

SÉMINAIRE DE PHILOSOPHIE ET MATHÉMATIQUES

JEAN LASSÈGUE

L'actualité de Turing : l'inférence dans la pensée, le calcul et les formes biologiques

Séminaire de Philosophie et Mathématiques, 1996, fascicule 2
« L'actualité de Turing », , p. 1-23

http://www.numdam.org/item?id=SPHM_1996__2_A1_0

© École normale supérieure – IREM Paris Nord – École centrale des arts et manufactures, 1996, tous droits réservés.

L'accès aux archives de la série « Séminaire de philosophie et mathématiques » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

**Actualité de Turing :
l'inférence dans la pensée,
le calcul et les
formes biologiques**

E.N.S. 12-2-96

Introduction

Turing (1912-1954) commence à être mieux connu, grâce à la publication systématique de ses œuvres, dont 3 volumes sur 4 ont été publiés chez North-Holland.

En France, 2 de ses articles les plus fameux ont été traduits et viennent d'être réunis en un seul volume au Seuil.

Ce regain d'intérêt n'est pas motivé seulement par son œuvre scientifique. On note que depuis les années 80, on s'est intéressé au personnage Turing. On a écrit sur lui une pièce de théâtre, une biographie, un roman et un documentaire. C'est assez rare pour un scientifique et je ne connais pas d'autres figures scientifiques du 20ème siècle qui ait reçu un tel traitement !

L'intérêt littéraire porté à Turing vient de sa vie non-conventionnelle, à la fois du point de vue académique (il change plusieurs fois de direction de recherche et à chaque fois obtient des résultats importants) et du point de vue de sa vie personnelle (il était homosexuel, a été condamné pour outrage aux mœurs et, suite au traitement hormonal qu'il a du subir comme peine, il s'est suicidé). Il y a bien de quoi favoriser l'imagination littéraire dans cette vie !

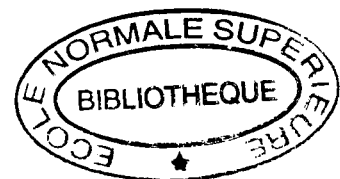
En quel sens peut-on parler d'une actualité de Turing ?

Turing a complètement transformé la façon dont on interprétait la notion d'inférence logique en se plaçant d'un point de vue inédit, que l'on appelle depuis lors, "cognitif". Pour ce faire, il a renouvelé 3 points de vue particuliers :

- Le point de vue psychologique
- Le point de vue calculatoire
- Le point de vue biologique

et il a montré leurs liens réciproques.

Il a effectué ces changements en s'appuyant sur deux principes épistémologiques qui dérivent de Leibniz : le principe de continuité qui postule une gradation indéfiniment complexe des êtres dans la nature et le principe des indiscernables qui postule l'existence de différences insensibles entre chaque être en fondant leur individualité.



Je vais passer en revue les changements qu'il a effectués dans ces trois domaines du point de vue de l'inférence logique. C'est ce qui me permettra de préciser ce que j'entends par "actualité" de Turing.

Du point de vue de l'inférence logique, on oppose habituellement le processus de vérification de l'inférence et le processus de découverte de l'inférence. Turing a obtenu des résultats dans les deux directions:

- Du point de vue de la vérification, il a montré ce que l'on pouvait attendre de sa mécanisation.
- Du point de vue de la découverte, il a montré quel itinéraire psychologique avait été le sien dans ses découvertes logiques.

Au vu de ces résultats, la grande actualité de Turing me paraît venir de ce qu'il a déplacé la frontière entre sciences dures et sciences humaines. C'est ce déplacement qui est à l'origine de son "modèle computationnel de l'esprit" et que son biographe, le mathématicien Andrew Hodges décrit en ces termes :

«Le modèle de la machine à états discrets est la description adéquate d'un des aspects du monde matériel - à savoir le fonctionnement du cerveau -.»

Mon argumentation se fera donc en 3 points :

- L'inférence dans la pensée
- L'inférence dans le calcul
- L'inférence dans les formes biologiques

1. Le point de vue psychologique : l'inférence dans la pensée

Du point de vue logique, la notion d'inférence a été longuement retravaillée depuis les débats qui ont agité la communauté des mathématiciens depuis Cantor. On a mis au jour les différentes manières de passer en revue les éléments des collections infinies et les problèmes comme l'induction mathématique, la définition imprédicative, la mise à la puissance d'un ensemble, le rejet intuitionniste du tiers-exclu ont été analysés et débattus.

L'une des questions posée au cours de ces débats est celle de savoir ce qui permet à la démonstration logique de revêtir un aspect fini puisque le lien entre prémisses et conséquences n'est pas composé d'une infinité d'étapes intermédiaires, ce qui serait nécessaire s'il fallait s'assurer de la validité de la démonstration pour une infinité d'objets. Sans entrer dans le débat proprement logique qui permet de faire avancer cette question, je vais m'intéresser à ce que cela suppose d'un point de vue psychologique.

1.1. Le jeu conscient / inconscient

De ce point de vue, ce qui se passe dans l'esprit est peu clair. On peut cependant accorder un point : puisqu'on ne passe pas en revue tous les objets d'une collection infinie et que cependant on arrive à tenir un raisonnement sur eux, c'est que, grâce à l'aspect logique du raisonnement, on est en mesure de faire comme si on avait parcouru les étapes en question.

D'un point de vue psychologique, on en conclut que certaines étapes du raisonnement logique peuvent être effectués inconsciemment.

J'entends "inconscient" au sens de Leibniz quand il parle des petites perceptions infinies qui produisent une perception unique. Son exemple favori est le bruit d'une vague qui s'étale sur le rivage et qui est composé d'une

multitude de petits bruits que nous percevons comme un tout¹.

Comment faut-il comprendre, chez Turing, ce rapport du conscient et de l'inconscient ?

Pour Turing, toute étape de l'inférence doit pouvoir être exhibée en propre sans passer par un raisonnement logique indirect et on comprend pourquoi quand on se rappelle ses principes épistémologiques : la continuité dans les étapes du raisonnement logique soit toujours être accessible pour sauvegarder le lien entre prémisses et conséquences mais cette continuité exige que toutes les étapes soient discernables, c'est-à-dire non confondues.

On peut donc dire qu'il y a une communauté de nature entre conscient et inconscient, communauté que l'on peut décrire en quatre points:

1. Le conscient n'est pas une propriété attachée à une étape particulière du raisonnement et en particulier pas aux prémisses et à la conclusion.
2. Toute étape d'un raisonnement doit pouvoir devenir consciente, même au prix «d'un gros effort» comme dit Turing².
3. Il n'y a rien de ce qui s'effectue consciemment qui soit qualitativement différent de ce qui s'effectue inconsciemment.
4. L'inconscient n'est qu'un effet de vitesse de fonctionnement : il suffit de ralentir le processus pour être capable d'en reconnaître le contenu³.

12. Rôle de la mémoire

Cette communauté de nature entre conscient et inconscient n'empêche pas Turing d'introduire une 2ème hypothèse psychologique qui prend la forme d'une restriction par rapport à la première :

Dans les faits, on ne peut pas s'arrêter sur chaque étape pour des raisons qui tiennent à la nature physique de notre mémoire : notre mémoire est en effet physiquement limitée.

Le stockage inconscient mais conscientisable des étapes de l'inférence - sous peine de rupture de la chaîne inférentielle - est une nécessité qui découle de cette limitation.

Cependant, on peut se demander en quoi le stockage pourrait faciliter le compactage des étapes intermédiaires de l'inférence dans le cas où il y en a une infinité. Dans ce cas-là, on ne voit pas comment le stockage, fût-il inconscient, pourrait faciliter la manipulation des étapes en question puisque la taille de la mémoire est finie. Il faut donc faire intervenir un autre facteur pour comprendre comment l'inférence est possible.

13. Rôle de l'écriture

Ce facteur c'est l'écriture des symboles et des règles qui joue un rôle fondamental dans le stockage en mémoire : le lien me paraît très fort ici entre l'écriture symbolique et la mémoire, c'est-à-dire le jeu entre conscient et inconscient.

Quel est en effet le rôle de l'écriture ? Son rôle est de compacter les étapes de l'inférence par le biais de

¹ Cf. par exemple l'Introduction des *Nouveaux Essais*.

² A. M. Turing, (1950), "Computing Machinery and Intelligence", *Mind*, vol LIX, n°236, 1950 : 459.

³ A. M. Turing, (1950), "Computing Machinery and Intelligence", *Mind*, vol LIX, n°236, 1950 : 459.

symboles et de règles. Là encore se trouve très proche de Leibniz qui accorde le même rôle à l'écriture.

L'inférence, d'un point de vue psychologique, est donc rendue possible par un effet d'écriture qui permet le stockage sous forme de symboles.

Les trois caractéristiques que je viens de mentionner doivent donc être mises en relation pour rendre compte de l'inférence d'un point de vue psychologique : l'aspect limité de la mémoire physique rend nécessaire un jeu entre le conscient et l'inconscient soutenu par une écriture symbolique.

Remarquons que la description psychologique de l'acte d'inférence fait appel à plusieurs niveaux d'intelligibilité :

1. Un niveau psychologique : c'est le rapport conscient/inconscient

2. Un niveau physique :: c'est l'aspect limité de la mémoire

3. Un niveau "intermédiaire" : l'écriture comme système de marques physiques joue un rôle à la fois du point de vue du physique puisqu'il s'agit bien de traces matérielles et du point de vue psychologique puisqu'elle a partie liée avec le rapport conscient / inconscient.

C'est la mise en rapport de ces trois niveaux qui caractérisent, pour Turing, l'inférence d'un point de vue psychologique. Cette analyse psychologique va avoir des conséquences importantes quand Turing voudra caractériser l'effectuation d'une inférence comme relevant du calcul, car il fait remarquer que, pour lui, «mécanisme et écriture sont presque synonymes»⁴.

J'en viens donc maintenant à la notion de calcul dans l'inférence.

2. Le calcul : le mécanisme de l'inférence

Quand on dit "effectuation des étapes de l'inférence", on dit en fait "calcul". pour bien le comprendre, il faut revenir à la façon dont le problème s'est posé à Turing dans les années 30 quand il suivait le cours de logique mathématique de Max Newman. à Cambridge.

21. Etat de la question

211. Hilbert

La recherche en logique mathématique était dominée par le programme de Hilbert, grand programme fondationnel, dont deux idées vont jouer un rôle déterminant pour Turing. La première idée est proprement logique et la seconde psychologique :

1°. On se rappelle que Hilbert voyait dans la notion de système formel le moyen de vérifier la validité des démonstrations; cette vérification des chaînes d'inférence est rendue possible par le jeu de règles logiques capables d'engendrer les répliques formelles des théorèmes et de leurs démonstrations. Dans un système formel, les énoncés mathématiques sont alors réduits à un stock de formules manipulées logiquement.

2°. Pour Hilbert, cette manipulation s'opère dans un cadre finitiste, le seul dont nous soyons épistémologiquement sûrs. L'hypothèse psychologique avancée par Hilbert est donc que la «pensée est finitiste;

⁴ A. M. Turing, (1950), "Computing Machinery and Intelligence", *Mind*, vol LIX, n°236, 1950 : 461

quand nous pensons se déroule un processus finitiste»⁵.

C'est donc dans ce cadre que l'on doit trouver tous les moyens de vérifier la validité des démonstrations. C'est dans ce cadre aussi que se pose un certain nombre de problèmes spécifiques touchant la nature des systèmes formels. L'un de ces problèmes est celui de la décision ou *Entscheidungsproblem* : peut-on décider pour tout énoncé formalisé s'il est ou non dérivable à partir des axiomes du système formel auquel il appartient ? Puisque les propositions mathématiques peuvent être représentées formellement, la question de la *dérivabilité* au sein d'un système formel, c'est-à-dire la question de l'inférence logique, vaut comme question de *décidabilité* pour tout énoncé mathématique, c'est-à-dire de savoir si on peut toujours décider pour un énoncé mathématique quelconque de sa valeur de vérité.

212. Gödel

Turing hérite de ce cadre de pensée sans être un hilbertien orthodoxe puisque sa réflexion sur la notion d'effectivité du calcul se situe après les résultats d'incomplétude de Gödel. Le point de vue de Turing est bien plutôt post-gödelien et Turing hérite en effet de la problématique de Gödel sur un point capital.

Gödel a en effet montré qu'il y avait un rapport étroit entre le concept logique de dérivabilité dans un système formel et celui de calculabilité.

L'instrument privilégié de cette mise en rapport est la numérotation de Gödel qui permet de désigner une proposition mathématique dans un système formel par un nombre spécifique qui lui sert de contremarque. L'idée générale de la mise en rapport de la dérivabilité et de la calculabilité est rendue possible par le fait que les deux processus ont en commun de devoir être effectif, c'est-à-dire d'être des processus que l'on doit suivre pas à pas.

Cette idée d'une mise en rapport par le biais d'une numérotation, c'est-à-dire par un jeu d'écriture, de la notion d'inférence dans un système formel; et de la notion de calcul va jouer un rôle décisif pour Turing puisque le cours de logique mathématique de Newman s'achevait précisément sur la question de savoir comment on pouvait généraliser pleinement les résultats de Gödel en précisant ce que l'on doit entendre par processus effectif, c'est-à-dire par calcul. En effet, la notion d'effectivité chez Gödel définissait une certaine classe de fonctions calculables dont on pouvait se demander si elle ne pouvait pas être étendue. Mais on n'avait, à l'époque, aucune idée de la façon dont on pouvait confirmer ou infirmer cette intuition.

C'est précisément ce à quoi Turing a contribué : montrer que l'essence du système formel réside dans son aspect mécanique, c'est-à-dire que le rapprochement opéré entre dérivabilité et calculabilité pouvait être pleinement justifié si l'on parvenait à montrer que ce qui caractérisait la notion de système formel, c'était précisément la notion de calcul mécanique qui permettait de construire la même classe de fonctions que les fonctions calculables dont Gödel avait parlé.

22. La généralité de la notion de calcul

Il restait donc une question à résoudre : comment caractériser dans toute sa généralité ce que l'on entend par calcul ?

⁵ D. Hilbert, "Les fondements des mathématiques", traduction française dans [J. Largeault éd., *Intuitionisme et théorie de la démonstration*, Paris, Vrin, 1992] : 140.

Cette question n'a d'intérêt que dans le contexte où elle s'inscrit puisque les mathématiciens et les logiciens s'étaient depuis toujours entendus sur ce que l'on devait entendre par calcul. Ce contexte est double :

1°. On cherche depuis Gödel à caractériser adéquatement comme calcul le lien d'inférence.

2°. On cherche à savoir si oui ou non Hilbert a raison de défendre l'hypothèse cognitive selon laquelle la pensée est un processus finitiste, c'est-à-dire finalement si la pensée calcule.

Turing répond à ces deux questions positivement :

- *L'inférence est un calcul*
- *La pensée calcule.*

On voit que les questions d'ordre logique et psychologique sont intimement mêlées dans cette réflexion sur la nature de la notion de système formel quand Turing cherche à caractériser l'essence du calcul comme effectuation mécanique. Comment Turing parvient-il à cette caractérisation ?

23. La caractérisation de la notion de calcul

Turing, dans son fameux article de 1936, "On Computable Numbers with an Application to the Entscheidungsproblem" établit un parallèle entre le fonctionnement de l'esprit et le fonctionnement d'une machine.

Dans son article de 1939, "Systems of Logic based on Ordinals", Turing décrit de la manière suivante la comparaison qu'il établit entre esprit et machine ⁶:

«Nous pouvons prendre cette proposition d'un point de vue littéral en comprenant par un "processus purement mécanique" celui qui pourrait être effectué par une machine.»

Contrairement à ses prédécesseurs qui utilisaient le terme de mécanique en soulignant l'aspect arbitraire du rapprochement effectué dans la métaphore, Turing prend la métaphore au sérieux : l'esprit dans l'acte de calcul ne ressemble pas plus ou moins à une machine, *c'est* une machine. Aussi le terme de "calculateur", [computer] s'adresse-t-il finalement aussi bien à l'homme qu'à la machine. A partir de cette identification, on peut constituer un corpus de remarques communes qui valent aussi bien pour l'esprit que pour la machine. C'est ce que fait Turing dans son article de 1936 dont je cite un large extrait :

«Le calcul est normalement opéré en écrivant certains symboles sur du papier. Nous pouvons supposer que ce papier est divisé en cases comme dans un cahier d'arithmétique pour enfant. En arithmétique élémentaire, le caractère bi-dimensionnel du papier est quelque fois utilisé. Mais un tel usage peut être évité [...]. Je fais donc l'hypothèse que le calcul peut être mené à bien sur un papier uni-dimensionnel, c'est-à-dire sur un papier divisé en cases. [...] Le comportement du calculateur à un moment quelconque est déterminé par les symboles qu'il est en train d'observer et son "état d'esprit" à ce moment. [C'est ce qui définit ce que Turing appelle une "m-configuration"]. Nous pouvons supposer qu'il y a une limite B au nombre des symboles ou cases qu'un calculateur peut observer à chaque moment. S'il souhaite en observer plus, il doit exécuter des observations successives. Nous ferons aussi la supposition que le nombre d'états d'esprit qui doivent être pris en considération est fini. [...]

⁶ A. M. Turing, (1939), 'Systems of logic based on Ordinals', Proceedings of the London Mathematical Society, ser. 2, vol. 45 : 161-228; republié dans [Davis M., *The Undecidable* Raven Press, Hewlett, New-York,] : 166.

Nous pouvons maintenant construire une machine qui fasse le travail du calculateur. A chaque état d'esprit du calculateur correspond une "*m*-configuration" de la machine. La machine vise les cases B correspondant aux case B observées par le calculateur. A tout mouvement, la machine peut changer un symbole sur une case visée ou peut changer chacune des cases visées en une autre case distante de moins de L cases de l'une des autres cases visées. Le mouvement qui est exécuté et la configuration qui s'en suit, sont déterminées par le symbole visé et la *m*-configuration. [...]; une machine à calculer peut être construite en vue de calculer [...] la suite calculée par le calculateur.»

24. Les principes de l'acte de calcul

On peut distinguer, dans cette longue citation, l'observation de la case du ruban et l'opération sur la case du ruban⁷.

241. Observation :

1. Il y a un nombre fini déterminé de symboles qu'un calculateur peut reconnaître immédiatement.
2. Il y a un nombre fini déterminé d'états internes que l'on a besoin de considérer.

Donc, il y a une finitude des *m*-configurations (c'est-à-dire du couple formé par l'état interne à un moment *t* et par la case observée).

242. Opération :

1. Seuls les symboles observés peuvent être modifiés.
2. Toute nouvelle case observée doit se situer à une distance finie de la case qui vient d'être observée.

Donc, il y a finitude de l'opération.

243. Choix du calcul d'un nombre réel

Muni de ces principes, Turing décrit la notion générale de nombre calculable en prenant en considération le calcul des nombres réels. Les machines qu'il décrit produisent donc en sortie la suite infinie des décimales du nombre réel à calculer : on voit donc bien, dans ce cas, que le calcul est à chaque étape fini, même s'il demanderait l'infinité du temps pour être exécuté entièrement.

Par quoi le choix du calcul d'un nombre réel est-il motivé, alors que Turing aurait pu choisir des calculs de valeurs individuelles de fonctions d'entiers ? A mon sens, le choix des réels s'explique pour des raisons qui touchent au cadre épistémologique choisi par Turing - et en particulier, le principe de continuité -, plus que par des raisons de commodité technique. Le principe de continuité est général et à ce titre, on doit le retrouver autant dans le domaine idéal des nombres que dans la nature et c'est d'ailleurs pourquoi, comme nous le verrons, les étapes de l'inférence doivent pouvoir s'incarner dans la nature. La gradation insensible qui approxime un nombre réel en convergeant vers lui fait donc partie de la même gradation que celle que l'on trouve dans la nature car, dit Turing, «tout est continu

⁷ Cf. Mundici et Sieg (1995), 'Paper Machines', *Philosophia Mathematica*, series III, vol. 3, 1995 : 12-13.

dans la nature»⁸.

De ce point de vue, ce n'est pas le continu mathématique qui intéresse Turing (d'ailleurs il n'a pas travaillé sur le problème ensembliste de l'hypothèse du continu, qu'il mentionne une seule fois dans son article de 1939, 'Systems of Logic based on Ordinals') ni non plus la controverse entre les tenants du continu ensembliste et ceux du continu réduit "à la Brouwer". C'est plutôt le rapport à la physique qui le motive et l'idée - que je dirais leibnizienne - qu'il y a une gradation insensible dans les êtres de la nature et que cette gradation est maîtrisable par le calcul. Le cas particulier du calcul d'un réel dont on peut engendrer l'expansion décimale entre donc dans un cadre beaucoup plus large et qui relève plus de l'épistémologie que de la mathématique.

C'est en tout cas dans ce cadre que Turing résout négativement le problème de la décision, l'*Entscheidungsproblem*, en rabattant ce problème sur un problème équivalent mais qui entre directement dans son cadre théorique, que Turing baptise le problème de l'arrêt. Ce problème s'énonce de la manière suivante : peut-on savoir mécaniquement si tout calcul de l'expansion décimale d'un nombre réel exécuté par une machine aura ou non une fin ? La réponse est *non* : il n'y a pas de machine capable de décider globalement du comportement d'une machine donnée au vu de sa seule table d'instructions. Turing montre dans la suite de son article que ce problème mécanique de décision dans son cadre théorique est équivalent au problème logique de la décision auquel il faut donc répondre aussi par la négative.

25. Les hypothèses nominalistes de Turing

Je voudrais aborder maintenant un point qui vise à compléter ce cadre épistémologique. Il s'agit des hypothèses nominalistes qui gouvernent l'analyse de l'acte de calcul tel qu'il est mené par Turing. Je voudrais pour ce faire partir d'une remarque de Judson Webb touchant l'argumentation générale de Turing⁹:

«Le cœur de son argument était une nouvelle analyse abstraite de ce que l'on entend logiquement par "se souvenir effectivement" des choses ayant rapport au calcul, tels que les symboles ou combien de fois on doit exécuter une sous-routine : pour ce faire, on doit être capable de passer d'un état discernable à un autre, *que vous soyez un humain ou une machine*. Nous faisons évidemment l'hypothèse que les "états d'esprit" peuvent aussi avoir des souvenirs qui dépassent les rêves les plus fous des machines, mais les seuls qui soient pertinents pour le calcul effectif sont ceux dans lesquels vous êtes placés par les symboles et les processus qui apparaissent au cours du calcul».

Si l'on suit l'argument de Webb, ce sont les symboles qui nous placent dans des états d'esprit particuliers et non les actes de l'esprit qui se matérialisent sous forme de symboles. Le fond de l'analyse de l'acte de calcul tel qu'il est présenté par Turing est donc qu'il y a une *efficacité intrinsèque aux symboles* : c'est ce que j'appelle une hypothèse nominaliste.

Comme le remarque Webb, cette hypothèse a une conséquence importante sur le rapport entre la notion

⁸ A. M. Turing, (1950), 'Computing Machinery and Intelligence', *Mind*, vol LIX, n°236, 1950 : 433-460.

⁹ J. C. Webb, "Introductory note to 1972a" in [K. Gödel, *Collected Works*, vol. II, Oxford University Press, Oxford, 1990], p. 302.

d'états d'esprit et celle de symbole. On peut imaginer en effet des états d'esprit extrêmement complexes mais comme Turing le fait remarquer :

«l'utilisation d'états d'esprit plus compliqués peut être remplacé par l'écriture de plus de symboles [...]».

L'exemple le plus probant de ce point de vue est celui des machines universelles. En effet, n'importe quelle machine, même celles dont les états internes sont très complexes, peut être imitée par une machine universelle, parce que le problème de la complexité des états peut toujours être contourné par le biais d'un surcroît d'écriture. La machine universelle apparaît donc comme le paradigme de cette efficacité symbolique qui compense toujours la complexité croissante des machines à simuler malgré son nombre fini d'états¹⁰.

Dès lors, le raisonnement de Turing est qu'il n'y a pas besoin de supposer une efficacité de l'objet sur l'esprit parce que l'écriture en dispense. Turing le dit d'ailleurs lui-même ¹¹:

«Il est toujours possible pour le calculateur de faire une pause dans son travail, de partir et de tout oublier à son sujet, puis de revenir plus tard et de s'y remettre».

L'être humain peut donc oublier ce qu'il faisait et changer entièrement les états de son esprit, il n'en reste pas moins capable, grâce aux instructions qu'il a écrites et conservées hors de lui, de reprendre ensuite son calcul pour retrouver immédiatement les états d'esprit finis qui étaient les siens.

On voit donc que la métaphore qui unit la pensée et la machine est rendue possible parce que ce qui gouverne l'analyse de l'acte de calcul, ce sont les effets d'écriture qui interviennent de façon semblable chez l'homme et la machine. Il y a un point de vue différent qui peut servir d'alternative : comme le faisait remarquer Gödel, au moins à un certain moment de sa réflexion puisque les textes conservés dans son *Nachlass* semblent faire état d'une position différente ultérieurement, on peut penser que le sens de l'influence est tout autre : c'est l'objet étudié qui a une efficacité sur l'esprit humain et non la seule écriture. Je ne veux pas clore le débat mais seulement montrer qu'il existe et que l'analyse de l'acte de calcul de Turing n'est sans doute pas le dernier mot touchant la nature de l'acte de calcul.

Je voudrais maintenant passer au dernier moment de mon étude et qui consiste à essayer de montrer en quoi l'inférence doit être conçue non seulement comme un processus psychologique et mécanique mais aussi comme un processus *naturel*.

3. Formes : l'inférence dans la nature

La taille limitée de la mémoire est un trait que l'espèce humaine partage avec les autres organismes de la nature. Il y a donc ici une contrainte naturelle qui joue dans deux registres : d'une part, nous l'avons vu, la limitation physique de notre mémoire structure la forme de notre intelligence qui doit s'aider de symboles; d'autre part on constate que cette limitation a partie liée avec la structure de l'organisme. Ainsi y-a-t-il un rapport, qu'il faut

¹⁰ Turing dit dans sa conférence de 1947 : «Ainsi la complexité de la machine à imiter est concentrée dans le ruban et n'apparaît en aucune façon dans la machine universelle elle-même». A. M. Turing, 'Lecture to the London Mathematical Society', 20 february 1947; republié dans [*Mechanical Intelligence, Collected Works*, D. C. Ince ed., North-Holland, Amsterdam, 1992 : 87-105].

¹¹ A. M. Turing, 'On Computable Numbers with an Application to the Entscheidungsproblem', *Proceedings of the London Mathematical Society*, 42 : 230-265; republié dans [M. Davis ed., *The Undecidable*, op. cit.], p. 139.

préciser, entre l'intelligence et la structure de l'organisme.

La thèse cognitive défendue par Turing n'est donc pas seulement liée à un certain contexte logico-mathématique : elle s'inscrit également dans un contexte plus large qui inclut des questions d'ordre physique et biologique. Je voudrais m'arrêter un instant sur les éléments les plus saillants de ce contexte. Le premier trait général que j'évoquerai est l'idée d'une continuité des organismes dans la nature.

31. Continuité des organismes dans la nature

Cette idée, dans sa forme moderne, remonte à Darwin bien qu'on la trouve déjà chez Aristote. On trouve en effet dans *L'origine des espèces*, publiée en 1859, l'idée selon laquelle seule la théorie de la sélection naturelle permet de rendre raison de l'adage aristotélicien : "La Nature ne fait pas de sauts". Voici le passage de Darwin¹² :

«Grâce à la théorie de la sélection naturelle, nous pouvons clairement comprendre la signification profonde du vieux canon de l'histoire naturelle "Natura non facit saltum". Ce canon, si nous n'observons que les habitants actuels du monde n'est pas à proprement parler exact mais si nous incluons tous ceux des temps passés, il doit être considéré, grâce à ma théorie, comme parfaitement exact.»

Cette thèse darwinienne a une contrepartie proprement cognitive. A la fin de *L'origine des espèces*, Darwin déclare¹³ :

«La psychologie sera basée sur un nouveau fondement, à savoir l'acquisition de chaque capacité mentale par gradation».

La sélection naturelle de variations apparaissant aux hasard permet donc de justifier l'idée d'une continuité dans la Nature qui s'exprime aussi bien dans la gradation entre *organismes* que dans la gradation entre capacités *intellectuelles*.

Mais la relation entre les deux types de gradation ne fait pas l'objet d'une analyse poussée de la part de Darwin et c'est l'école behavioriste en psychologie qui, confrontée à cette question, a tenté d'élaborer une réponse dont Turing hérite directement¹⁴.

32. La notion behavioriste d'apprentissage par stimulus-réponse

La théorie behavioriste vise à justifier l'idée selon laquelle il y a une continuité entre le signal de type *causal* que l'on rencontre chez les animaux et le symbole à contenu *sémantique* que l'on rencontre chez les humains. La pièce maîtresse de l'argumentation consiste à dire que *toute* communication, animale ou humaine, se fait par le biais d'un apprentissage qui repose sur la structure du stimulus-réponse. Cette structure permet de rendre compte de toute forme d'intelligence, de la plus simple à la plus complexe, parce que la structure stimulus-réponse ne requiert pas, dans son fonctionnement, une quelconque capacité intellectuelle supérieure. On voit bien ici comment la thèse

¹² Charles Darwin, (1859), *The Origin of Species*, The Penguin English Library, Harmondsworth, 1968, p. 233

¹³ Charles Darwin, (1859), *The Origin of Species*, The Penguin English Library, Harmondsworth, 1968, p. 458

¹⁴ Historiquement, cette réponse date de 1840. Elle a été élaborée ensuite durant tout le dix-neuvième siècle avant d'être reprise par l'école behavioriste. Mais le relais, pour Turing, est bien l'école behavioriste. Cf. sur l'histoire de la notion de stimulus-réponse en psychologie et en psychiatrie, M. Gauchet, *L'inconscient cérébral*, Seuil, Paris, 1992, particulièrement le chapitre 3.

cognitive de Turing hérite de ce point de vue : on se rappelle que, chez lui, chaque étape de l'inférence ne requiert en effet aucune capacité intellectuelle supérieure pour pouvoir être effectuée et qu'en tant que calcul, elle peut être effectuée par n'importe quel organisme.

Dans la lignée de Darwin et des behavioristes, mais en se démarquant de leur évolutionnisme matérialiste quelque peu simpliste, Turing pose la question suivante : *Quelles contraintes émanant de la structure physico-chimique de l'organisme pèsent sur l'intelligence ?*

Sa réponse ne me paraît pas définitivement arrêtée et il semble avoir hésité entre deux réponses, l'une clairement énoncée, l'autre à peine thématisée. Étudions ces deux réponses en commençant par la plus claire.

33. La première réponse : les formes organisatrices dans l'organisme et dans l'intelligence

33. 1. Le rapport à Darwin

En posant la question des contraintes physico-chimiques qui pèsent sur la gradation des formes des organismes, Turing se heurte à une grande question laissée quelque peu en suspens par la théorie darwinienne (et qui alimente encore aujourd'hui le débat entre néo-darwinien et partisans de la genèse causale des formes). Il y a en effet des limites à l'intérieur desquelles des variations peuvent apparaître. C'est à la détermination de ces limites qu'il s'est attaché.

Par exemple, la théorie de la sélection naturelle a du mal à expliquer pourquoi deux papillons de famille différente ont développé exactement le même motif coloré sur leurs ailes. On peut fournir une explication *ad hoc* en disant qu'il s'agit d'un effet mimétique et que ces papillons n'étant pas comestibles, adopter le même motif permet d'éviter plus facilement que les oiseaux n'en fassent leurs proies : en réduisant le temps d'apprentissage des oiseaux, l'apparition d'un même motif réduit du même coup le nombre de papillons tués. Mais on peut aussi remarquer qu'une grande proportion des motifs d'ailes de papillon font partie d'un même ensemble, appelé "plan nymphalide" et que les motifs en question sont des variations comprises dans ce plan. Ces variations sont obtenues en changeant la constante de vitesse de certaines réactions chimiques¹⁵. Il est donc plus facile de rendre compte de la ressemblance de deux familles de papillons par leur ressemblance dans leur structure physico-chimique que par un raisonnement finaliste mettant en action un concept de mimétisme difficile à utiliser parce que susceptible d'être taxé d'anthropomorphe.

On peut donc expliquer la variation des formes du vivant par des déterminations physico-chimiques tout en restant dans un cadre darwinien : il ne s'agit pas de supprimer les effets des variations aléatoires mais de replacer ces effets dans un éventail de possibilités déterminé par des conditions physico-chimiques déterminables *a priori*. L'idée darwinienne d'une continuité des formes dans l'évolution est donc à la fois préservée et modifiée puisqu'on tente de remplacer l'utilisation d'arguments finalistes *ad hoc* portant sur les cycles écologiques dans la nature par des contraintes naturelles. Turing, en lecteur assidu du *On Growth and Form* de D'arcy Thompson, a ainsi construit des modèles informatiques permettant de rendre raison des formes du vivant tout en acceptant que leur apparition se fasse de façon aléatoire.

¹⁵ P. T. Saunders, (1993), "L'évolution des formes biologiques", *Intellectica*, n° 16, 1993/1 : 70.

Il y a donc bien, pour Turing, un moyen de rendre compte *causalement* de l'idée darwinienne de gradation dans la nature.

33. 2. Le rapport au béhaviorisme

Le point de vue béhavioriste consistait à rendre raison de l'intelligence en la replaçant dans la continuité de la nature grâce à la notion de stimulus-réponse. Turing considère que c'est par le biais de l'apprentissage par stimulus-réponse que l'on acquière la forme universelle de l'intelligence, à savoir l'inférence computationnelle. Il y a donc une forme universelle d'intelligence que l'on doit retrouver plus ou moins développée au sein de la nature, quel que soit le type d'organisme considéré. De ce point de vue, la notion d'intelligence n'est pas l'apanage de l'humanité mais est bien universellement distribuée dans la nature.

Là où le point de vue de Turing se distingue de la théorie béhavioriste, c'est que cette théorie suppose un lien causal *direct* entre la forme de l'intelligence et la structure de l'organisme. Pour Turing, il n'y a qu'un rapport négatif entre monde physique et monde de l'esprit : c'est la limitation de la mémoire physique qui a des conséquences sur la forme de l'intelligence. Il n'y a pas de positivité de ce rapport au sens où on pourrait dire que tel caractéristique physique a des conséquences sur la forme de l'intelligence. C'est ce point que nous allons étudier en détail maintenant.

33. 3. Le rapport entre la gradation entre les formes d'organisme et entre les formes d'intelligence

Au lieu de vouloir préciser le rapport de causalité que le béhaviorisme supposait entre une certaine structure de l'organisme et une certaine forme d'intelligence, le point de vue de Turing vise au contraire à montrer que la théorie béhavioriste doit présupposer l'existence de formes structurant l'organisme d'une part et l'intelligence de l'autre sans qu'il y ait un rapport de causalité direct entre les deux. Alors que le point de vue béhavioriste cherchait à réduire le mental au physique par la découverte de la causalité *physique* qui mène de l'un à l'autre, Turing veut montrer au contraire que l'idée d'un continuum des organismes et de l'intelligence dans la nature implique deux conditions :

1°. elle *présuppose* l'existence d'un concept abstrait de forme et d'un concept abstrait d'intelligence : de même qu'une forme identique se retrouve dans plusieurs organismes (par exemple, un même motif sur la peau d'animaux très différents), de même il existe une forme générale de l'intelligence développée de façon plus ou moins complexe partout dans les organismes naturels. Bref, s'il y a bien un continuum dans la nature, on doit pouvoir extraire de l'organisme le concept de *structure* et abstraire de l'intelligence celui de *forme*, les deux concepts n'étant pas liés aux substrats particuliers qui leur sont associés.

2°. elle rend les deux concepts *autonomes* l'un par rapport à l'autre, c'est-à-dire par rapport à la détermination d'une causalité qui les réduiraient au physique. Dès lors, *on peut étudier de façon autonome et les structures des organismes et la forme de l'intelligence*.

Turing subvertit donc le point de vue béhavioriste en tentant de montrer que le "continuisme" darwinien n'implique pas un réductionnisme vulgaire qui voit un lien causal direct entre une certaine structure physique et une certaine forme intellectuelle. La réponse à la question : *Quelles contraintes émanent de la structure physico-chimique*

de l'organisme pèsent sur l'intelligence ? est donc qu'il n'y a pas de contraintes physico-chimiques puisque les notions de structure de l'organisme et de forme de l'intelligence sont en fait des concepts abstraits qui de ce fait, ont une autonomie l'un par rapport à l'autre.

34. La seconde réponse : l'expérience symbolique du corps

La seconde réponse à la question : quelles contraintes émanant de la structure physico-chimique de l'organisme pèsent sur l'intelligence ?, jamais clairement thématifiée, vise plutôt à reconnaître l'existence d'un lien secret de type symbolique - non plus dans le sens de l'écriture dans un espace logique mais dans un sens psychanalytique - entre l'existence des corps propres et la forme que prend l'intelligence. La réponse à la question est posée est alors : *oui*, il y a des contraintes physiques qui pèsent sur la forme de l'intelligence parce que l'intelligence est l'expression du corps incarné.

Je crois que cette deuxième réponse n'a jamais été thématifiée par Turing parce qu'il s'agit d'une réponse *fantasmatique*. Au départ, quand j'ai lu le texte de Turing dans lequel cette réponse a été élaborée, - c'est l'article de 1950, "Computing Machinery and Intelligence" - j'ai cru que l'intention de Turing était de faire une défense et illustration du projet d'intelligence artificielle. Mais dans le cours même de mon investigation, j'ai été frappé par un certain nombre d'indices concordants qui m'ont fait dire que le texte devait sans doute se lire autrement, non pas selon la structure rationnelle de l'inférence logique mais plutôt de manière associative parce que Turing parlait en fait beaucoup de lui dans ce texte et finalement assez peu du thème avoué de l'article.

J'ai donc délibérément fait une lecture *hors-sujet* en partant d'une hypothèse : les indices qui peuvent paraître arbitraires s'il sont pris séparément possèdent une cohérence quand on les réunit et qu'on les prend au sérieux. C'est cette cohérence que j'ai voulu dégager.

Je vais donc d'abord prendre le texte pour ce qu'il dit être, une défense et illustration du projet d'intelligence artificielle puis dans un deuxième temps, je m'intéresserai aux détails du texte et à leurs rapports mutuels.

34. 1. Le but déclaré du jeu de l'imitation et ses conséquences

L'article se veut être un plaidoyer pour ce que l'on appelle aujourd'hui "l'intelligence artificielle". Plutôt que de parler dans le vide et d'assurer dogmatiquement la chose possible, Turing propose un montage expérimental fondé sur un jeu, baptisé "jeu de l'imitation". Le jeu a pour but de montrer que la notion d'intelligence peut être attribuée aussi bien aux êtres humains qu'aux ordinateurs digitaux. Cela a deux conséquences :

1°. D'une part, l'intelligence est un concept abstrait indépendant des substrats physiques, corps biologique ou structure électrique et électronique.

2°. D'autre part, il est possible d'étendre son universalité à un continuum *encore plus large* que celui qui se trouve dans la nature. Idée presque darwinienne en elle-même, si l'on se rappelle que Darwin avait eu l'idée de la théorie de l'évolution en comparant les effets naturels de l'évolution découverts lors de son voyage autour du monde sur le *Beagle* et les effets artificiels de l'évolution chez les animaux progressivement sélectionnés dans les élevages anglais.

Le jeu se joue à trois : un homme (appelé A), une femme (appelée B) et un interrogateur dont le sexe n'est pas précisé. L'interrogateur est séparé de l'homme et de la femme (disons qu'il est dans une autre pièce). Le but du

jeu pour l'interrogateur est de poser des questions aux joueurs pour essayer de déterminer, des deux personnes qui se trouvent hors de sa vue, quel est l'homme et quelle est la femme. La communication se fait sans contact physique direct, c'est-à-dire par l'intermédiaire de l'écrit (disons qu'elle s'établit par un terminal d'ordinateur - Turing en 1950 parlait d'un télétype). Chaque joueur peut induire l'interrogateur en erreur si cela lui semble la meilleure stratégie pour que l'interrogateur ne parvienne pas à deviner son identité. Dans ce cas, le joueur doit *imiter* les réponses dont il suppose qu'elles seraient celles de l'autre joueur. D'où le nom du jeu. Le jeu est terminé quand l'interrogateur parvient à fournir la bonne réponse. Appelons ce jeu, le "jeu n°1".

Turing transforme ensuite les règles du jeu n°1. Dans la nouvelle formule, appelons la "jeu n°2", le joueur masculin A sera remplacé, à l'insu de l'interrogateur, par un ordinateur digital convenablement programmé pour qu'il imite le type de réponse du joueur masculin qu'il remplace.

Turing pose alors une question : l'interrogateur peut-il se rendre compte *par lui-même* que l'on est passé du jeu n°1 au jeu n°2, c'est-à-dire que les joueurs qu'il interroge sont d'une part un ordinateur et d'autre part une femme ? La réponse de Turing est *non*. La conséquence est que, pour un observateur extérieur au jeu, l'ordinateur digital est semblable aux humains du point de vue des réponses données, c'est-à-dire des manifestations de son intelligence. L'intelligence est donc bien un concept qui n'est pas lié à un substrat particulier mais peut s'incarner dans tout substrat physique du continuum naturel ou artificiel.

34. 2. But atteint par Turing

Je vais passer maintenant à une autre lecture du texte, lecture qui sera plus attentive aux détails qui sont habituellement éliminés de la description du jeu de l'imitation (certaines présentations du jeu en sont tellement éloignées que le jeu en devient méconnaissable).

34. 21. La viabilité du jeu

Turing parvient-il à justifier cette conclusion par le biais du jeu de l'imitation ? Je ne crois pas. Observons en effet les conditions effectives d'une partie. Puisque pour apporter la preuve de l'identité des êtres humains et des ordinateurs, l'interrogateur ne doit pas parvenir à déceler de différence dans leurs réponses, le jeu n'a aucune raison de s'arrêter : l'interrogateur doit donc continuer à poser des questions indéfiniment. Pour que le jeu s'arrête, il faut donc une intervention extérieure à la partie qui décide après un temps fini que l'interrogateur ne pourra jamais parvenir à faire la distinction entre les joueurs. Mais qui à l'extérieur du jeu décide de l'arrêter ? quelqu'un de convaincu du fait que l'interrogateur ne parviendra pas à ses fins. Mais par quel biais a-t-il été convaincu ? Cela ne peut pas être par le jeu puisque le jeu est en train de se dérouler. C'est donc que le personnage extérieur qui décide de l'arrêt d'une partie - appelons le l'arbitre - a d'autres moyens que le jeu d'être convaincu de l'échec de l'interrogateur c'est-à-dire du fait que l'intelligence peut s'incarner dans les substrats les plus divers. Mais à aucun moment Turing ne donne directement les raisons pour lesquelles l'arbitre pourrait être convaincu d'une chose pareille. Cette conviction doit donc être présupposée pour que le jeu fonctionne : Turing la présuppose donc pour lui-même qui a eu l'idée du jeu et tente par le jeu de renforcer cette conviction chez le lecteur.

Le jeu ne me paraît donc pas viable d'un point de vue logique. En revanche, la façon dont Turing le décrit du point de vue des stratégies mises en place par l'homme, la femme puis l'ordinateur me paraît très instructive de

l'idée cachée qui gouverne secrètement le texte. Étudions un instant ces stratégies qui n'ont rien, à mon sens, de rationnelles.

34. 22. Les stratégies des joueurs

La stratégie de la femme

Pour Turing, la stratégie de la femme consiste à dire la vérité sans chercher à cacher son identité. Si la femme ne bluffe pas, ce qui caractérise finalement l'absence de stratégie, c'est elle que l'interrogateur va réussir à identifier le plus vite. Mais pourquoi la femme devrait-elle dire la vérité ? Turing ne donne *aucune raison* qui justifierait un tel choix de la part de la femme. Une fois fait le choix d'une absence de stratégie pour la femme, les autres stratégies en découlent.

La stratégie de l'homme

Pour Turing, l'homme applique une stratégie qui consiste à tenter de se faire passer pour une femme et à tromper de ce fait l'interrogateur.

La stratégie de la machine

Dans le jeu n°2 dans lequel l'ordinateur remplace l'homme, Turing donne l'exemple d'un dialogue entre l'interrogateur et l'ordinateur ¹⁶:

«Question : Écrivez-moi, s'il vous plaît, un sonnet au sujet du pont sur le Forth.

Réponse : Ne comptez pas sur moi pour ça. Je n'ai jamais été capable d'écrire des poèmes.

Q : Ajoutez 34 957 et 70 764.

R : (En silence pendant 30 secondes puis donne la réponse) 105 621.

Q : Jouez-vous aux échecs ?

R : Oui.

Q: J'ai mon roi en C8 et aucune autre pièce. Il vous reste seulement votre roi en C6 et une tour en

A1. A vous de jouer. Que jouez-vous ?

R : La tour en A8 , échec et mat.»

Examinons les réponses données par la machine. La première réponse ne donne guère d'indication sur la différence entre les êtres humains et les ordinateurs : beaucoup d'humains se diraient incapables d'écrire des vers au pied levé. La troisième réponse paraît aussi sans grand intérêt : les pièces restantes dans cette partie d'échecs en font une partie tellement triviale que n'importe qui, être humain ou machine, pourrait facilement gagner. En revanche, la deuxième réponse est très intéressante : le calcul arithmétique est justement le point où les ordinateurs sont incontestablement meilleurs que les êtres humains. Or la réponse de l'ordinateur a ceci de particulier qu'elle est *fausse* : 34 957 et 70 764 font 105 721 et non pas 105 621. A ce sujet, la traduction française fait un contresens¹⁷ puisque le traducteur a cru qu'il s'agissait d'une *coquille* dans le texte original et qu'il a corrigé le texte sans mentionner sa correction ! Or cette prétendue coquille manifeste la stratégie de la machine : il s'agit de camoufler sa supériorité en calcul arithmétique en introduisant des fautes qui ressemblent à des fautes d'inattention *humaine*, c'est-

¹⁶ A. M. Turing, "Computing Machinery and Intelligence", *Mind*, vol LIX, n°236, 1950 : 434-435.

¹⁷ A. M. Turing et J.-Y. Girard, *La machine de Turing*, coll. Source du Savoir, Le Seuil, Paris, 1995 : 137.

à-dire liée à l'un ou l'autre sexe. Dans le cas d'un problème d'arithmétique, la machine n'imité plus la capacité de travestissement de l'homme mais la vérité de l'être humain, c'est-à-dire sa capacité d'erreur quel que soit son sexe. La machine dépasse donc le cas particulier masculin ou féminin tout en simulant l'être humain et c'est en ce sens qu'elle fait preuve d'une intelligence artificielle.

Mais toute l'analyse repose sur le fait que Turing assigne à la femme une absence de stratégie. Or rien ne justifie ce choix dans le texte, sinon des considérations touchant l'intelligence des femmes que seul Turing partage. Le rapport entre la gradation corporelle et la gradation intellectuelle est donc le produit d'une "théorie" sexuelle construite par Turing et dont l'article de 1950 témoigne (j'emploie le mot de théorie sexuelle dans le sens psychanalytique quand on dit que les enfants échafaudent des théories sexuelles fantasmatiques). Ce rapport établit la hiérarchie suivante du point de vue du corps et de l'intelligence : les femmes, les hommes, Turing et les ordinateurs digitaux.

34. 3. Interprétation du texte

Comment Turing en est-il venu à développer des idées pareilles ? Le texte, si on le lit entre les lignes, donne un certain nombre d'éléments qu'il serait trop long de détailler ici : différence entre l'engendrement par l'intérieur du corps chez les femmes et l'engendrement par la surface du corps chez les hommes; assimilation de la peau à une surface écrite; enfin identification de la machine à cette surface symbolique une fois réalisé le sacrifice de la surface du corps (j'emploie à dessein le terme de symbolique dans ses deux acceptions : symbolique comme écriture dans un espace logique et symbolique comme manifestation d'une vie fantasmatique). Je ne peux que renvoyer à mon article sur ce sujet dont je donne la référence en bibliographie.

34. 1. Le sacrifice

Je ne mentionnerai en détail ici que l'idée de sacrifice du corps qui a un rapport avec la conquête de l'universalité abstraite.

34. 11. L'apprentissage

Pour Turing, l'invention de son concept universel de machine est le fruit d'une abstraction qui est le résultat d'un long apprentissage. L'apprentissage est ce qui permet d'appliquer le principe d'induction et de parvenir à des résultats universels.

Turing fait un parallèle dans le texte entre l'apprentissage des enfants et l'apprentissage des machines et il remarque dans un autre texte sans dire si ce qu'il dit a à voir avec les humains ou avec les machines, que parvenir à devenir une machine universelle représente la forme la plus extrême de l'apprentissage¹⁸.

34. 12. L'induction

Il y a donc une induction à l'œuvre dans l'apprentissage dont Turing donne dans le texte de 1950 2 exemples

¹⁸A. M. Turing, *Intelligent Machinery*, p. 21 ; republié dans [*Mechanical Intelligence, Collected Works*, D. C. Ince ed., North-Holland, Amsterdam, 1992 : 87-105].

qui ont rapport tous les deux avec l'apprentissage des enfants :

Le premier exemple est le suivant : quand un enfant qui s'est brûlé a peur du feu et manifeste cette peur en évitant le feu, il applique à bon escient le principe d'induction. La peur du feu que l'enfant ressent lui est transmise comme un signal, par un canal émotionnel : la passion de la peur.

Le deuxième exemple, comme le fait remarquer Turing, porte sur des données culturelles et a pour objet un mauvais usage du principe d'induction. La plupart des petits Anglais pensent qu'il est idiot d'apprendre le français puisque tout le monde parle anglais ...

Dans le cas de données culturelles, il est facile de se méprendre sur l'universalité de la proposition dégagée par induction (en l'occurrence "tout le monde parle anglais") puisque l'anglais est une langue parmi d'autres et non pas la langue universelle. La mauvaise application du principe d'induction est liée à une mauvaise connaissance de l'usage des symboles (contrairement aux signaux qui passent par le canal du corps, la portée des symboles peut être mal appréciée, comme c'est le cas chez les petits Anglais monolingues).

Turing en conclut p. 445 :

«Les travaux et les coutumes de l'humanité ne semblent pas être un matériau très adapté à l'application de l'induction scientifique. On doit mener une enquête sur une très grande partie de l'espace-temps si on veut obtenir des résultats fiables.»

Et pourtant, cette induction scientifique est possible puisque Turing a inventé le concept universel de machine qui fonctionne grâce à un langage universel de nature symbolique. Quel a donc été l'apprentissage nécessaire pour aboutir à ce résultat ?

34. 13. Les punitions et les récompenses

On se rappelle que Turing souscrivait à l'idée behavioriste selon laquelle l'apprentissage se faisait par le biais du stimulus-réponse. Turing indique que l'apprentissage exige dans ce cas un système de récompense et de punition que l'on peut transposer dans le cas de l'apprentissage des machines. Turing fait remarquer à ce propos¹⁹ :

«L'utilisation de punitions et de récompense peut au mieux faire partie du processus d'éducation. Grossièrement parlant, si le maître n'a pas d'autres moyens de communication avec l'élève, la quantité d'information s qui peut lui parvenir ne dépasse pas le nombre total des récompenses et punitions utilisées. D'ici à ce qu'un enfant ait appris à répéter, "Casabianca", il éprouverait sans doute quelques désagréments s'il ne pouvait découvrir le texte que par la méthode des 20 questions", où chaque "non" impliquerait qu'il reçoive un coup. Il est donc nécessaire d'avoir d'autres canaux que des canaux émotionnels. Si ces derniers sont disponibles, il est possible d'enseigner à la machine, par un système de sanctions et de récompenses, à obéir aux ordres donnés dans un certain langage, par exemple; un langage symbolique. L'utilisation de ce langage diminuera de beaucoup le nombre de punitions et de récompenses requises.»

Ce paragraphe paraît bien anodin et pourtant, à y regarder de près, il est plus complexe qu'il n'y paraît.

Le paragraphe décrit l'apprentissage par cœur d'un texte appelé "Casabianca". Ce texte est une poésie écrite

¹⁹ A. M. Turing, "Computing Machinery and Intelligence", *Mind*, vol LIX, n°236, 1950 : 457.

par une poétesse du nom de Felicia Hemans (1793-1835) qui raconte un fait d'armes français : on rapporte que pendant la campagne d'Égypte de 1798, lors de la bataille du Nil en rade d'Aboukir où les Français furent vaincus par la flotte anglaise commandée par Nelson, le vaisseau amiral français était commandé par le capitaine Louis Casabianca et que, même après la mort de l'amiral français Brueys d'Aigalliers, il continua à se battre jusqu'au bout. Son fils âgé de treize ans, Giacomo, refusa de quitter le navire et mourut avec son père.

La strophe la plus souvent citée de ce poème est celle-ci :

«Le garçon se tint debout sur le pont en feu
duquel tous sauf lui avaient fui;
Les flammes éclairant les débris de la bataille
brillaient autour de lui au dessus des morts.»

Si l'on analyse le contenu de la poésie, on voit qu'il s'agit aussi d'un exemple d'application du principe d'induction.

Giacomo Casabianca aurait du fuir le navire avec tous les autres marins, appliquant en cela à juste titre le principe d'induction qui consiste à avoir peur du feu (c'était le premier exemple dans "Computing Machinery and Intelligence"). Pourquoi refuse-t-il d'appliquer le principe d'induction ? Par fidélité envers son père et pour suivre l'exemple de son sacrifice. Dès lors, cet exemple devient plus qu'un exemple particulier : il prend la valeur d'un symbole, c'est-à-dire qu'il joue le rôle d'un message universel.

Ainsi, dans le poème, l'expérience portant sur ce que Turing appelait "une grande partie de l'espace-temps" nécessaire pour devenir une machine universelle est une expérience radicale, celle de la mort. Giacomo Casabianca suit l'exemple de son père en mettant en péril son corps en en faisant un sacrifice par le feu : son sacrifice a pour but d'assurer la transmission d'un message linguistique universel en mettant en péril son corps qui sert de simple contenant au message.

Dans le deuxième exemple, on se rappelle que les petits Anglais au contraire, considèrent que leur langue maternelle est une langue universelle parce qu'ils n'ont pas fait, comme Giacomo, l'épreuve du sacrifice.

C'est sans doute la raison pour laquelle Turing dénie à la femme une stratégie au jeu de l'imitation : sa langue est toujours trop maternelle, dans la mesure où elle a été apprise par un canal émotionnel, et c'est ce qui la rend incapable de se détacher du corps pour affronter l'abstraction de la langue universelle symbolique.

34. 13. Indices dans la vie de Turing

Certains indices dans la vie de Turing renvoient à ce poème et à l'expérience qu'il décrit et on ne peut pas manquer de se demander ce qu'il y a derrière l'expression "une grande enquête dans l'espace-temps", enquête qui a été nécessaire à Turing pour inventer son concept de machine.

La biographie de Turing, écrite par le mathématicien Andrew Hodges, donne à ce sujet d'utiles renseignements.

Turing est le fils d'un administrateur colonial qui a vécu en Inde. Turing est né à Londres et a été séparé à peu près immédiatement de son père, retourné travailler à Madras puis au bout de quelques mois de sa mère retournée auprès de son mari. Il a donc eu une enfance très solitaire dans une famille chez qui il avait été, lui et son frère, mis en placement. Un autre détail révélé par Hodges a aussi son importance : Turing a été circoncis dans son enfance et il

dira par la suite que cette circoncision a décidé en grande partie de son orientation sexuelle.

34. 14. Rapprochement entre la vie de Turing et le poème

Le rapprochement que j'opère entre ces détails biographiques et le poème est le suivant : il faut sans doute voir dans cette circoncision une atteinte portée à l'intégrité du corps par une blessure faite à la peau comme le feu dans le poème portait atteinte à l'intégrité physique de Giacomo. On trouve d'ailleurs dans une lettre à ses parents (11 février 1923) une allusion au poème. L'allusion au poème est en rapport avec l'une de ses inventions : celle-ci consiste à construire un film sous forme d'un ruban de 16 images successives sur lesquelles étaient dessinées les images correspondant au récit. Mais le récit représenté n'est pas celui du poème de "Casabianca" mais celui d'une parodie de "Casabianca". Il faut insister un instant sur cette parodie. Le poème de Casabianca a eu un curieux destin dans le système éducatif britannique : appris par cœur par des générations d'écoliers jusqu'à aujourd'hui, on oublie généralement le poème original ainsi que le nom de l'auteur du poème, mais, en revanche, on fait, à partir du poème, de nombreux vers parodiques, souvent à consonance grossière ou grivoise²⁰. La parodie du poème telle qu'elle est rapportée par Turing "le garçon se tint à la table à thé" ["The boy stood at the tea table"] fait donc partie de la tradition écolière, que les parents de Turing connaissent certainement, ainsi que les connotations souvent sexuelles de ce genre de parodie, même si ce n'est pas celles-là que Turing rapporte dans sa lettre.

Turing construit donc un film de 16 images où sont reproduites les images correspondant au récit parodié. La construction du film associe donc, si on met ensemble tous ces détails, l'idée du sacrifice selon Turing et la question de la sexualité, grâce à la construction d'une machine qui décompose en étapes successives le récit parodié du poème.

Comment ne pas retrouver ici, dans ce souvenir d'enfance, une trace des constructions futures de Turing, le jeu de l'imitation et sa parodie-imitation des comportements de l'autre sexe, ainsi que l'idée même d'une machine ayant dépassé la différence des sexes dont la fonction est de décomposer un récit parodié d'un poème écrit par une femme en étapes successives représentées dans des cases ?

Le ruban du film a peut-être pour fonction de retrouver abstraitement l'intégrité perdue du corps, brûlé dans la circoncision et dans le poème et donner un sens par le bais d'un récit à cette perte d'intégrité.

34. 15. Conséquences de cette analyse

Je ne veux pas, par cette analyse, réduire la portée des découvertes de Turing mais seulement montrer qu'elles s'appuient sur un matériau fantasmatique qui lui est éminemment personnel.

Je pourrais montrer comment fonctionnent d'autres détails du texte. Mais les multiples indices que j'ai pu retrouver semblent concorder et montrer que l'inférence logique n'est pas le tout du psychisme humain et qu'en

²⁰ Je cite une parodie, parmi beaucoup d'autres :
"Le garçon se tient sur le pont en feu
Ses pieds couverts de cloques
Il avait déchiré sa chemise de nuit jusqu'en bas
Et avait dû porter celle de sa sœur."
["The boy stood at the burning deck
His feet all covered in blisters
He split his nightshirt down the back
and had to wear his sister's."]

particulier la question de l'invention liée à celle de l'apprentissage semble receler encore bien des mystères qui tournent autour du statut à accorder à la langue, à ses figures de style et à ses double sens.

Pour Turing, la question fondamentale qui a nourri son imaginaire de savant pendant tout sa vie, c'est celle de la nature de l'abstraction, ce qui la rend possible et ce qui nous la rend accessible. Les indices que donne le texte de 1950 donnent à penser que, chez Turing, l'abstraction est une conquête sur le corps : elle exige de sacrifier son corps pour elle. L'invention du concept de machine de Turing a donc partie liée avec la vie fantasmatique de son auteur, ce qui serait une banalité invérifiable si Turing n'avait pas laissé dans son texte suffisamment d'indices qui permettent de reconstituer ce qui, d'un point de vue psychologique, a présidé à l'élaboration du concept de machine de Turing.

Que faut-il en conclure du point de vue de l'invention de ce concept ?

Le problème qui se pose est celui du rapport entre l'universalité d'un concept (en l'occurrence celui de machine de Turing) et la particularité du point de vue psychologique qui l'a rendu possible (à savoir la vie fantasmatique hautement irrationnelle de l'individu Turing). C'est ce point que je vais aborder en conclusion.

Conclusion

J'avais commencé en disant que j'essaierai de préciser les apports de Turing à la question de la nature de l'inférence. Ces apports sont doubles, dans les sciences dures et dans les sciences humaines et donc inévitablement, dans leur rapport.

- Dans les sciences dures, les différents points de ce que j'ai appelé la "thèse cognitive de Turing" ont montré quels étaient ces apports personnels : Turing a précisé ce qu'il fallait entendre par "mécanique" dans le rapport d'inférence; il a intégré cette idée à un corpus scientifique extrêmement large, allant de la calculabilité des nombres réels à des questions proprement logiques touchant la nature de la déductibilité dans les systèmes formels, mais aussi à la théorie des jeux (dont je n'ai pas eu le temps de parler) et à des modèles informatiques de processus chimiques à l'œuvre dans la morphogenèse.

- Dans les sciences humaines, le bilan est, comme toujours, plus controversé. Le modèle computationnel de l'esprit qui identifie les processus psychologiques du cerveau à des processus modélisables par machine de Turing est incontestablement une idée révolutionnaire dans la mesure où, pour la première fois, il semble que l'on ait trouvé le *bon point de vue* pour étudier mathématiquement des processus psychologiques. Que faut-il entendre par le "bon point de vue" ? A mon sens, le bon point de vue c'est celui qui étudie mathématiquement des processus psychologiques *sans confondre* le modèle mathématique et la réalité de l'esprit puisque l'on sait, si l'on accepte ma lecture de l'article de 1950, que le modèle a une portée *limitée* à un certain type de processus, ceux qui ont une forme inférentielle. Mais il y a beaucoup d'autres processus psychologiques qui n'ont pas cette forme et en particulier les processus liés à l'*invention* des concepts, puisque ces processus psychologiques sont éminemment personnels et fantasmatiques. Il faut faire jouer alors, pour en rendre compte, une structure autre que la structure inférentielle, celle des associations sémantiques qui ne relèvent pas de l'inférence logique.

- On voit donc mieux, je l'espère, en quel sens on peut dire que Turing a déplacé la frontière entre sciences dures et sciences humaines : cette frontière s'est déplacée parce qu'il a élaboré un *modèle*, général mais limité - ce qui en fait un modèle véritablement scientifique - de certains processus psychologiques. Mais l'apport de Turing, à mon sens, ne s'arrête pas là. Il a en effet montré quels processus psychologiques étaient susceptibles d'être analysés par

son modèle, à savoir les processus de type inférentiel, c'est-à-dire qu'il a établi une limite entre ceux qui sont modélisables et ceux qui ne le sont pas dans le cadre de son modèle.

Je voudrais faire une dernière remarque : on est toujours étonné du nombre des processus psychologiques modélisables par l'intelligence artificielle, qui vont de la perception aux fonctions intellectuelles supérieures. Mais j'avoue que mon étonnement a été aussi grand quand j'ai découvert que l'article qui a servi d'acte fondateur à l'intelligence artificielle manifestait la présence de processus si éloignés de la forme inférentielle. Ce double étonnement ne doit pas nous quitter et doit renforcer notre conviction à tous qu'il y a encore beaucoup de travail à faire pour parvenir à avoir une idée juste de la globalité des processus psychologiques à l'œuvre dans la pensée.

Jean Lassègue

Bibliographie

1. Œuvres de Turing :

Turing A. M., *Collected Works of A. M. Turing*, vol 1- 4, North-Holland Publishing Company, 1992, 1993. (Trois volumes parus; vol. 1 : "Pure mathematics", J. L. Britton ed.; vol. 3 : "Mechanical Intelligence", D. C. Ince ed.; vol 4 : "Morphogenesis", P. T. Saunders ed.; un autre à paraître : vol. 2 : "Mathematical Logic").

- (1936), 'On Computable Numbers with an Application to the Entscheidungsproblem', Proceedings of the London Mathematical Society, 42 : 230-265; republié dans [Davis M. ed., *The Undecidable* Raven Press, Hewlett, New-York, 1965, p. 115-154]; traduit en français dans A. M. Turing, J.-Y. Girard, *La machine de Turing*, coll. Source du Savoir, Le Seuil, Paris, 1995.
- (1937), 'Computability and lambda-definability', Journal of Symbolic Logic, 2 : 153-163.
- (1939), 'Systems of logic based on Ordinals', Proceedings of the London Mathematical Society, ser. 2, vol. 45 : 161-228 republié dans [Davis M., *The Undecidable* Raven Press, Hewlett, New-York, p. 154 sq]
- (1945), 'Proposals for Development in the Mathematics Division of an Automatic Computing Engine (ACE)', Executive Committee NPL, 1946, Crown Copyright Reserved, H. M. S. O.; republié dans [*Mechanical Intelligence, Collected Works*, D. C. Ince ed., North-Holland, Amsterdam, 1992 : 1-86].
- (1947), 'Lecture to the London Mathematical Society', 20 february 1947; republié dans [*Mechanical Intelligence, Collected Works*, D. C. Ince ed., North-Holland, Amsterdam, 1992 : 87-105].
- (1948), 'Intelligent Machinery', Executive Committee NPL, 1948, 1-20, Crown Copyright Reserved, H. M. S. O.; republié dans [*Mechanical Intelligence, Collected Works*, D. C. Ince ed., North-Holland, Amsterdam, 1992 : 107-127].
- (1948) 'Practical Forms of Type Theory', Journal of Symbolic Logic, 13 (1948)
- (1950), 'Computing Machinery and Intelligence', Mind, vol LIX, n°236, 1950 : 433-460; republié dans [*Mechanical Intelligence, Collected Works*, D. C. Ince ed., North-Holland, Amsterdam, 1992 : 133-160]; traduit en français sous le titre "Les ordinateurs et l'Intelligence" dans [*Pensée et Machine* Anderson A. R. ed., Champ Vallon, Seyssel, 1983] : 39-67 et dans A. M. Turing, J.-Y. Girard, *La machine de Turing*, coll. Sources du Savoir, Le Seuil, Paris, 1995.

- (1950), 'The Word Problem in Semi-Groups with Cancellation', *Ann. Math.* (Princeton) 52, revu, clarifié et corrigé par W. W. Boone, *Journal of Symbolic Logic*, 17 (1952)
- (1952), 'The Chemical Basis of Morphogenesis', *Phil. Trans. Roy. Soc. B* 237 : 37-72; republié dans [*Morphogenesis, Collected Works*, North-Holland, 1992] : 1-36.
- (1953), 'Digital Computers Applied to Games', Bowden B. V. ed., *Faster Than Thought*, Pitman, London, 1953 : 286-310; republié dans [*Mechanical Intelligence, Collected Works*, D. C. Ince ed., North-Holland, Amsterdam, 1992 : 161-185].
- (1953), 'Some Calculations of the Riemann Zeta Function', *Proceedings of the London Mathematical Society*, (3) 3.
- (1954), 'Solvable and Unsolvable Problems', *Science News*, 31 : 7-23, republié dans [*Mechanical Intelligence, Collected Works*, D. C. Ince ed., North-Holland, Amsterdam, 1992 : 187-203].

2. Sur Turing :

- Davis M., (1982), 'Why Gödel didnt Have Church's Thesis', *Information and Control*, 54, (1982) : 3-24.
- (1987), 'Mathematical Logic and the Origin of Modern Computers', *Studies in the Working of Mathematics, The Mathematical Association of America*, 1987, reprint in [*The Universal Turing machine*, Herken R. ed., Oxford Science Publications, Oxford University Press, 1988] : 149 -174.
- (1988), 'Influences of Mathematical Logic on Computer Science' in [*The Universal Turing machine*, Herken R. ed., Oxford Science Publications, Oxford University Press, 1988] : 315-326
- Delahaye J.-P., (1993), "Formulations mathématiques de la question : le monde est-il récursif ?", *Cahiers du Crea*, n°15 : Méthodologie de la science empirique (2), Alain Boyer dir., Ecole Polytechnique, Paris, janvier 1993 : 185-225.
- Detlefsen M., (1990), 'Brouwerian intuitionism', *Mind*, 99 : 501-534.
- Dubucs J. et Mosconi J., (1991) "Le mécanisme et la thèse de Church-Turing", *Courrier du CNRS*, n°79 : 116.
- Gandy R. O., (1969), 'The Concept of Computability' in [*Scientific Thought 1900-1960, a Selective Survey*, ed. Harré R., Clarendon Press, Oxford]: 1-4.
- (1980), 'Church's thesis and principles for mechanisms' in [*The Kleene Symposium*, J. Barwise, J.J Keisler and K. Kuchen eds., 123-145, Amsterdam, North-Holland Publ. Co.] : 123-145.
- (1988), 'The Confluence of Ideas in 1936' in [*The Universal Turing machine*, Herken R. ed., Oxford Science Publications, Oxford University Press, 1988] : 55-110.
- Gödel K., (1972), 'Some remarks on the undecidability results', reproduit dans [*Collected Works*, t. II, Oxford University Press, Oxford, 1990] : 305-306.
- Hao Wang, (1993), 'On Physicalism and Algorithmism : Can Machines Think ?', *Philosophia Mathematica*, series III, vol. 1, 1993 : 97-138.
- Hodges A., (1983), *Alan Turing, the Enigma of Intelligence*, Unwin Paperbacks, London.
- (1988), 'Alan Turing and The Turing Machine' in [*The Universal Turing machine*, Herken R. ed., Oxford Science Publications, Oxford University Press, 1988] : 3-15.
- Kleene S. C., (1981), 'Origins of recursive function theory', *Ann. Hist. Comp.* 3 (1981) : 52-67.
- (1988), 'Turing's Analysis of Computability and Major Applications of It', in [*The Universal Turing machine*, Herken R. ed., Oxford Science Publications, Oxford University Press, 1988] : 17-54.
- Kreisel G., (1987), 'Church's Thesis and the Ideal of Informal Rigour', *Notre Dame Journal of Symbolic Logic*, 28, 4, October 1987 : 499-519.
- Lassègue J., (1993), "Le test de Turing et l'énigme de la différence des sexes" in [*Les contenants de pensée*, Anzieu D. et coll., Dunod, Paris, 1993] : 145-195.
- Leibniz G. W., (1703), *Nouveaux Essais sur l'entendement humain*, in [*die Philosophischen Schriften*, ed. Gerhardt, Band V, Georg Olms Verlag, Hildesheim, reprint 1978].
- *Opuscules et fragments inédits de Leibniz*, ed. Couturat, Felix Alcan, Paris, 1901.
- Michie D., (1993), 'Turing's Test and conscious thought', *Artificial Intelligence*, 60 (1993), 1-22.

- Mosconi J., (1989), *La constitution de la théorie des automates*, thèse de doctorat d'Etat, Université de Paris I, imprimé par l'Atelier National de reproduction des thèses, Université de Lille III.
- (1991), "Sur quelques capacités et incapacités des machines", Bulletin de la société française de philosophie, t. LXXXV, Armand Colin, Paris : 81-115.
- Mundici D. et Sieg W., (1995), 'Paper Machines', Philosophia Mathematica, series III, vol. 3, 1995 : 5-30.
- Nelson R. J., (1987) 'Church Thesis and Cognitive Science', Notre-Dame Journal of Symbolic Logic, XXVIII-4, 1987 : 581-614
- Penrose R., (1988), 'On physics and mathematics of thought' in [*The Universal Turing machine*, Herken R. ed., Oxford Science Publications, Oxford University Press, 1988] : 491-522.
- Rosen R., (1988), 'Effective Processes and Natural Law' in [*The Universal Turing machine*, Herken R. ed., Oxford Science Publications, Oxford University Press, 1988] : 523-537.
- Saunders P. T., (1993), "L'évolution des formes biologiques", Intellectica, n° 16, 1993/1 : 61- 83.
- Shanker S., (1995), 'Turing and the Origins of AI', Philosophia Mathematica, series III, vol. 3, 1995 : 52-86.
- Shapiro S., (1995), 'Reasoning, Logic and Computation', Philosophia Mathematica, series III, vol. 3, 1995 : 31-52.
- Webb J. C., (1980) *Mechanism, Mentalism and Metamathematics : an Essay on Finitism*, Reidel Publishing Company, Dordrecht.
- (1990), "Introductory note to 1972a" in [*Gödel K., Collected Works*, vol. II, Oxford University Press, Oxford, 1990] : 281-304.