements de \mathcal{M} à la disposition de niveau (ii).

Dans ces conditions, un stimulus particulier est la *cause* du processus que suit la machine, en particulier de l'inscription par lequel il se termine éventuellement, et l'interprétation canonique du stimulus est la *raison* de l'interprétation canonique de l'inscription finale. L'interprétation est arbitraire, en un sens ; mais elle est canonique dans la mesure où elle est telle que la raison dont il vient d'être question vaut, moyennant l'architecture de la machine, explication du processus physique.

Comparons : un hâchoir à viande (d'un type donné) peut être en acier ou en fonte (accordons-le) — fonctionnellement donc l'objet d'acier et l'objet de fonte sont indiscernables. L'un comme l'autre peut transformer un morceau de 100 grammes de bœuf (resp. de porc) en une masse de 100 grammes de hâché de bœuf (resp. de porc). Le niveau (i') correspondant ici au niveau (i) pour la machine de Turing est occupé par la fonction « hâcher », vue comme un ensemble de propriétés ; le niveau (ii') par la disposition

qui en est la conséquence à transformer des morceaux de viande en hâché. L'appartenance d'une machine donnée \mathcal{M}' à la classe C' des hâchoirs du type considéré est assurée par le fait que les propriétés pertinentes de \mathcal{M}' instancient les propriétés de niveau (i') ; et de cela résulte que le processus que provoque dans \mathcal{M}' l'intromission d'un morceau de viande est un hâchage particulier qui actualise la disposition à hâcher de niveau (ii').

Qu'est-ce qui distingue, dans ces conditions, une (classe) C de machines de Turing de table \mathcal{T} d'une classe C' de hâchoirs d'un type donné ? A vrai dire, la question n'est pas des plus faciles. On ne peut évidemment pas faire appel à l'idée fonctionnaliste, puisqu'elle s'applique dans les deux cas. Quantitativement, il est intuitivement clair que la distance entre classe et élément est beaucoup plus grande dans le cas des calculateurs que dans celui des hâchoirs, et corrélativement que la classe contient une bien plus grande variété de dispositifs matériels dans le premier cas que dans le second. De même, la classe des stimuli est plus vaste, et la classe des couples entrée-sortie plus variée. Pour passer de ces intuitions quantitatives à des critères qualitatifs, deux voies semblent prometteuses : l'idée que la machine de Turing, contrairement au hâchoir, opère sur des symboles, et l'idée que le plan — c'est-à-dire la caractérisation de la classe — est de nature mathématique, dans le premier cas, non du second.

Pour donner corps à la première de ces idées, il serait bon de disposer d'une conception claire de ce qu'est un symbole, ou encore de ce qu'est l'information — nous n'en sommes pas là ; cela n'oblige pas à renoncer à l'idée : elle reste seulement inaboutie. Quant au caractère mathématique du plan, il est vrai qu'il suffit à différencier le calculateur du hâchoir, mais non pas d'autres dispositifs, ceux par exemple qui sont régis par un système d'équations différentielles, comme le système solaire. Pour répondre à cette objection, il faut sans doute rapprocher les deux idées : ni le hâchoir ni le

système solaire ne se prêtent à une combinatoire productive¹⁷, contrairement à la machine de Turing, et cette dernière le doit à la propriété qu'elle a d'être naturellement descriptible comme machine à un nombre fini d'états discrets.

Ce n'est là qu'une ébauche, mais elle suffit à marquer l'importance — et la difficulté inattendue — du concept de machine introduit par Turing. Ce double aspect apparaît mieux encore lorsqu'on pose une autre question : qu'est-ce qui distingue la classe C des machines de Turing de la classe C' des hâchoirs ? Que celle-là soit, intuitivement, infiniment plus riche que celle-ci, et qu'elle soit clairement délimitée contrairement à l'autre, c'est certain. L'essentiel est cependant ailleurs : C est uniformément paramétrée, en un sens qui permet — notamment¹⁸ — de définir une machine de Turing universelle ; de là aussi la démonstration de l'indécidabilité du problème de l'arrêt et par voie de conséquence la solution négative de l'*Entscheidungsproblem*. Il est frappant que Turing n'ait pas mis en vedette ces résultats¹⁹, mais leur portée, on le sait, est immense. D'un côté Turing confirme le résultat de Church, en l'asseyant

¹⁷ Au sens technique : l'ensemble des combinaisons est infini et finiment engendré.

¹⁸ Nous taisons les propriétés plus profondes de ce paramétrage, elles ne semblent pas nécessaires dans une discussion élémentaire. Voir à ce sujet, entre autres, Sieg (*loc. cit.*).

¹⁹ Même le titre de l'article de 1936 le montre : il n'annonce pas la résolution de l'*Entscheidungsproblem* par le recours à une notion nouvelle de calcul ; la résolution du fameux problème est présentée comme une application. L'intuition de Turing sur le résultat d'indécidabilité ne l'a pas égaré : si important qu'il soit, on ne peut que reconnaître aujourd'hui, je crois, qu'il pâlit en comparaison avec la simple invention du concept de machine de Turing et ses conséquences pour les notions d'effectivité et de système formel. En revanche, peut-être voudra-t-on considérer qu'il n'a pas mesuré en 1936 toute l'importance de l'uniformité du paramétrage et de la machine universelle ; si c'est le cas, son travail sur l'ordinateur britannique à la fin de la guerre montre qu'il s'est rapidement rattrapé sur ce point.

mieux encore, et assigne ainsi une limite nouvelle au formalisme, après les résultats d'incomplétude de Gödel. De l'autre il prépare le terrain pour l'article de 1950 : l'existence d'une machine universelle sera un argument de poids, entre ses mains comme celles de ses continuateurs, en faveur de la possibilité d'une machine intelligente. Mais le résultat d'indécidabilité lui-même montre la machine sous un jour qui déjoue les attentes d'hier, et peut-être parfois celles d'aujourd'hui : que la machine soit incapable de prévoir l'arrêt de ses congénères ne marque pas tant ses limites qu'elle n'illustre la complexité des machines, leur « épaisseur » qui les rend imprédictibles autant par l'homme que par la machine elle-même²⁰ !

L'auteur de l'article de 1936 apparaît en fin de compte bien davantage comme un précurseur du cognitivisme que comme l'homme de la limitation du formalisme. Dans ce dernier rôle, il n'est d'abord pas le seul ni le premier. Ensuite, comme on vient de le dire, son résultat de limitation se présente comme une arme à double tranchant²¹ : la limitation du formalisme a pour raison profonde son « épaisseur » intrinsèque, et ne le place pas de manière évidente en position d'infériorité devant l'homme en tant que l'autre du formalisme. Enfin et surtout, Turing fait apparaître en un seul coup une forme nouvelle de psychologie, ni behavioriste ni essentialiste, une idée nouvelle de la machine, et un type inédit de rapport entre esprit et machine — trois conceptions qui forment

²⁰ Ce mélange de singulier et de pluriel à propos de machines(s) est l'indice d'un problème qu'il n'est pas possible d'aborder dans ce court article : celui des rapports d'une machine quelconque à la population de toutes les machines et à d'autres machines particulières ; l'existence de machines universelles ne constitue à cet égard qu'une (importante) complication supplémentaire.

²¹ Webb (1980) fait de cette ambivalence l'axe de sa réfutation des arguments anti-formalistes développés par Lucas (1961) à partir des théorèmes d'incomplétude de Gödel.

solidairement le socle sur lequel les sciences cognitives seront érigées.

3. L'ARTICLE DE 1950 : INTELLIGENCE ET MATIÈRE

Entre l'article de 1936 et celui de 1950, c'est peu dire que le temps a passé : dans l'intervalle Turing est devenu, pour le compte des services secrets britanniques, le « cerveau » du décryptage du chiffre des forces navales allemandes pendant la guerre, et l'inventeur d'une série de machines de décryptage automatique²². Dès la guerre finie, il participe à la conception des premiers ordinateurs, d'abord dans le cadre militaire, puis à l'université de Manchester. Tenu par le secret pendant de longues années, placé au centre d'opérations vitales pour son pays, Turing s'est éloigné définitivement de l'idéal du mathématicien pur, mais il reste un penseur.

Cette digression biographique, qui peut sembler déplacée, permet en peu de lignes de signaler l'extraordinaire changement de climat survenu entre les deux publications. La continuité des préoccupations de Turing n'en apparaît que plus frappante : les changements de discipline, de problématique, de cadre institutionnel ne font finalement que répondre à une nécessité intellectuelle — c'est le même projet que Turing poursuit avec de nouveaux moyens : fonder une psychologie de la pensée, ou, comme on dirait aujourd'hui, une psychologie cognitive.

L'article paru en 1950 dans *Mind*, intitulé « Computing machinery and intelligence », s'attaque à un problème différent à plusieurs égards de celui de 1936. D'une part, il ne comporte pas d'aspect directement mathématique ; d'autre part, il est à l'inverse entièrement immergé dans la psychologie théorique, et dans ce domaine il aborde une question beaucoup plus vaste. En 1936, en effet, Turing

²² Cf. Hodges (1983).

considérait l'homme calculant, le ordinateur, et se demandait comment décrire de manière systématique et exhaustive son activité mentale idéalisée. En 1950, c'est toute l'étendue des activités ou manifestations de l'intelligence qu'il prétend subsumer sous un schéma explicatif général.

Dans sa première entreprise, Turing pouvait s'appuyer sur l'expérience millénaire des algorithmes, dont s'était dégagée depuis longtemps une vision assez claire du genre d'activité *externe* à laquelle se livre le ordinateur, ou pour le dire d'un mot, de son comportement. Il restait à imaginer une contrepartie mentale, interne, « pilotant » ce comportement. La nature séquentielle discrète de l'algorithme conduisait de manière assez naturelle à l'hypothèse d'une suite d'états mentaux discrets, s'enchaînant selon des transformations élémentaires diverses. L'idée qu'un algorithme peut être appliqué « mécaniquement », sans intervention de l'intelligence ou de l'imagination, se traduisait par l'hypothèse d'un répertoire fini, commun à tous les algorithmes, d'opérations élémentaires, chaque algorithme particulier se caractérisant par un certain ordonnancement de ces opérations.

Dans quelle mesure cette conception pouvait-elle s'étendre aux comportements intelligents ? En aucune façon, semblait-il : le ordinateur n'est-il pas l'homme qui choisit de se passer de son intelligence pour accomplir une tâche mécanique ? Sans doute, mais depuis qu'en 1936 Turing avait procuré à la machine, précisément à partir d'une réflexion sur le caractère mécanique de l'exécution d'un algorithme, un deuxième bulletin de naissance, la machine nouvelle était devenue une réalité, et, entre les mains des hommes, avait commencé à manifester des potentialités insoupçonnées : elle se comportait moins « mécaniquement » que son modèle, et pouvait servir à son tour de modèle de l'homme *non* calculant.

Mais avant de pouvoir proposer une pareille hypothèse, il fallait préparer les esprits à la recevoir. C'est à cette propédeutique qu'est consacré le début de l'article de 1950.

Il s'agissait de rendre séparable par la pensée l'intelligence de quelque chose d'autre, qui appartiendrait à l'homme et serait la cause de son intelligence, en sorte de rendre imaginable qu'une entité autre que l'homme puisse posséder ce quelque chose, lequel serait la cause que cette autre entité posséderait l'intelligence. Cette séparation accomplie, la question de savoir si l'idée d'intelligence non humaine²³ est douée de sens — ce dont alors, et sans doute encore aujourd'hui, beaucoup doutaient — serait remplacée par une question *empirique* : est-il matériellement possible qu'une entité non humaine possède le « facteur causal » de l'intelligence ?

Turing s'inspira-t-il, pour résoudre la difficulté, de l'exemple si particulier qui lui avait réussi en 1936 ? Toujours est-il qu'il imagina d'inventer ce qui serait à l'intelligence ce que les gestes de l'homme calculant sont à ses processus mentaux. Remarquons que les deux termes moyens (intelligence, gestes du ordinateur) étaient identifiés avant Turing, même si l'on n'en possédait pas une idée claire et distincte. En 1936, Turing inventait, ou découvrait, l'un des termes extrêmes : le processus mental du ordinateur. Il peut maintenant compléter l'homologie et rechercher le dernier terme : ce sera le comportement intelligent. Par quoi il ne faudra pas comprendre un comportement qui a la qualité d'être intelligent, comme il pourrait être distingué ou incohérent ou névrotique — de même que par comportement mécanique, on ne doit pas entendre (simplement) un comportement qui a la qualité

²³ Qu'il ne faut surtout pas entendre au sens où elle s'opposerait à l'intelligence divine, ou à l'intelligence animale, etc. : il s'agit d'une intelligence qui a toutes les qualités de l'intelligence humaine, sans être une qualité ou une capacité *de* l'homme. On voit combien la question du sens, abruptement posée, est délicate : comment parler d'une intelligence humaine non humaine ? Un procédé devenu commun consiste à comparer l'intelligence à une substance chimique : l'insuline de synthèse a toutes les qualités de l'insuline humaine, sans être un produit du corps humain. Mais pour admettre la pertinence de la comparaison, ne faut-il pas avoir déjà accompli la séparation en question ?

phénoménale des mouvements d'une machine, qui ressemble au comportement d'une machine, mais d'abord et de manière essentielle un comportement qui est *la manifestation visible d'un mécanisme caché*. Un comportement intelligent sera compris comme la manifestation visible de l'intelligence : ce sera l'effet dont la cause est ce quelque chose qui produit l'intelligence.

L'exemple le moins discutable de comportement intelligent — à vrai dire le seul dont Turing disposait, et le meilleur dont nous-mêmes disposons encore aujourd'hui — est le comportement de l'homme en pleine possession de ses facultés. Il restait donc à montrer qu'il n'y a pas d'incohérence à imaginer une situation dans laquelle on observerait un comportement intelligent produit par une entité non humaine. Encore fallait-il procéder à une nouvelle séparation, au sein même du comportement humain : que ce soit une corde vocale qui vibre lorsqu'un homme dit « bonjour » ne contribue pas à l'intelligence de son comportement (l'homme sans cordes vocales qui se fait comprendre par l'intermédiaire d'un appareil exhibe un comportement équivalent sur le plan qui nous concerne). Il fallait donc que le comportement de l'entité non humaine ne risque pas d'être jugé défaillant sur le plan de l'intelligence du seul fait de différences non pertinentes.

C'est pour résoudre en une seule opération les deux séparations — l'effet de la cause, et le pertinent du non-pertinent — que Turing imagina l'une des plus célèbres expériences de pensée de la philosophie contemporaine, le « jeu de l'imitation ». Ce jeu se joue à trois : un être humain, une machine et un arbitre, isolés dans trois cabines opaques et communiquant entre eux par télétype. Chacun des deux premiers cherche à persuader le troisième qu'il est l'être humain, en répondant comme il l'entend aux questions que pose l'arbitre. Le jeu se termine lorsque l'arbitre propose une identification (« l'être humain occupe la cabine X »), ou s'il tarde, après un nombre d'échanges fixé à l'avance. Supposons que nous constatons, après une longue série de

parties, que dans une proportion importante, l'arbitre n'a pas donné la bonne réponse ; alors, demande Turing, ne serons-nous pas tentés de modifier nos habitudes de langage et dire de la machine qu'elle est intelligente ? Cette modification s'imposera, pense-t-il, quelles que puissent être aujourd'hui les objections à une pareille idée. Une telle rectification de l'usage courant se justifiera par le fait que l'arbitre s'est montré incapable de distinguer à coup sûr l'homme de la machine, alors que chacun a pu au cours du jeu exhiber toute forme imaginable de comportement intelligent sans devoir se trahir (dans le cas de la machine) ou pouvoir s'identifier (dans le cas de l'être humain) en montrant pour ainsi dire le bout de l'oreille. Bref, l'intelligence sera redéfinie sans violence contre le bon sens comme propriété causale d'un certain type de comportement, qu'il soit celui d'un être humain ou d'une autre sorte d'entité.

Naturellement, cette expérience est d'une grande complexité, et appelle des réglages conceptuels minutieux : Turing en est conscient et fournit quantité de précisions ingénieuses. L'interprétation des résultats (imaginaires) de l'expérience est encore plus délicate : d'innombrables commentateurs s'y sont employés — c'est devenu une sorte d'exercice obligé dans tout travail de nature épistémologique sur les fondements de l'intelligence artificielle. Mais trop souvent ils n'ont retenu de l'expérience que les deux questions suivantes : (1) Est-il concevable qu'une machine passe avec succès le « test de Turing »²⁴ — qu'elle trompe régulièrement l'arbitre au jeu de l'imitation ? (2) A supposer que ce soit le cas, serons-nous fondés à la considérer comme intelligente ?

²⁴ Expression que Turing lui-même n'emploie jamais, comme me l'a fait observer Jean Lassègue. Je lui dois de m'avoir mieux fait comprendre toute la difficulté de l'expérience de pensée de Turing, même si nous sommes en désaccord sur la nature des problèmes qu'elle pose. Les longs développements qu'il lui consacre dans sa thèse (1994) sont fort stimulants.

En un sens, ce sont bien les deux questions que Turing veut voir son lecteur se poser. Mais c'est sur le sens profond de ces questions que trop souvent à la fois l'on se méprend et l'on est infidèle à Turing.

Prenons la première. Il faut d'abord rappeler combien le contexte a changé depuis le moment où Turing écrit, et combien ce changement doit précisément à son article. Peu de gens savent en 1950 ce qu'est un ordinateur ; ils ne savent même pas que cela existe : les premières machines sont couvertes par le secret militaire ; les ordinateurs civils se comptent sur les doigts. Quant à imaginer ce dont ils sont capables, hormis les calculs de physique appliquée pour lesquels ils ont été conçus, seuls seraient en mesure de le faire quelques cybernéticiens et autres lecteurs de l'article de 1936, et sans en excepter les hypothétiques admirateurs de Babbage et Menabrea. C'est donc à tenter de faire comprendre ce que sont les « *digital computers* », et quelles sont leurs capacités potentielles, que s'emploie longuement Turing dans la seconde partie de son article. Seuls quelques pionniers sont en mesure de même saisir l'idée générale.

Dix ans plus tard, une partie du monde scientifique a pris conscience de l'existence des ordinateurs et de leurs extraordinaires capacités de calcul. Néanmoins, en dehors des fondateurs de l'intelligence artificielle, et peut-être de certains linguistes ou psychologues de leur entourage, personne n'imagine que les ordinateurs puissent jamais être autre chose que de monstrueux bouliers automatiques. La question de Turing n'est vraiment comprise de personne, hormis les initiés qui tout en répondant par l'affirmative sont déjà, sans en être bien conscients, décalés par rapport à ce qui, aux yeux de Turing, donnait son intérêt à la question, à savoir une certaine conception générale de la psychologie dont ils ont commencé à s'éloigner²⁵.

C'est seulement au début des années 1970 que le contexte change au point de modifier la perception de

²⁵ Cf. Newell (1983).

l'article de Turing et de le rendre véritablement audible. Les ordinateurs sont devenus les outils indispensables des sciences et de la gestion. L'intelligence artificielle est sortie des limbes : elle a ses manifestes, ses résultats, ses critiques. Enfin, les sciences cognitives ont commencé à réaliser une jonction effective en dehors d'une stricte allégeance à l'intelligence artificielle. La question de Turing devient alors : est-il concevable qu'on puisse un jour programmer un ordinateur en sorte qu'il passe le « test » de Turing ?, les méthodes de programmation étant relativement bien circonscrites et constituant l'appareil conceptuel de l'intelligence artificielle. C'est à ce moment-là seulement que la question prend un sens qui, sans être exactement celui de Turing, est relativement clair et saisissable en dehors du cercle des *afficionados* de l'intelligence artificielle.

Enfin, au début des années 1980, la conception dite classique ou symbolique de l'intelligence artificielle cesse d'être considérée comme l'unique manière de réaliser une intelligence de synthèse, elle est concurrencée par le « neurocalcul »²⁶. Une nouvelle complication surgit alors : peut-on considérer que le neurocalcul est la poursuite du projet de Turing par d'autres moyens, ou faut-il au contraire y voir un démenti potentiel à ses hypothèses ?

Ainsi, la question de la plausibilité d'une éventuelle victoire de la machine dans le jeu de l'imitation ne peut être considérée indépendamment du cadre dans lequel elle est posée, en particulier des hypothèses quant aux capacités informationnelles de la machine : pour Turing, c'était ses conceptions *psychologiques* bien particulières qu'il s'agissait de mettre à l'épreuve. Beaucoup de ses lecteurs en ont fait abstraction, se reposant sans doute sur leurs conceptions personnelles en matière de psychologie et les confrontant directement avec ce qu'ils peuvent deviner des capacités de

²⁶ Cf. pour les textes fondateurs Anderson & Rosenfeld (1988) et Pélissier & Tete (1995), et pour une mise en perspective épistémologique l'essai « Connexionnisme et cognition : à la recherche des bonnes questions », ce volume.

l'ordinateur. D'autres ont partiellement modifié les hypothèses de Turing sans y prendre garde, faisant de lui le premier représentant de l'intelligence artificielle classique. Il faut insister ici sur la trahison que constituent, chacune à sa façon, ces deux attitudes, car l'enjeu est celui du formalisme de Turing en tant que principe de psychologie générale.

Turing affirme bien, et défend dans la dernière partie de son article, une thèse formaliste, que la première catégorie de lecteurs rejette sans l'avoir seulement considérée. Mais ce formalisme est moins radical que celui des théoriciens de l'intelligence artificielle classique auxquels la seconde catégorie de lecteurs tend à l'assimiler. Cette affirmation ne sera certainement pas sans susciter des objections, et un article entier serait nécessaire pour tenter de la justifier. Il faudra se contenter ici d'en indiquer deux raisons. *Primo* Turing, anticipant sur ce point les objections de H. Dreyfus²⁷, notamment, admet qu'il puisse être impossible de formuler un ensemble complet de règles prescrivant le comportement dans toutes les situations possibles, même en se restreignant à une compétence particulière (telle que la conduite à tenir face à un feu rouge). C'est ce qu'il appelle « *the argument from informality of behaviour* »²⁸. Mais il distingue — déjà — les règles que l'on suit des lois auxquelles le « comportement » (mental) se conforme ; puisque il n'en déduit pas que l'affaire est désespérée ou absurde (ce qui le placerait soit dans le camp de Dreyfus, soit dans celui de Searle), c'est qu'il pressent que l'explication pourrait se situer à un niveau « subpersonnel », qu'on l'entende au sens « classique » ou au sens « subsymbolique » des connexionnistes à la Smolensky. En tout cas, cette intuition-là n'est pas partagée par les fondateurs de l'intelligence artificielle.

Secundo, Turing accorde une importance primordiale à l'apprentissage : il spécule que l'entreprise à laquelle il songe n'a de chance d'aboutir que si le bagage de l'enfant nouveau-

²⁷ Dreyfus (1972/79/93).

²⁸ Turing (1950), p. 452.

né est suffisamment léger pour être « programmé » à la main, l'essentiel de la compétence de l'adulte résultant de l'apprentissage en milieu naturel. On ne pourra s'étendre sur les faiblesses de la conception turingienne en la matière. L'important pour nous est que tant par l'importance qu'il accorde à l'apprentissage que par l'idée qu'il se fait de ce processus, Turing s'éloigne une seconde fois de l'intelligence artificielle et du formalisme strict : c'est plongé on ne peut plus directement dans la vie même, dans la recherche de voies évitant les punitions et recherchant les récompenses, que l'enfant acquiert les compétences qu'il lui manque. S'il est vrai que l'on *peut* ériger des théories formelles de l'apprentissage²⁹, l'idée que semble s'en faire Turing est fort peu formelle : elle est on ne peut plus concrète et située, sans qu'il soit possible de préjuger de la façon dont il eût précisé son intuition s'il en avait eu le loisir.

Venons-en maintenant à la seconde question : que pourrait-on conclure d'un éventuel succès de la machine au jeu de l'imitation ? Les réponses que l'on trouve en général sont de deux types. Certains pensent que quoi qu'il en soit des capacités manifestées par la machine, et qui lui permettent notamment de tromper l'arbitre, il lui manquera toujours quelque chose d'essentiel dont tout être humain est capable. C'est ce que Turing, qui a anticipé cette objection également, appelle le « X » ; il propose même une liste réjouissante de candidats à la dignité de X, liste qui témoigne, pour qui aurait la naïveté d'en douter, qu'il n'avait rien d'un théoricien désincarné. Comment entend-il répondre à ce genre d'objections ? Au cas par cas, en comptant que la succession des réfutations engendrera une conviction, en même temps qu'une méthode argumentative générale, qui aura pratiquement le même effet qu'une réfutation en bloc, jugée peu plausible.

Cette situation dialectique, curieusement, Turing l'a rencontrée en 1936 (mais il n'en fait pas mention en 1950 : le

²⁹ Cf. Osherson, Stob & Weinstein (1986).

parallèle ne lui paraissait peut-être pas digne d'être signalé). Lorsqu'il présente sa « thèse » selon laquelle toute fonction calculable par *computeur* (humain) l'est par machine (de Turing), il a dû faire face à des objections de la forme suivante : telle fonction X que l'homme sait calculer, je vous défie de trouver une machine qui la calcule. Après qu'une série suffisamment longue de tels défis ont été relevés, un « coup de main » s'acquiert qui permet de relever presque à coup sûr les suivants ; et c'est alors que l'on est fondé à conjecturer inductivement qu'il existe une correspondance régulière sous-jacente. Cette conjecture est précisément la thèse que Turing formule en 1936. Dans le contexte de l'article de 1950, il ne va pas aussi loin. Ce n'est pas, semble-t-il, qu'il ait personnellement des doutes quant à la vérité d'une thèse semblable, mais que le critique lui paraît être en proie à une confusion : si on lui montre un programme accomplissant son « X », « *one will not make much of an impression* »³⁰, car ce à quoi le critique songe en fait, c'est que la machine n'accomplit pas X *de manière consciente*. Autrement dit, ce n'est pas accomplir X dont le critique doute que la machine ait la capacité, mais de posséder la conscience. Or cette critique-là est d'une autre nature, et Turing lui consacre un développement séparé.

L'autre genre de réponses, qu'elles soient à visée négative ou positive, porte sur la question de la *simulation*. En simulant, la machine montre dans le jeu de l'imitation qu'elle est capable de simuler l'homme³¹. Nombre de critiques, avant et après Searle³², ont pensé pouvoir affirmer

³⁰ Turing (1950), p. 449-50.

³¹ D'ailleurs, Turing propose pour simplifier l'analyse que la meilleure stratégie de la machine, au cours du jeu, est d'*imiter* ce que ferait l'homme en pareille circonstance : les deux sens de simulation se trouvent donc confondus.

³² Searle (1980) présente une critique célèbre de ce qu'il appelle « intelligence artificielle forte », c'est-à-dire l'idée qu'un ordinateur convenablement programmé en sorte de produire les effets d'intelligence

que de la simulation à l'identité, la conséquence ne suit pas. Leurs adversaires ont répliqué que le dogme essentialiste est encore moins recevable. Sans rouvrir ce débat, remarquons qu'il remet en cause le principe même de l'expérience de pensée du Turing. De pouvoir simuler un ouragan sur ordinateur nous ne déduisons pas, comme le souligne Searle, l'hypothèse que l'ouragan est « essentiellement » un certain programme d'ordinateur : mais c'est que nous avons de bonnes raisons de penser qu'une part importante de ce qui fait d'un phénomène un ouragan tient précisément à la composition chimique, la structure physique des matières qui participent au phénomène. Il serait donc incohérent de filtrer ces propriétés-là en espérant conserver l'essentiel de l'ouragan. Toute l'idée de Turing est que nous ne savons pas définir ou même circonscrire l'intelligence³³ autrement que de manière ostensive, et que dans cette complète ignorance il est raisonnable de *proposer* de filtrer conceptuellement certaines propriétés matérielles et de voir si ce qui reste est suffisamment proche de la notion ostensive. Reprocher à Turing de commettre un sophisme et de *présupposer* que simulation vaut identité est tout simplement absurde.

La difficulté qu'il faut en revanche soulever ici ressemble superficiellement à la précédente. Il s'agit bien d'une question de filtrage, et ce qui est éliminé est bien de nature organique. Mais ce n'est pas le soubassement cérébral dont Turing, encore une fois, *propose* la mise entre parenthèses. Il s'agit d'une autre séparation, qu'il doit *présupposer* en effet pour « monter » son expérience³⁴.

voulus soit intelligent, ou encore puisse être réputé posséder un « esprit ». Voir aussi la discussion de Hofstadter & Dennett (1981).

³³ Il faudrait distinguer plus nettement esprit et intelligence. La discussion n'en serait pas beaucoup modifiée, en raison du rôle particulier assigné à ces concepts dans le contexte : que dans l'usage commun esprit et intelligence aient des sens dissemblables à de nombreux égards est une évidence, mais dont rien d'important ne découle ici.

³⁴ Je crois devoir à Jean Lassègue de m'avoir mis sur cette piste. Je lui dois en tout cas d'avoir pu discuter longuement de nos conceptions respectives, et différentes, de l'enjeu véritable.

Turing cherche à isoler, puis à reproduire, ce qui dans un comportement est d'ordre informationnel ou intentionnel, en éliminant tout le reste, c'est-à-dire la matérialité. Le *corps* humain doit être mis hors jeu : l'homme salue en soulevant sa casquette ou en émettant le son « Bonjour », la machine fait apparaître à l'écran la suite de caractères « B-o-n-j-o-u-r » : cette différence-là doit être annulée. Turing doit bel et bien présupposer que cela est possible, et non pas seulement dans un cas particulier, mais de manière systématique et principielle ; bref, il doit postuler l'existence d'un niveau informationnel (et conjointement d'un niveau « purement » matériel). Un comportement n'est pas un simple enchaînement de mouvements matériels : c'est un enchaînement de gestes significatifs agencés par une intelligence. Mais une telle description repose sur l'hypothèse de séparabilité.

Cette hypothèse n'est pas posée explicitement, et il est probable que Turing n'était pas clairement conscient de sa nécessité. En revanche, il fait valoir sa plausibilité indirectement de deux manières. D'une part, l'idée du télétype, présentée comme un moyen d'empêcher l'identification par les traits physiques (l'aspect corporel, la voix), a pour effet presque subreptice de rabattre tout comportement sur l'émission de messages imprimés. D'autre part, le caractère ouvert, non contraint, du jeu de l'imitation, dans lequel n'importe quelle question peut être posée, et n'importe quelle stratégie de dissimulation déployée, implique, sans l'affirmer, la portée générale de la réduction. Les deux procédés se complètent admirablement : réduction uniforme, puis universalité du domaine d'application.

Examinons-les rapidement, en commençant par le second. Dans un précédent article³⁵, Turing avait proposé un dispositif beaucoup plus simple, dans lequel une machine affrontait un joueur humain aux échecs. Le jeu de 1950 constitue un triple progrès : il écarte le risque de

³⁵ Turing (1947).

spécialisation — c'est bien l'intelligence « générale », comme diront plus tard Newell et Simon³⁶, que vise Turing (ce que confirme d'autre part l'accent mis sur l'existence d'une machine universelle); il empêche qu'on attribue le succès possible de la machine, et de l'expérience de pensée elle-même, à la nature intuitivement « formelle » du jeu d'échecs, que l'on serait tenté de voir comme un exemple taillé sur mesure pour Turing; il inclut expressément une aptitude qui semble au contraire la moins accessible à un traitement formel, à savoir la *métis*, la ruse. Pour autant, on reste très en deçà d'une universalité véritable: ce n'est pas tant qu'il manque les « savoirs tacites » ou « habiletés », les variétés de « *know how* », ou encore les comportements affectifs, sur lesquels les critiques de l'IA insisteront plus tard — Turing n'est pas sans arguments sur ce point (qui s'apparente aussi bien à l'argument du facteur manquant qu'à l'argument de l'informalité du comportement, mentionnés plus haut); c'est surtout que les comportements eux-mêmes ne sont pas directement mis à l'épreuve, mais seulement les témoignages des agents sur leurs propres comportements. Non seulement s'introduit une distorsion systématique, comme dans toute enquête par questionnaire; mais la nature même du comportement est voilée; affirmer que rien d'essentiel n'est occulté est faire appel, en dernière analyse, à l'hypothèse de séparabilité.

Examinons maintenant le recours au téléscrip-teur. Il soulève deux objections. La première s'inscrit dans le prolongement du questionnement précédent: le comportement auquel l'arbitre a accès, et sur lequel il fonde sa tentative d'identification, quel est-il? S'agit-il de la suite des actions sur le clavier du téléscrip-teur qu'il infère à partir des inscriptions qui apparaissent sur la bande de papier qu'il déchiffre? Ou bien de quelque chose de plus abstrait, comme l'intention scripturale qui guide ces gestes? Ou, plus loin (plus « haut »?), de l'intention communicative qui

³⁶ Newell & Simon (1975).

anime l'entité dissimulée dans sa cabine ? Bref, pour normalisé qu'il soit, le comportement « téléscriptionnel » n'en soulève pas moins tout le problème de la séparabilité. La deuxième objection concerne la nature particulière du comportement linguistique : comme le souligne D. Michie, la désinvolture de Turing sur ce chapitre semble aujourd'hui incroyable : rien n'indique qu'il entrevoie la difficulté pour la machine (ou son programmeur) de passer d'une « pensée » ou d'une « intention » communicative ou informative à une expression linguistique correcte et pragmatiquement adéquate. On pourrait le soupçonner de penser que la chose n'est pas sensiblement plus difficile, en général, que de taper sur le clavier « glf3 » lorsqu'on a déterminé dans une partie d'échecs le coup que l'on va jouer. C'est peut-être là que l'intuition de Turing, un demi-siècle plus tard, semble le plus manifestement prise en défaut.

Mais ne faut-il pas aller plus loin, et voir en cet « oubli du langage » l'indice d'un oubli plus général encore, qu'on pourrait appeler l'oubli de la représentation ? Tout se passe comme si pour Turing, la codification des entités dans un langage de calcul ne posait de problème ni technique ni conceptuel ; comme s'il s'agissait d'une immense gödelisation, d'un passage massif au code-barre des supermarchés. Ce n'est pas chez lui l'indice d'une foi formaliste de charbonnier : il ne défend pas une conception syntaxique de la saisie du sens, il ne se réclame pas d'une forme d'atomisme logique ou sémantique. Au contraire, il adopte une théorie des concepts, de leur nature et de leur acquisition, qui est antiformaliste dans la mesure où, comme le veut la psychologie de son époque, elle est associationniste et behavioriste. La représentation n'est pas absente, elle est transparente : que les croix sur le ruban de la machine de Turing puissent figurer n'importe quoi, et pas seulement des nombres, comme s'en sont avisé avec émerveillement, selon leur propre témoignage, les pionniers de l'intelligence artificielle dans les années qui suivent la publication de l'article, voilà ce que Turing ne pouvait ignorer. Il pensait

seulement sans doute que la décision d'interpréter les croix par des notes de musique ou des significations lexicales pouvait être laissée à un Dieu impartial, ou à un programmeur.

Le problème est bien trop important pour être traité convenablement en ces quelques lignes. Mentionnons seulement une tentative pour faire remonter cet « oubli de la représentation » à la première cybernétique, celle de McCulloch et Pitts³⁷. Rappelons aussi que la désinvolture de Turing sera celle des pionniers de l'intelligence artificielle : ils suivront Turing dans la lettre et dans l'esprit pendant une bonne décennie. Puis ils commenceront à découvrir les tourments de la « représentation des connaissances » et du « traitement du langage naturel ». La question ne sera cependant vraiment abordée que dans le cadre des sciences cognitives naissantes : le cognitivisme³⁸ est contemporain de l'émergence d'une problématique de la représentation.

*

Le Turing cher aux logiciens et le Turing précurseur de l'intelligence artificielle sont le même homme, le même penseur. Qu'il n'y ait pas continuité pure et simple entre les articles de 1936 et de 1950 est certain, mais aucune tension sensible ne se perçoit à l'examen attentif. L'antiformalisme que l'on pourrait être tenté d'attribuer au premier Turing n'a rien de radical ; nous avons vu qu'il se présente sous une forme ambivalente et ne joue qu'un rôle accessoire. Inversement, le formalisme du second Turing demeure

³⁷ Cf. l'essai « From paleo to neo-connectionism », ce volume ; il s'agit évidemment de McCulloch & Pitts (1943).

³⁸ Par quoi j'entends invariablement le programme computationnel dans les sciences cognitives, et non pas celles-ci dans leur entier. Cf. Andler (1992a) ou l'essai « Qu'est-ce que le cognitivisme ? », ce volume.

inachevé, et rien n'indique, dans les quelques années qui lui restaient à vivre, qu'il se souciait de lui donner la forme extrême qu'il prendrait entre les mains de certains de ses héritiers. On peut penser au contraire, vu l'orientation biologique de ses derniers travaux, qu'il aurait recherché une voie médiane dans les sciences cognitives d'aujourd'hui. Qu'il se soit néanmoins largement mépris sur les difficultés qui attendaient ceux qui entreprirent de réaliser le programme esquissé en 1950, sans diminuer en rien son immense mérite, ne doit pas être tu : il nous reste en effet à déterminer, un demi-siècle plus tard, si c'est à avoir été trop, ou trop peu, formaliste qu'il doit de s'être trompé.

BIBLIOGRAPHIE

- Anderson, J. & Rosenfeld, E., eds. (1988), *Neurocomputing. Foundations of Research*, Cambridge, Mass. : MIT Press.
- Andler, D. (1990), « Connexionnisme et cognition : à la recherche des bonnes questions », *Revue de Synthèse*, série générale CXI, n° 1-2, p. 95-127.
- Andler, D. (1992), « From paleo to neo-connectionism », in G. van der Vijver, ed., *New Perspectives on Cybernetics*, Kluwer, Dordrecht, p. 125-146.
- Andler (1992a), « Calcul et représentations : les sources », in Andler, D., dir., *Introduction aux sciences cognitives*, Paris : Folio Gallimard.
- Boolos, G. & Jeffrey, R. (1980), *Computability and Logic*, Cambridge : Cambridge University Press, 2d ed. (ou postér.)
- Church A. (1936), « An unsolvable problem of elementary number theory », *Amer. J. of Math.* **58**, p. 345-363. Repris in : Davis (1965), p. 88-107.
- Church A. (1936), « A note on the Entscheidungsproblem », *J. Symbolic Logic* **1**, p. 40-41. Réimp. in Davis (1965).
- Davis M., ed. (1965), *The Undecidable*, Hewlett, NJ : Raven Press.

- Davis M. (1982), « Why Gödel didn't have Church's Thesis », *Information and Control* **54**, p. 3-24.
- Dreyfus, H.L. (1972/79/93), *What Computers [Still] Can't Do. A Critique of Artificial Intelligence*, New York : Harper Torchbooks ; 3e éd., Cambridge, Mass. : MIT Press.
- Feferman, S. (1988), « Turing in the land of $O(z)$ », in Herken (1988).
- Gandy, R. (1988), « The confluence of ideas in 1936 », in Herken (1988).
- Haugeland, J., ed. (1981), *Mind Design*, Cambridge, Mass. : MIT Press.
- Herken, R., ed. (1988), *The Universal Turing Machine. A Half-Century Survey*, Oxford : Oxford University Press.
- Hodges, A. (1983), *Alan Turing : The Enigma of Intelligence*, Londres : Burnett ; trad. fr. abrégée : *Alan Turing ou l'énigme de l'intelligence*, Paris : Payot, 1988.
- Hofstadter, D. & Dennett, D., eds. (1981), *The Mind's I*, New York : Basic Books ; trad. fr. par Henry, J., *Vues de l'esprit*, Paris : InterEditions, 1987.
- Kleene S. C. (1936), « General recursive functions of natural numbers », *Math. Annalen* **112**, p. 340-353.
- Kleene S. C. (1981), « Origins of recursive function theory », *Annals of History of Computing* **3** (1), p. 52-67.
- Lassègue, J. (1994), « L'intelligence artificielle et le problème du continu », thèse de philosophie soutenue à l'Université Paris X — Nanterre.
- Lucas, J.R. (1961), « Minds, machines, and Gödel », *Philosophy* **36**, p. 112-127.
- McCulloch W. S. & Pitts W. H. (1943), « A logical calculus of ideas immanent in nervous activity », *Bulletin of Mathematical Biophysics* **5** ; trad. fr. in Pélissier & Tête (1995).
- Michie, D. (1993), « Turing's test », *Artificial Intelligence* **60**, n°1, p. 1-22.
- Newell, A. & Simon, H.A. (1976), « Computer science and empirical enquiry : symbols and search », *Comm. Ass.*

- Computing Machinery* **19**, p.113-126. Réimp. in Haugeland (1981).
- Newell, A. (1983), « Intellectual issues in the history of artificial intelligence », in Machlup, F. & Mansfield, U., eds., *The Study of Information : Interdisciplinary Messages*, New York : Wiley.
- Osherson, D., Stob, M., & Weinstein, S. (1986), *Systems That Learn*, Cambridge, Mass. : MIT Press.
- Pélissier, A. & Tête, A., dir. (1995), *Sciences cognitives. Textes fondateurs (1943-1950)*, Paris : PUF.
- Post, E. (1936), « Finitary combinatory processes. Formulation 1 », *J. Symbolic Logic* **1**, p. 103-105. Aussi in : Davis (1965), p. 289-291. Compte rendu par A. Church, *J. Symbolic Logic* **2** (1937), p. 43.
- Randell, B. (1982), *The Origins of Digital Computers*, Berlin : Springer, 3e éd.
- Searle, J. (1980), « Minds, brains and programs », *Behavioral and Brain Sciences* **3**, p. 417-458 ; trad. fr. (sans les objections complémentaires et la réponse de l'auteur) : « Esprits, cerveaux et programmes », in *Vues de l'esprit* (1983), p. 354-373.
- Shannon, C. E. (1938), « A symbolic analysis of relay and switching circuits », *AIEE Transactions* **57**, p. 713-723.
- Shannon, C. E., Weaver W. (1949), *The Mathematical Theory of Communication*, Urbana : University of Illinois Press.
- Sieg, W. (1991), « Mechanical procedures and mathematical experience », communication au colloque « Mathematics and Mind », Amherst, Mass., avril 1991.
- Smolensky, P. (1988), « On the proper treatment of connectionism », *Behavioral & Brain Sciences* **11**, p. 1-74.
- Turing, A. M. (1936-7), « On computable numbers, with an application to the Entscheidungsproblem », *Proc. London Math. Soc. ser. 2*, **42**, p. 230-265. « A correction », *ibid.*, **43**, 1937, p. 544-546. Trad. fr. par Basch J., in Girard J.-Y., *La machine de Turing*, Paris : Seuil, 1995, p. 49-102. Repris in : Davis (1965), p. 115-154. Compte rendu par A. Church, *J. Symbolic Logic* **2**, 1937, p. 42.

- Turing, A. M. (1947), « Intelligent Machinery », *Executive Committee NPL*, 1948, p. 1-20, H. M. S. O. Publications, London. Réimp. in B. Melzer & D. Michie, eds., *Machine Intelligence* 5, 1969, p. 3-23.
- Turing, A. M. (1950), « Computing machinery and intelligence », *Mind* LIX, 236, p. 433-460. Réimp. in Anderson A. R., ed., *Minds and machines* et dans de nombreuses autres anthologies, en particulier Hofstadter & Dennett (1981), mais tronqué ; trad. fr. par Blanchard P., in *Pensée et machine*, dir. Anderson A. R., Seyssel : Champ Vallon, 1983, p. 39-67 ; et in Pélissier & Tête (1995).
- Stern, J. (1991), *Fondements mathématiques de l'informatique*, Paris : McGraw Hill.
- Wagner, P. (1994), « Machine et pensée. L'importance philosophique de l'informatique et de l'intelligence artificielle », thèse de philosophie soutenue à l'Université Paris I.
- Webb, J.C. (1980), *Mechanism, Mentalism, and Mathematics*, Dordrecht : Reidel.