Une méthode pour le calcul de la fonction zeta

PAR A. M. TURING.

Une série asymptotique pour la fonction zeta a été trouvée par Riemann et a été publiée par Siegel [1], et appliquée par Titchmarsh [2] pour calculer les positions approximatives de quelques-uns des zéros de la fonction. Il est difficile d'obtenir des estimations satisfaisantes pour les restes avec les séries asymptotiques, comme on peut le voir dans le premier de ces deux articles de Titchmarsh, à moins que t ne soit très grand. Dans le présent article, une méthode sera décrite, qui, comme la formule asymptotique, est basée sur une approximation de l'équation fonctionnelle ; elle est applicable pour toutes les valeurs de s. Il est vraisemblable qu'elle soit plus intéressante pour un domaine de t dans lequel t n'est ni trop petit de telle façon que la méthode de sommation d'Euler-Maclaurin puisse être utilisée (e.g. > 30) ni trop grand pour que la formule asymptotique de Riemann-Siegel puisse l'être (e.g. $t < 1000$).

Pour le dire grossièrement, la méthode consiste à utiliser l'approximation de l'équation fonctionnelle pour la fonction zeta, avec le reste exprimé comme une intégrale, que pour le moment nous écrivons comme $\int_{-\infty}^{\infty} h(x)dx$. Nous approximons l'intégrale par la somme triviale $\sum_{k=-K}^{K} \frac{1}{\kappa} h\left(\frac{k}{\kappa}\right)$ et nous trouvons que, si certaines modifications sont faites dans la "série principale", cela donne un résultat remarquablement précis ; quand le nombre de termes pris est $T = 2K+1$, l'erreur est d'un ordre de grandeur de $e^{-\frac{1}{2}\pi T}$. Les fonctions theta fournissent un autre cas de ce phénomène. Nous avons l'identité

$$\sum_{n=-\infty}^{\infty} e^{-\pi n^2 x} = \frac{1}{\sqrt{x}} \sum_{n=-\infty}^{\infty} e^{-\pi n^2/x}$$

pour tout x positif. Par conséquent,

$$\left| \frac{1}{\kappa} \sum_{n=-\infty}^{\infty} e^{-\pi n^2/\kappa^2} - \int_{-\infty}^{\infty} e^{-\pi v^2} dv \right| = 2 \sum_{n=1}^{\infty} e^{-\pi n^2 \kappa^2} < \frac{2}{1-e^{-2\pi}} e^{-\pi\kappa^2} \qquad (\text{si } \kappa \geqslant 1),$$

$$\left| \frac{1}{\kappa} \sum_{n=-K}^{K} e^{-\pi n^2/\kappa^2} - \int_{-\infty}^{\infty} e^{-\pi v^2} dv \right| < \frac{2}{1-e^{-2\pi}} e^{-\pi\kappa^2} + \frac{\kappa e^{-\pi K^2/\kappa^2}}{\pi K},$$

$$< e^{-\pi K} \left(\frac{2}{1-e^{-2\pi}} + \frac{1}{\pi} \right) \qquad (\text{si } K = \kappa^2).$$

Comme nous l'avons prouvé ci-dessus, cette inégalité dépend entièrement de la forme spéciale de la fonction $e^{-\pi u^2}$. Mais nous pouvons aussi le prouver de cette manière. Nous intégrons la fonction $e^{-\pi u^2}/(1 - e^{-2\pi i u \kappa})$ autour du rectangle de sommets $\pm R \pm i\kappa$. À la limite $R \to \infty$, nous obtenons,

[Reçu le 7 Mars 1939. Lu le 16 Mars 1939]
dans *Proceedings of the London Mathematical Society. Second Series*, vol. 48, p. 180-197, 1943.
Traduction de l'anglais au français : Denise Vella-Chemla, mars 2021.

1. C. L. Siegel, "Über Riemanns Nachlass zur analytischen Zahlentheorie", Quell. Gesch. Math., B, 2 (1931), p. 45-80.
2. E. C. Titchmarsh, "The zeros of the Riemann zeta-function", Proc. Royal Soc. (A), 151 (1935), p. 234-255 ; également 157 (1936), p. 261-263.

par le théorème des résidus,

$$\left(\int_{-\infty-i\kappa}^{\infty-i\kappa}+\int_{\infty+i\kappa}^{-\infty+i\kappa}\right)\frac{e^{-\pi u^2}}{1-e^{-2\pi iu\kappa}}du=\sum_{k=-\infty}^{\infty}\frac{1}{\kappa}e^{-\pi k^2/\kappa^2},$$

i.e.

$$\sum_{k=-\infty}^{\infty}\frac{1}{\kappa}e^{-\pi k^2/\kappa^2}-\int_{-\infty-i\kappa}^{\infty-i\kappa}e^{-\pi u^2}du=\int_{-\infty-i\kappa}^{\infty-i\kappa}\frac{e^{-\pi u^2}e^{-2\pi iu\kappa}}{1-e^{-2\pi iu\kappa}}du+\int_{\infty+i\kappa}^{-\infty+i\kappa}\frac{e^{-\pi u^2}e^{-2\pi iu\kappa}}{1-e^{2\pi iu\kappa}}du$$

Le chemin d'intégration du côté gauche de l'équation peut être remplacé par l'axe réel, alors que le côté droit est inférieur en module à

$$\frac{2}{1-e^{-2\pi\kappa^2}}\int_{-\infty}^{\infty}\Big|\exp[-\pi(u-i\kappa)^2-2\pi i(u-i\kappa)\kappa]\Big|du$$

$$\leqslant\frac{2}{1-e^{-2\pi\kappa^2}}\int_{-\infty}^{\infty}\exp(-\pi u^2-\pi\kappa^2)du\leqslant\frac{2e^{-\pi\kappa^2}}{1-e^{-2\pi}}\qquad(\text{si }\kappa\geqslant 1).$$

Cet argument peut être utilisé dans des cas plus généraux, mais en bougeant le chemin d'intégration, nous pouvons rencontrer des singularités de l'intégrande, ce qui modifiera le résultat.

Nous basons notre calcul sur la représentation intégrale de la fonction zeta due à Riemann [3].

1. *Évaluation d'une intégrale définie.*

Posons

$$G(u)=\int_{0\nearrow 1}\frac{\exp(i\pi z^2+2\pi iuz)dz}{e^{i\pi z}-e^{-i\pi z}},$$

où $0\nearrow 1$ signifie que l'intégration s'effectue selon une droite de $-\epsilon\infty$ à $\epsilon\infty$ en coupant l'axe réel entre 0 et 1, et $\epsilon=e^{\frac{1}{2}\pi i}$. Nous dénotons une intégrale similaire sur $-1\nearrow 0$ par $G_1(u)$. Alors, par un changement de variables, nous obtenons

$$G_1(u)=G(u-1)e^{-2\pi iu},$$

et, par le théorème des résidus,

$$G(u)=G_1(u)+1.$$

En multipliant l'identité

$$\frac{1}{e^{i\pi z}-e^{-i\pi z}}=e^{-i\pi z}+\frac{e^{-2\pi iz}}{e^{i\pi z}-e^{-i\pi z}}$$

par $\exp(i\pi z^2+2\pi iuz)$, et en intégrant sur $0\nearrow 1$, on obtient

$$G(u)=\epsilon e^{-i\pi(u-\frac{1}{2})^2}+G(u-1).$$

En combinant ces résultats, on a

$$G(u)=\frac{1}{1-e^{-2\pi iu}}-\frac{e^{-i\pi u^2}}{e^{i\pi u}-e^{-i\pi u}}\tag{1.1}$$

3. Siegel, *loc. cit.* 24.

2. Une représentation intégrale de la fonction zeta.

On intègre $z^{-s}/(1-e^{-2\pi iz})$ autour d'une courbe L qui peut être prise comme étant constituée d'une ligne droite de $i\infty$ à $\frac{1}{2}$, d'un semi-cercle de $\frac{1}{2}$ à $-\frac{1}{2}$ qui se trouve dans le demi-plan inférieur et d'une ligne droite de $-\frac{1}{2}$ à $i\infty$. Nous définissons z^{-s} de telle façon qu'il ait sa valeur habituelle sur l'axe réel positif, et soit continu excepté sur l'axe imaginaire positif. L'intégrale autour d'une quelconque partie d'un cercle centré sur 0 tend vers 0 lorsque le rayon tend vers l'infini avec des valeurs appropriées, sous la condition que $\Re s > 1$, et ainsi dans ce cas, on a

$$\begin{aligned}\int_L \frac{z^{-s}dz}{1-e^{-2\pi iz}} &= 2\pi i\,(\text{somme des résidus de} \frac{z^{-s}}{1-e^{-2\pi iz}} \text{ aux entiers autres que } 0)\\ &= \sum_{m=1}^{\infty}(m^{-s}+e^{i\pi s}m^{-s}) = \zeta(s)(1+e^{i\pi s}).\end{aligned}$$

Cela donne le prolongement analytique de $\zeta(s)$ sur le plan entier excepté possiblement sur les entiers pairs. Maintenant, par (1.1),

$$\begin{aligned}\int_L u^{-s}\left(\frac{1}{1-e^{-2\pi iu}} - \frac{e^{-i\pi u^2}}{e^{i\pi u}-e^{-i\pi u}}\right)du &= \int_L u^{-s}\left[\int_{0\nearrow 1}\frac{\exp(i\pi z^2+2\pi iuz)dz}{e^{i\pi z}-e^{-i\pi z}}\right]du\\ &= \int_{0\nearrow 1}\frac{e^{i\pi z^2}}{e^{i\pi z}-e^{-i\pi z}}\left(\int_L u^{-s}e^{2\pi iuz}du\right)dz\\ &= -2(2\pi)^{s-1}\sin s\pi\Gamma(1-s)e^{\frac{1}{2}is\pi}\int_{0\nearrow 1}\frac{e^{i\pi z^2}z^{s-1}}{e^{i\pi z}-e^{-i\pi z}}dz,\end{aligned}$$

i.e.

$$\zeta(s)(1+e^{i\pi s}) = \int_L \frac{e^{-i\pi u^2}u^{-s}du}{e^{i\pi u}-e^{-i\pi u}} - 2(2\pi)^{s-1}\sin s\pi\Gamma(1-s)e^{\frac{1}{2}is\pi}\int_{0\nearrow 1}\frac{e^{i\pi z^2}z^{s-1}dz}{e^{i\pi z}-e^{-i\pi z}}.$$

Dans la première intégrale, on peut remplacer L par les deux lignes $0\downarrow 1$ et $-1\uparrow 0$, et on obtient

$$\begin{aligned}\int_L \frac{e^{-i\pi u^2}u^{-s}du}{e^{i\pi u}-e^{-i\pi u}} &= \int_{0\downarrow 1}\frac{e^{-i\pi u^2}u^{-s}du}{e^{i\pi u}-e^{-i\pi u}} + \int_{-1\uparrow 0}\frac{e^{-i\pi u^2}u^{-s}du}{e^{i\pi u}-e^{-i\pi u}}\\ &= \int_{0\downarrow 1}\left(\frac{e^{-i\pi u^2}u^{-s}}{e^{i\pi u}-e^{-i\pi u}} - \frac{e^{i\pi u^2}u^{-s}e^{i\pi s}}{e^{-i\pi u}-e^{i\pi u}}\right)du.\end{aligned}$$

La courbe peut maintenant être remplacée par $0\nwarrow 1$ si on change le signe, et ainsi on obtient

$$-\int_{0\nwarrow 1}(1+e^{i\pi s})\frac{e^{-i\pi u^2}u^{-s}du}{e^{i\pi u}-e^{-i\pi u}}.$$

La fonction zeta s'exprime alors sous la forme

$$\zeta(s) = -\int_{0\nwarrow 1}\frac{e^{-i\pi u^2}u^{-s}du}{e^{i\pi u}-e^{-i\pi u}} - 2(2\pi)^{s-1}\sin\frac{1}{2}s\pi\Gamma(1-s)\int_{0\nearrow 1}\frac{e^{i\pi z^2}z^{s-1}dz}{e^{i\pi z}-e^{-i\pi z}} \tag{2.1}$$

et le calcul de $\zeta(s)$ se réduit à celui de l'intégrale

$$I(s) = \int_{0\nearrow 1}\frac{e^{i\pi z^2}z^{-s}dz}{e^{i\pi z}-e^{-i\pi z}} = \int_{0\nearrow 1}h(z)dz\ ; \tag{2.2}$$

pour

$$\overline{I(s)} = -\int_{0\nwarrow 1} \frac{e^{-i\pi z^2} z^{-\overline{s}} dz}{e^{-i\pi z} - e^{i\pi z}}$$

Si on multiplie les deux côtés de (2.1) par $\Gamma(\frac{1}{2}s)\pi^{-\frac{1}{2}s}$, et si on utilise la relation

$$\Gamma(1-s) = \frac{2^{-s}\pi^{\frac{1}{2}}\Gamma(\frac{1}{2}-\frac{1}{2}s)}{\sin\frac{1}{2}s\pi\Gamma(\frac{1}{2}s)},$$

on obtient

$$\zeta(s)\Gamma\left(\frac{1}{2}s\right)\pi^{-\frac{1}{2}s} = \Gamma\left(\frac{1}{2}s\right)\pi^{-\frac{1}{2}s}\overline{I(\overline{s}} - \Gamma\left(\frac{1}{2}-\frac{1}{2}s\right)\pi^{is-\frac{1}{2}}I(1-s), \tag{2.3}$$

et sur la droite critique, cela est égal à

$$-2\Re\Gamma\left(\frac{1}{2}s\right)\pi^{-\frac{1}{2}s}I(\overline{s}). \tag{2.4}$$

Pour les points sur cette droite, il y a, ainsi, seulement une intégrale réelle à calculer. Pour les points à l'extérieur de cette droite, il y a quatre intégrales réelles.

3. La méthode de calcul.

Soit μ un nombre réel positif compris entre les entiers m et $m+1$. Alors

$$I(s) = \int_{0\nearrow 1} h(z)dz = \int_{m\nearrow m+1} h(z)dz - \sum_{r=1}^{m} r^{-s}.$$

Maintenant, soit κ un nombre réel positif et posons

$$g(z) = \frac{h(z)}{1-\exp\left[-2\pi\kappa\epsilon(z-\mu)\right]}.$$

La fonction g a des pôles simples aux points entiers autres que 0, et en les points $p_k = \mu + \epsilon k/\kappa$, où k est un entier ; sinon, elle est régulière, excepté sur l'axe imaginaire positif fermé. Le résidu sur un entier non nul r est

$$\frac{r^{-s}}{2\pi i\{1-\exp\left[-2\pi\kappa\epsilon(r-\mu)\right]\}},$$

et sur p_k, il vaut

$$\frac{h(p_k)}{2\pi\kappa\epsilon}$$

La droite sur laquelle les pôles p_k se tiennent, prise de gauche à droite, sera appelé P.

Soient J et J' deux courbes allant du troisième quadrant au premier, et du premier au troisième respectivement, J étant entièrement sur la droite de P et J' entièrement sur sa gauche et sur la droite de l'origine. Supposons également qu'il y a un nombre réel positif a tel que, à une distance suffisamment grande de l'origine, les courbes se tiennent dans la région dans laquelle soit $a < \arg z < \frac{1}{2}\pi - a$, soit $\frac{1}{2}\pi - a > \arg z > -\pi + a$, et que la longueur de la courbe avec $|z| < R$ est $O(R)$. Alors on voit aisément que

$$\int_{J+J'} g(z)dz = 2\pi i \text{ (somme des résidus de } g \text{ entre } J \text{ et } J')$$

$$=\sum \frac{r^{-s}}{1-\exp[-2\pi\kappa\epsilon(r-\mu)]}+\sum_k \frac{\epsilon}{\kappa}h(p_k),$$

où la première somme est calculée sur les entiers r se trouvant entre J et J'. Maintenant

$$\begin{aligned}\int_J g(z)dz &= \int_J \big(g(z)-h(z)\big)dz+\int_J h(z)dz\\ &= \int_J h(z)\frac{\exp[-2\pi\kappa\epsilon(z-\mu)]}{1-\exp[-2\pi\kappa\epsilon(z-\mu)]}dz+\int_P h(z)dz,\end{aligned}$$

et

$$\int_{J'} g(z)dz=-\int_{J'} h(z)\frac{\exp[2\pi\kappa\epsilon(z-\mu)]}{1-\exp[2\pi\kappa\epsilon(z-\mu)]}dz$$

Si les courbes J et J' sont toujours distantes de plus de $\frac{1}{4}\kappa^{-1}$ de la ligne P, alors on a pour eux

$$\left|\frac{1}{1-\exp[2\pi\kappa\epsilon\ \mathrm{sg}(z)(z-\mu)]}\right|<1.27,$$

où $\mathrm{sg}(z)$ prend la valeur 1 ou -1 selon que z est à gauche ou à droite de P. Nous pouvons maintenant rassembler nos résultats sous la forme

$$I(s)=\sum_k \frac{\epsilon}{\kappa}h(p_k)-\sum_{r=1}^{\infty} r^{-s}\theta_r+R_0,$$

où θ_r prend la valeur 1 si r est sur la gauche de J', la valeur

$$\{1-exp[2\pi\kappa\epsilon(r-\mu)]\}^{-1}$$

si r est entre J et J', et la valeur 0 sinon. Le résidu R_0 satisfait

$$|R_0|<\int_{J+J'} 1.27\exp[-\sqrt{2}\pi\kappa\mu\ \mathrm{sg}(z)]\frac{|z|^{-\sigma}}{e^{i\pi z}-e^{-i\pi z}}e^{\Re\phi(z)}|dz|, \tag{3.2}$$

où

$$\phi(z)=i\pi z^2+2\pi\kappa\epsilon z\ \mathrm{sg}(z)-it\log z. \tag{3.3}$$

Nous pouvons aussi écrire la formule pour $I(s)$ sous la forme

$$I(s)=\sum_{k=-\infty}^{\infty}\frac{\epsilon}{\kappa}\frac{e^{i\pi p_k^2}p_k^{-s}}{e^{i\pi p_k}-e^{-i\pi p_k}}-\sum_{r=1}^{\infty}\frac{r^{-s}}{1-\exp[2\pi\kappa\epsilon(r-\mu)]}+R, \tag{3.4}$$

où $R=R_0+R_1$, $p_k=\mu+\epsilon k/\kappa$ et

$$|R_1|<1.27\sum r^{-s}\exp[-\sqrt{2}\pi\kappa|r-\mu|],$$

la sommation s'effectuant sur les entiers positifs non entre J et J'. En exprimant $I(s)$ sous cette forme, nous pouvons éliminer des calculs numériques toute référence à la position des courbes J, J', et le résidu n'est pas augmenté de façon appréciable. Dans le § 5, nous choisissons les courbes de telle façon que $R_1=0$. Bien sûr, dans le calcul, le facteur $\{1-\exp[2\pi\kappa\epsilon(r-\mu)]\}^{-1}$ sera rendu égal à 1 ou à 0 excepté pour un nombre relativement petit de termes.

En estimant le reste, on suppose que $\sigma \geqslant 0$, mais cela n'est pas nécessaire.

4. Remarques générales sur l'estimation du reste.

Supposons que

$$U = \int_{C_0} e^{w(z)} k(z) dz.$$

Alors, pour toute courbe C déformable en C_0 dans le domaine de régularité de $e^{w(z)}k(z)$,

$$|U| \leqslant \int_C e^{\Re w(z)} |k(z)| dz. \tag{4.1}$$

Maintenant supposons que $\Re w(z)$ a de grandes et rapides variations, alors que $|k(z)|$ est comparativement stable. Alors la valeur de l'intégrale (4.1) est principalement affectée par la valeur maximum de $\Re w(z)$ sur C, et une bonne inégalité pour $|U|$ est obtenue en minimisant ce maximum. On voit aisément que s'il y a une courbe pour laquelle le maximum est minimisé, et si z_0 est le point auquel le maximum de la courbe est atteint, alors $w'(z_0) = 0$, i.e. z_0 est un "point de selle" de w. Supposons que dans le voisinage de ce point de selle la courbe est $z = z_0 + le^{ia}$, l étant un paramètre de longueur d'arc. Alors la contribution à l'intégrale du voisinage du point de selle est approximativement

$$e^{\Re w(z_0)} \left| k(z_0) \right| \int_{-\infty}^{\infty} \left| \exp\left[\frac{1}{2} w''(z_0) e^{2ia} l^2 \right] \right| dl = \frac{(2\pi)^{\frac{1}{2}} |k(z_0)| e^{\Re w(z_0)}}{\surd\{-\Re[w''(z_0)e^{2ia}]\}}.$$

Nous choisissons naturellement a comme valant $\frac{1}{2}\pi - \frac{1}{2} \arg w''(z_0)$, et alors l'expression devient

$$\surd\left\{ \frac{2\pi}{w''(z_0)} \right\} \left| k(z_0) e^{w(z_0)} \right|.$$

Dans l'estimation de R_0, nous pourrions prendre $w(z)$ comme étant soit égal à $\phi(z)$ soit à $\phi(z) \pm i\pi z$, dans ce dernier cas des signes différents étant pris sur les deux courbes J, J'. Avec la première de ces formes, l'analyse est plus simple, mais la seconde fournit un meilleur résultat ; nous ne traitons que la forme la plus simple.

Nous utilisons l'idée de l'intégration par point de selle dans la forme suivante. Supposons que nous ayons une courbe C avec un paramètre d'arc de longueur l commençant en $l = 0$. Alors

(a) si $\Re \psi'(z) \dfrac{dz}{dl} \leqslant -al$ sur la courbe, où $a > 0$, alors

$$e^{\Re\psi[z(l)]} \leqslant e^{-\frac{1}{2}al^2} e^{\Re\psi[z(0)]}$$

sur la courbe, et par conséquent

$$\int_C e^{\Re\psi[(z(l)]} |dz| \leqslant \surd\left(\frac{\pi}{2a}\right) e^{\Re\psi[z(0)]}.$$

Ceci nous permet d'estimer la contribution à l'intégrale du voisinage du point de selle ; nous pouvons estimer la contribution du reste de la courbe par une ou plusieurs applications de

(b) si $\Re \psi'(z) \dfrac{dz}{dl} \leqslant -a < 0$ sur la courbe, alors $e^{\Re\psi[z(l)]} \leqslant e^{-al} e^{\Re\psi[z(0)]}$ et

$$\int_C e^{\Re\psi[z(l)]} |dz| \leqslant \frac{1}{a} e^{\Re\psi[z(0)]}.$$

5. Estimation détaillée du reste.

Pour les deux courbes J, J', nous avons deux points de selle différents z_0, z_0', ceux-ci étant des zéros de $\phi'(z)$ des deux côtés de la ligne P. Nous pouvons poser

$$\phi'(z) = 2\pi i z + 2\pi\kappa\epsilon - it/z = \frac{2\pi i}{z}(z - z_0)(z - z_1)$$

sur la gauche de P, et

$$\phi'(z) = 2\pi i z + 2\pi\kappa\epsilon - it/z = \frac{2\pi i}{z}(z - z_0')(z - z_1')$$

sur la droite de P ; les points z_0, z_0' sont dans le demi-plan droit et z_1, z_1' dans le gauche. Si l'on pose de plus $\tau = t/2\pi$, $\rho = \kappa\tau^{-\frac{1}{2}}$, $\zeta = z\tau^{-\frac{1}{2}}$, on peut écrire ces équations ainsi

$$(\zeta - \zeta_0)(\zeta - \zeta_1) = \zeta^2 + \bar{\epsilon}\rho\zeta - 1, \qquad (\zeta - \zeta_0')(\zeta - \zeta_1') = \zeta^2 - \bar{\epsilon}\rho\zeta - 1.$$

Les racines satisfont

$$\zeta_0 + \zeta_1' = 0, \qquad \zeta_1 + \zeta_0' = 0, \qquad \arg\zeta_0 + \arg\zeta_0' = 0, \qquad \frac{1}{4}\pi > \arg\zeta_0' > 0.$$

Il y a une courbe cubique indépendante de ρ sur laquelle se trouvent quatre racines. Nous aurons besoin d'un certain nombre d'autres propriétés des racines, et nous les mentionnerons lorsque nous le nécessiterons ; ce sont principalement des inégalités qui peuvent être prouvées par des méthodes évidentes mais fastidieuses.

Le comportement d'un certain nombre de fonctions de ρ est montré dans la Fig. 1.

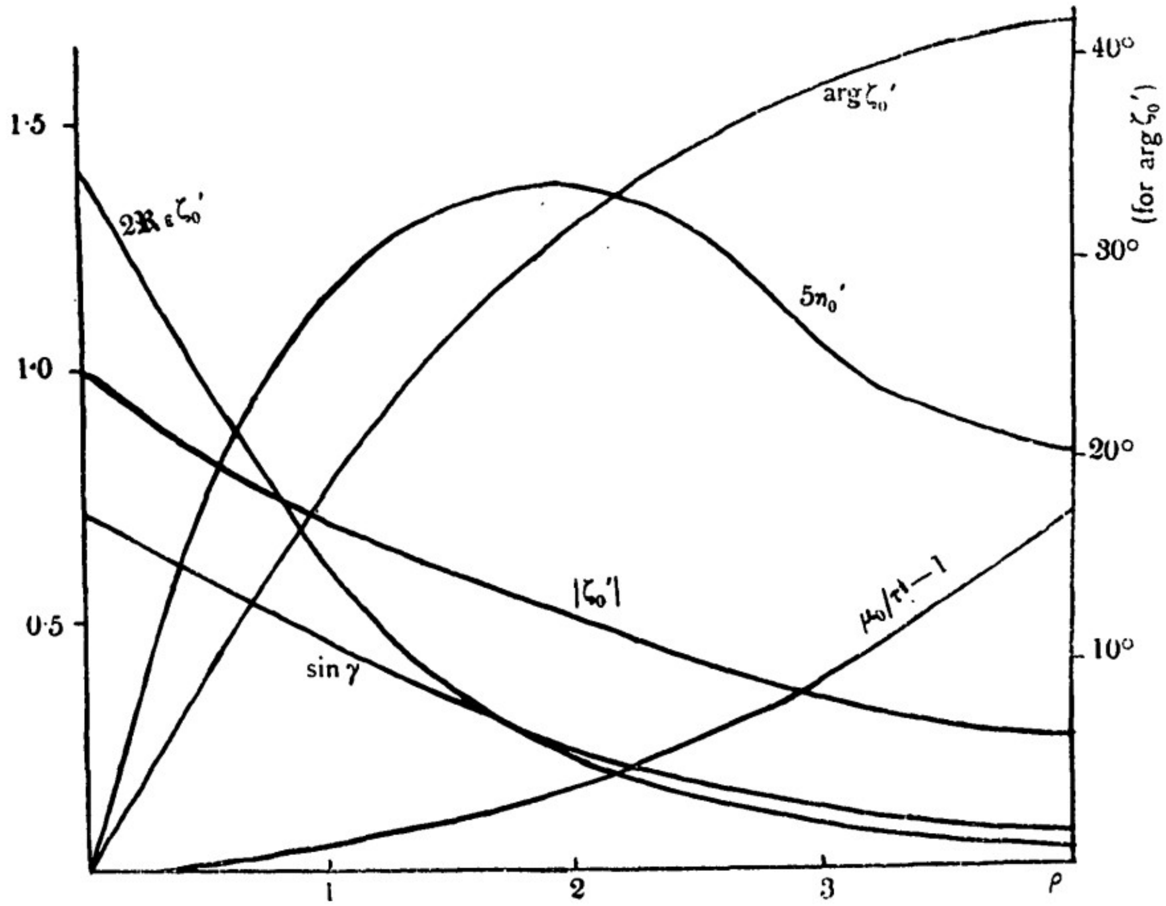

FIG. 1

Nous choisissons les courbes d'intégration J, J' comme suit. J est constitué de trois parties droites J_1, J_2, J_3, avec J_1 une ligne droite de $-\epsilon\infty$ via z_0 vers b, où $b = z_0 - \frac{1}{2}y_o(1+i)$; J_2 relie b et $b+\beta$, où β est réel et positif ; et J_3 relie $b+\beta$ et $\epsilon\infty$, en passant à travers un nombre moitié d'un entier impair. Lorsque β tend vers l'infini, la contribution au reste de J_3 tend vers 0 et la contribution de J_2 tend vers une limite. Nous pouvons de ce fait omettre J_3 et supposer que J est constitué seulement de J_1 et J_2, pris comme se prolongeant à l'infini ; il n'y a alors aucun pôle sur la droite de J qui puisse contribuer à R_1.

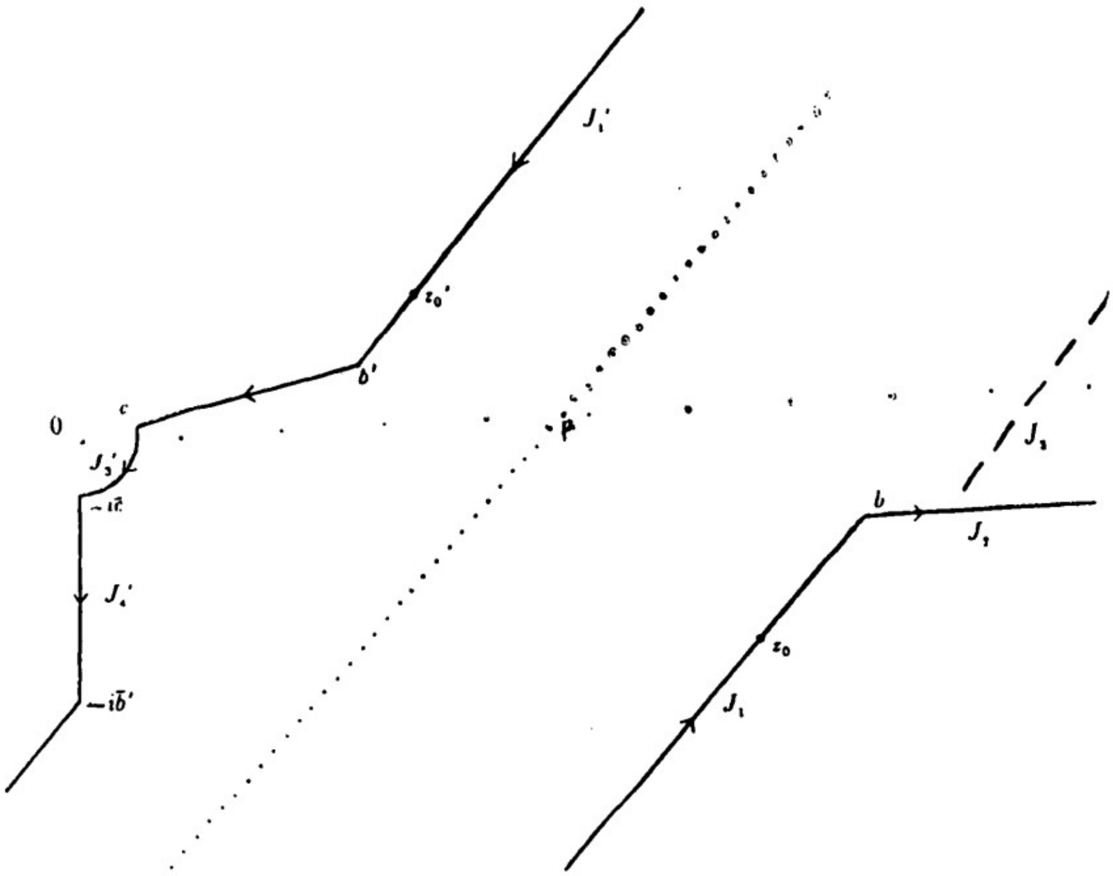

FIG. 2

La courbe J' est constituée des quatre parties J_1', J_2', J_3', J_4', et on divise à nouveau J_2' en J_5' et J_6'. Posons $b' = z_0' - \frac{1}{2}y_0'(1+i)$. Alors J_1' est une ligne droite de $\epsilon\infty$ à b' via z_0' ; J_2' est une ligne droite de b' à c, où c/b' est réel, $0 < c/b' < 1$, et $|c| = \min(\frac{1}{2}, \frac{1}{3}|z_0'|)$. La courbe J_3' est une portion de cercle se trouvant dans le demi-plan $\mathfrak{Re} z > 0$ et reliant c à $-i\bar{c}$. Nous obtenons J_4' comme symétrique de J_1' et J_2' selon la droite $\mathfrak{Je} z = 0$ et en inversant la direction. Nous divisons J_2' en J_5' et J_6', avec J_5' qui est une partie sur laquelle $y > \frac{1}{3}y_0'$. Ces deux parties sont de longueur positive, J_5' étant au moins de longueur $\frac{1}{6}|z_0'|$.

Sur J_1 et J_2, nous avons $|z| > \tau^{\frac{1}{2}}\mathfrak{Re}\zeta$ et

$$\left|\frac{1}{e^{i\pi z} - e^{-i\pi z}}\right| \leqslant \frac{1}{2} \text{ cosech} \left|\frac{1}{2}\pi y_0\right|.$$

La contribution à R_0 provenant de J est ainsi au plus

$$\frac{1}{2}(1.27)\tau^{-\frac{1}{2}\sigma}(\mathfrak{Re}\zeta_0)^{-\sigma}\exp[\sqrt{2\pi\kappa\mu}] \text{ cosech } |\frac{1}{2}\pi y_0| \int_J e^{\mathfrak{Re}\phi(z)}|dz|. \tag{5.1}$$

Sur J_1, on a $0 > \arg\,(z - z_1)/z > -\cos^{-1}\frac{1}{3}$, une simple conséquence du fait que la distance à J_1 à partir de 0 est plus grande que $2^{-\frac{1}{2}}|z_1|$ et que $\arg\, z_1 < -\frac{3}{4}\pi$. Aussi

$$|2\pi(z - z_1)(z - z_0)/z| > \pi|z - z_0|,$$

de telle façon que

$$\Re\frac{2\pi i(z - z_1)(z - z_0)}{z}\frac{dz}{dl} < -\frac{1}{3}\pi l = -\frac{1}{3}\pi|z - z_0|,$$

si l décroît toujours lorsqu'on approche de z_0. Alors, en appliquant (a) du § 4, on a

$$\int_{J_1} e^{\Re\phi(z)}|dz| < \sqrt{6}e^{\Re\phi(z_0)}. \qquad (5.2)$$

Également

$$\Re\phi(b) < \Re\phi(z_0) - \frac{1}{6}\pi y_0^2. \qquad (5.3)$$

Sur J_2, $dz/dl = 1$ et

$$\Re\phi'(z) = \Re\left(2\pi iz - 2\pi\kappa\epsilon - \frac{2\pi i\tau}{z}\right) = -\pi y_0 - \sqrt{2\pi\kappa} - \frac{\pi\tau y_0}{x^2 + \frac{1}{4}y_0^2}.$$

Mais $\tau < x_0^2$, puisque $\xi_0 > 1$ pour tout ρ, et par conséquent

$$\Re\phi'(z)\frac{dz}{dl} < \pi(-\sqrt{2\kappa} - 2y_0) = -2\pi y_0'. \qquad (5.4)$$

Il s'ensuit que, par (b) du § 4, (5.3) et (5.4),

$$\int_{J_2} e^{\Re\phi(z)}|dz| < \frac{e^{\Re\phi(z)}}{2\pi y_0'} < \frac{\exp\left[\Re\phi(z_0) - \frac{1}{6}\pi y_0^2\right]}{2\pi y_0'}. \qquad (5.5)$$

Maintenant étudions J'. Si z est un point de $J_1' + J_2'$, alors $-i\bar{z}$ est le point correspondant de J_4', et

$$\begin{aligned}\Re\phi(-i\bar{z}) &= \Re[i\pi(-i\bar{z})^2 + 2\kappa\epsilon(-i\bar{z}) - it\log z - it\log\,(-i\bar{z}/z)]\\ &\leqslant \Re[-i\pi(iz)^2 + 2\kappa\bar{\epsilon}iz - it\log z - \tfrac{1}{2}t\pi],\end{aligned}$$

i.e.

$$\Re\phi(-i\bar{z}) \leqslant \Re\phi(z) - \frac{1}{2}t\pi.$$

Aussi $|-i\bar{z}| = |z|$ et

$$|\operatorname{cosec}\,(-i\pi z)| < |\operatorname{cosec}\,\pi z|,$$

comme cela peut être démontré en utilisant la représentation par produit de la fonction sinus, en se rappelant que $|\arg z| < \frac{1}{4}\pi$. Par conséquent,

$$\int_{J'} \frac{|z|^{-\sigma}}{|e^{i\pi z} - e^{-i\pi z}|}e^{\Re\phi(z)}|dz| < (1 + e^{\frac{1}{2}i\pi})\int_{J_1'+J_2'+J_3'} \frac{|z|^{-\sigma}}{|e^{i\pi z} - e^{-i\pi z}|}e^{\Re\phi(z)}|dz|. \qquad (5.6)$$

Sur J_1', on a $0 > \arg\,(z - z_1')/z > \gamma - \frac{1}{2}\pi$, où $\gamma = \frac{1}{4}\pi - \arg\zeta_0'$, et

$$|2\pi(z - z_1')(z - z_0')/z| > 2\pi|z - z_0'|.$$

Par conséquent

$$\Re\phi'(z)\frac{dz}{dl} < -2\pi\sin\gamma$$

sur J_1', si l décroît lorsqu'on approche de z_0' de n'importe quel côté. En appliquant (a) du § 4, on a

$$\int_{J_1'} e^{\Re\phi(z)}|dz| < \frac{e^{\Re\phi(z_0')}}{(\sin\gamma)^{\frac{1}{2}}} \tag{5.7}$$

et

$$\Re\phi(b) < \Re\phi(z_0') - \frac{1}{2}\pi y_0'^2\sin\gamma. \tag{5.8}$$

Sur J_2'

$$\begin{aligned}\Re\phi'(z)\frac{dz}{dl} &= \Re\frac{-2\pi iz}{|z|}\left(z+\bar\epsilon\kappa-\frac{\tau}{z}\right)\\ &\leqslant \Re\frac{-2\pi ib'}{|b'|}\left(b'+\bar\epsilon\kappa-\frac{\tau}{b'}\right)\\ &= \mathrm{R}\frac{-2\pi i}{|b'|}(b'-z_0')(b'-z_1').\end{aligned} \tag{5.9}$$

Mais $\arg(z_0'-b') = \frac{1}{4}\pi$ et

$$0 \ > \ \arg(b'-z_1') \ > \ \arg(-z_1') = \gamma - \frac{1}{4}\pi,$$

de telle façon que

$$\frac{1}{2}\pi+\gamma \ < \ \arg\frac{-2\pi i}{|b'|}(b'-z_0')(b'-z_1') \ < \ \frac{3}{4}\pi.$$

Également

$$\Big|-2\pi i(b'-z_0')(b'-z_1')/|b'|\Big| > \sqrt{2}\pi y_0',$$

de telle façon que

$$\Re\frac{-2\pi i}{|b'|}(b'-z_0')(b'-z_1') < -\sqrt{2}\pi y_0'\sin\gamma, \tag{5.10}$$

et en appliquant (b) du § 4, et en utilisant (5.9), (5.10) et (5.8), on obtient

$$\int_{J_3'} e^{\Re\phi(z)}|dz| < (\sqrt{2}\pi y_0'\sin\gamma)^{-1}\exp\,[\Re\phi(z_0') - \frac{1}{2}\pi y_0'^2\sin\gamma]. \tag{5.11}$$

Aussi

$$\Re\phi(\frac{2}{3}b') < \Re\phi(z_0') - \frac{1}{2}\pi y_0'^2\sin\gamma - \frac{1}{6}\sqrt{2}\pi y_0'|z_0'|\sin\gamma. \tag{5.12}$$

Sur J_3'

$$\Big|\arg\left\{\frac{-2\pi i}{z}(z-z_0')(z-z_1')\frac{dz}{dl}\right\}\Big| < 2\sin^{-1}\frac{1}{3} < \frac{1}{4}\pi < \frac{1}{2}\pi - \gamma\ ;$$

de telle façon qu'à la fois sur J_6' et J_3'

$$\Re \frac{2\pi i}{z}(z - z_0')(z - z_1')\frac{dz}{dl} < -\sqrt{2\pi} y_0' \sin\gamma.$$

Cela donne, en utilisant également (5.12),

$$\int_{J_2'+J_3'} e^{\Re\phi(z)}|dz| < (\sqrt{2\pi} y_0' \sin\gamma)^{-1} \exp\,[\Re\phi(z_0') - \frac{1}{2}\pi y_0'^2 \sin\gamma - \frac{1}{6}\sqrt{2\pi} y_0' |z_0'| \sin\gamma]. \quad (5.13)$$

Sur J_1', on a

$$\left|\frac{z^{-\sigma}}{e^{i\pi z} - e^{-i\pi z}}\right| < \frac{1}{2}\tau^{-\frac{1}{2}\sigma}\left|\frac{1}{2}\zeta_0'\right|^{-\sigma} \text{cosech}\left|\frac{1}{2}\pi y_0'\right| \ ; \quad (5.15)$$

sur J_5'

$$\left|\frac{z^{-\sigma}}{e^{i\pi z} - e^{-i\pi z}}\right| < \frac{1}{3}\tau^{-\frac{1}{2}\sigma}\left|\frac{1}{3}\zeta_0'\right|^{-\sigma} \text{cosech}\left|\frac{1}{2}\pi y_0'\right| \ ; \quad (5.16)$$

et sur J_6', J_3'

$$\left|\frac{z^{-\sigma}}{e^{i\pi z} - e^{-i\pi z}}\right| < \frac{1}{2\pi}\left[\min\left(\frac{1}{2}, \frac{1}{3}\tau^{\frac{1}{2}}|\zeta_0'|\right)\right]^{-\sigma-1} \text{cosech arg } \zeta_0' \ ; \quad (5.17)$$

Nous pouvons maintenant rassembler nos résultats pour donner une inégalité pour $|R|$. On utilise (5.1), (5.2), (6.6), (5.6), (5.7), (6.11), (5.18), (5.15), (5.16) et (5.17), et nous rendons les exposants $\Re\phi(z_0) + \sqrt{2\pi}\kappa\mu$ et $\Re\phi(z_0') - \sqrt{2\pi}\kappa\mu$ plus explicites en utilisant la relation $z_0^2 = \bar{\epsilon}\kappa z_0 + \tau$:

$$\begin{aligned}
|R| = |R_0| &< 0.635\ (\Re\,\epsilon\zeta_0)^{-\sigma}\tau^{-\frac{1}{2}\sigma}\ \text{cosech}\ |\tfrac{1}{2}\pi y_0| \left\{2.45 + \frac{0.160}{y_0'}e^{-\frac{1}{2}\pi y_0^2}\right\} \\
&\times \exp\,[-\pi\kappa\Re\,\epsilon(z_0 - 2\mu) + 2\pi\tau\ \text{arg}\ \zeta_0] + 0.635(1 + e^{-\frac{1}{2}\pi i})\tau^{-\frac{1}{2}\sigma}|\zeta_0'|^{-\sigma} \\
&\times \Big\{(\sin\gamma)^{-\frac{1}{2}}2^{\sigma}\ \text{cosech}|\tfrac{1}{2}\pi y_0'| + (\sqrt{2\pi} y_0' \sin\gamma)^{-1}\ \exp[-\tfrac{1}{2}\pi y_0'^2 \sin\gamma] \\
&\times \left[3^{\sigma}\ \text{cosech}|\tfrac{1}{3}\pi \text{y}_0'| + \frac{1}{\pi}|\zeta_0'|^{\sigma}\tau^{\frac{1}{2}\sigma}\left[\min\left(\frac{1}{2}, \frac{1}{3}\tau^{\frac{1}{2}}|\zeta_0'|\right)\right]^{-\sigma-1}\ \text{cosec arg}\zeta_0' \right. \\
&\left. \times \exp[-\tfrac{1}{6}\sqrt{2\pi} y_0'|z_0'| \sin\gamma]\right]\Big\}\ \exp\,[\,-\pi\kappa\Re\,\epsilon(2\mu - z_0') + 2\pi\tau\text{arg}\ \zeta_0'\,]. \quad (5.18)
\end{aligned}$$

À ce point, il peut être aussi bien de répéter les définitions des quantités variables apparaissant sur le côté droit de cette inégalité, en fonction de s, κ :

$$\sigma = \Re s, \qquad t = \Im s, \qquad \tau = t/2\pi, \qquad \rho = \kappa\tau^{-\frac{1}{2}}, \qquad \epsilon = e^{\frac{1}{2}i\pi}.$$

Les nombres complexes $\zeta_0', \zeta_1', \zeta_1, \zeta_0$ sont les racines de l'équation

$$(\zeta^2 - 1)^2 = -i\rho^2\zeta^2$$

se trouvant respectivement dans les premier, second, troisième et quatrième quadrants, et

$$\zeta = \xi + i\eta, \qquad z = x + iy = \tau^{\frac{1}{2}}\zeta, \qquad \gamma = \frac{1}{4}\pi - \arg\, \zeta_0',$$

où ξ, η, x, y sont réels.

L'estimation (5.18) pour R est assez compliquée. Elle peut être considérablement simplifiée quand d n'est pas trop grand. Je fournis une estimation dans le cas $\rho \leqslant \frac{1}{2}$; on a alors

$$\begin{cases} \sqrt{2}\mathfrak{R}\,\epsilon\zeta_0' > 1, & \mathfrak{R}\,\epsilon(1-\zeta_0') > 0.45\rho, \\ -\eta_0 > \rho/2\sqrt{2}, & \sin\gamma > 0.55, \\ \eta_0' > 0.29\rho, & |\zeta_0'| > 0.81. \end{cases} \tag{5.19}$$

Le résultat est que, pour $\rho \leqslant \frac{1}{2}, t \geqslant 25$,

$$\begin{aligned} |R| = |R_0| < r^{-\frac{1}{2}\sigma}\Big[&0.76\ .\ 2^{\frac{1}{4}-\frac{1}{2}\sigma} \operatorname{cosech} 0.78(\kappa/\sqrt{2}) \\ &\times \{2.45 + 0.40(\kappa/\sqrt{2}^{-1} \exp\,(-0.13(\kappa/\sqrt{2})^2)\}e^{-A} + 0.71(0.81)^{\frac{1}{2}-\sigma} \\ &\times \Big[1.91.2^{\sigma-\frac{1}{2}} \operatorname{cosech} 0.64(\kappa/\sqrt{2}) + 1.00(\kappa/\sqrt{2})^{-1} \exp\,(-0.14(\kappa/\sqrt{2})^2) \\ &\times \Big\{1.74\ .\ 3^{\sigma-\frac{1}{2}} \operatorname{cosech} 0.42(\kappa/\sqrt{2}) + 1.7(1.62)^{\sigma-\frac{1}{2}}\tau^{\frac{1}{2}\sigma+\frac{1}{2}} \\ &\qquad \times (\kappa/\sqrt{2}) \exp\,\Big(-0.13\tau^{\frac{1}{2}}(\kappa/\sqrt{2})\Big)\Big\}\Big]e^{-B}\Big], \end{aligned} \tag{5.20}$$

où

$$A = \pi\kappa\mathfrak{R}\,\epsilon(z_0 - 2\mu) + 2\pi\tau \arg\, \zeta_0', \qquad B = \pi\kappa\mathfrak{R}\,\epsilon(2\mu - z_0') - 2\pi\tau \arg\, \zeta_0'.$$

Dans le cas $\rho \geqslant \frac{1}{2}$, l'estimation du reste à partir de (5.18) peut être rendue plus facile en utilisant les inégalités suivantes, qui sont valides pour tout ρ positif :

$$\mathfrak{R}\,\epsilon\zeta_0' > \frac{1}{\rho^3+\rho^2+\rho+\sqrt{2}}, \qquad \eta_0' > \frac{0.65\rho}{\rho^2+2}, \qquad \sin\gamma > \frac{1}{\rho^2+\frac{1}{2}\rho+\sqrt{2}}. \tag{5.21}$$

6. Choix des paramètres. Le reste avec une série finie.

Le reste R comme donné par (5. 18) est la somme de deux termes dans lesquels les principaux facteurs sont e^{-A} et e^{-B}. Le choix le plus favorable du paramètre μ est probablement approximativement ce qui rendra ces deux facteurs égaux. En appelant cette valeur μ_0, nous avons

$$\begin{aligned} \mu_0 &= \frac{1}{2\sqrt{2}}\left[\mathfrak{R}\,\epsilon(z_0' + z_0) + \frac{4\tau^{\frac{1}{2}}}{\rho} \arg\, \zeta_0'\right] \\ &= \frac{\tau^{\frac{1}{2}}}{\sqrt{2}}\left[\mathfrak{R}\sqrt{\left(i + \frac{1}{4}\rho^2\right)} + \frac{2}{\rho} \arg\, \zeta_0'\right] \end{aligned}$$

Lorsque ρ tend vers l'infini, $\mu_0 \sim \kappa/2\sqrt{2}$; et, lorsque ρ tend vers 0, $\mu_0 \sim \tau^{\frac{1}{2}}$. Aussi

$$\begin{aligned}\Re\, \epsilon(\mu_0 - z_0') \quad &= \frac{1}{4}\left[\Re\, \epsilon(z_0 - 3z_0') + \frac{4\tau^{\frac{1}{2}}}{\rho}\ \arg \zeta_0'\right] \\ &= \frac{1}{2}\kappa\left[1 - \Re\sqrt{\left(\frac{1}{4} + \frac{i}{\rho^2}\right)} + \frac{2}{\rho^2}\ \arg \zeta_0'\right].\end{aligned}$$

Lorsque ρ tend vers 0, le facteur

$$1 - \Re\sqrt{\left(\frac{1}{4} + \frac{i}{\rho^2}\right)} + \frac{2}{\rho^2}\ \arg \zeta_0'$$

tend vers 1 ; et lorsque ρ tend vers l'infini, ce facteur tend vers $\frac{1}{2}$. Pour toutes les valeurs positives de ρ, il est plus grand que $\frac{1}{2}$, et par conséquent

$$\Re\, \epsilon(\mu_0 - z_0') > \frac{1}{4}\kappa.$$

Aussi

$$\begin{aligned}\Re\, \epsilon(z_0 - \mu_0) \quad &= \frac{1}{4}\left[\Re\, \epsilon(3z_0 - z_0') + \frac{4\tau^{\frac{1}{2}}}{\rho}\ \arg \zeta_0'\right] \\ &= \frac{1}{2}\kappa\left[1 - \Re\sqrt{\left(\frac{1}{4} + \frac{i}{\rho^2}\right)} - \frac{2}{\rho^2}\ \arg \zeta_0'\right].\end{aligned}$$

et on a

$$\Re\, \epsilon(z_0 - \mu_0) > \frac{1}{2}\kappa.$$

Si, pour μ, nous choisissons $\mu_0 \pm \delta$, où $\delta > 0$, alors le plus grand des facteurs exponentiels e^{-A}, e^{-B} est

$$\exp\ \left[-\frac{1}{2}\pi\kappa^2 + \sqrt{2}\pi\kappa\delta\right].$$

Les valeurs de μ que nous pouvons choisir sont restreintes seulement par la condition que les courbes J, J' doivent être distantes au moins de la quantité $1/(4\kappa)$ de P. Si $\kappa \geqslant \sqrt{2}$, nous pouvons alors choisir μ dans l'intervalle $(\mu_0, \mu_0 + \frac{1}{2})$; et, si $\kappa \geqslant 2$, nous pouvons le choisir dans l'intervalle $(\mu_0 - \frac{1}{2}, \mu_0 + \frac{1}{2})$. Pourtant, pour de si petites valeurs de κ, ce sera probablement un meilleur choix que de choisir μ plutôt proche de μ_0. Nous n'avons pas à considérer le cas de valeurs de κ plus petites que $\sqrt{2}$, puisque, comme cela apparaîtra, ce n'est pas désavantageux de prendre de telles petites valeurs, même quand il n'y a qu'un terme pris dans la série $\sum h(p_k)$.
Quand ρ est petit, μ_0 est proche de $\tau^{\frac{1}{2}}$, et il est alors probablement plus simple de choisir une valeur de μ qui est proche de $\tau^{\frac{1}{2}}$ sans vraiment calculer μ_0. L'inégalité

$$0 < \mu_0/\tau^{\frac{1}{2}} - 1 < \rho^2/3 \qquad (0 < \rho < 1)$$

devrait alors être de qualité.

Pour des grandes valeurs de κ (e.g. $\kappa > 3$), nous ferions bien de prendre μ soit entier soit demi entier impair. Dans le cas où μ est un entier, la fonction g a un double pôle en μ ; à la place des termes

$$\frac{\epsilon}{\kappa}h(\mu) - \frac{r^{-s}}{1 - \exp\left[2\pi\kappa\epsilon(\mu - \mu)\right]}$$

nous devons alors mettre le reste en 0 de

$$\frac{2\pi i(-)^{\mu}(z+\mu)^{-s}\exp\left[i\pi(z+\mu)^2 - i\pi z\right]}{(1-e^{-2\pi i z})(1-\rho^{-2\pi\kappa\epsilon z})}$$

et cela est égal à

$$\mu^{-s}\left\{-\frac{1}{2} + \frac{\sigma}{2\pi\kappa\epsilon\mu} - \frac{\epsilon}{\kappa}\left(\mu - \frac{t}{2\pi\mu}\right)\right\}.$$

Dans les applications pratiques, bien sûr, nous ne prenons qu'un nombre fini de termes de la série $\sum(\epsilon/\kappa)h(p_k)$. Nous voulons alors une estimation de l'erreur provenant de ce choix. Si nous décidons la plus grande erreur totale η que nous puissions admettre dans notre résultat, nous pouvons procéder de cette manière. On choisit κ de telle façon que $|R| < \frac{1}{2}\eta$ et nous prenons alors suffisamment de termes de la série pour que l'erreur provenant de cette seconde source n'excède pas $\frac{1}{2}\eta$. Maintenant estimons le second reste. Dans ce but, nous prouvons le lemme suivant :

Lemme. *La fonction $|e^{i\pi z^2}z^{-it}|$ a seulement un maximum sur la ligne P.*

Posons $z = \mu\Big(1+\theta(1+i)big)$, $a = t/(2\pi\mu^2)$; alors θ est réel et

$$\log|e^{i\pi z^2}z^{-it}| = \Re\left[i\pi\Big(1+\theta(1+i)\Big)^2 - 2\pi i a\log\Big\{\mu\Big(1+\theta(1+i)\Big)\Big\}\right]\mu^2$$

Abrégeons le côté droit par $H(\theta)$. Alors

$$H'(\theta) = 2\pi\mu^2\Big(-1-2\theta+\frac{a}{(1+\theta)^2+\theta^2}\Big) = -2\pi\mu^2\frac{\Big((1+\theta)^2+\theta^2\Big)(1+2\theta)-a}{(1+\theta)^2+\theta^2}.$$

Maintenant

$$\frac{d}{d\theta}\Big[\Big((1+\theta)^2+\theta^2\Big)(1+2\theta)-a\Big] = 3(2\theta+1)^2+1 > 0,$$

et par conséquent $\Big((1+\theta)^2+\theta^2\Big)(1+2\theta)-a$ ne peut s'évanouir pour plus d'une valeur de θ ; cela s'évanouit clairement pour au moins une valeur. Alors $H(\theta)$ a juste une valeur stationnaire, qui est clairement vue comme étant un maximum. Cela complète la preuve du lemme. Si $a < 1$, la valeur θ fournissant le maximum satisfait $0 > \theta > a^{\frac{1}{2}} - 1$.

Posons $u_k = \kappa^{-1}|h(p_k)|$; alors, si $\sigma \geqslant 0$,

$$\sum_{k=K+1}^{\infty} u_{-k} < \frac{(\mu/\sqrt{2})^{-\sigma}}{1-\exp\left[\sqrt{2}\pi(K+1)/\kappa\right]}\sum_{k=K+1}^{\infty}\Big(-\frac{\pi k}{\sqrt{2}\kappa}\Big)\Big|e^{i\pi p_{-k}^2}p_{-k}^{-it}\Big|.$$

Si nous supposons que $\Im(p_{-K-1})$ est inférieur à la valeur de y pour laquelle le maximum de $|e^{i\pi z^2}z^{-it}|$ est atteint, et si $k \geqslant K+1$, alors

$$|e^{i\pi p_{-k}^2}p_{-k}^{-it}| \leqslant |e^{i\pi p_{-K-1}^2}p_{-K-1}^{-it}|$$

et, par conséquent

$$\begin{aligned}\sum_{K+1}^{\infty} u_{-k} &< \Big(\frac{\mu}{\sqrt{2}}\Big)^{-\sigma} \frac{\exp\left[-\pi(K+1)/\kappa\sqrt{2} + \Re(i\pi p_{-K-1}^2)\right] |p_{-K-1}^{-it}|}{\Big(1-\exp[-\sqrt{2}\pi(K+1)/\kappa]\Big)(1-\exp[-\pi/\kappa\sqrt{2}])} \\ &< \frac{2^{\frac{1}{2}\sigma} u_{-K-1}}{\Big(1-\exp[-\sqrt{2}\pi(K+1)/\kappa]\Big)^2 (1-\exp[-\pi/\kappa\sqrt{2}])}\end{aligned}$$

De façon similaire, si (ce qui est toujours le cas si $K' \geqslant 0$ et $a \leqslant 1$) $\Im(p_{K'+1})$ est plus grand que la valeur de y pour laquelle le maximum advient,

$$\sum_{K'+1}^{\infty} u_k < \frac{u_{K'+1}}{\Big(1-\exp[-\sqrt{2}\pi(K'+1)/\kappa]\Big)^2 (1-\exp[-\pi/\kappa\sqrt{2}])}$$

de telle façon que

$$\begin{aligned}|R^*| &= \Big|\Big(\sum_{-\infty}^{\infty} - \sum_{-K}^{K'}\Big)\frac{\epsilon}{\kappa} h(p_k)\Big| \\ &< \frac{2^{\frac{1}{2}\sigma}|h(p_{-K-1})| + |h(p_{K'+1})|}{\kappa\Big(1-\exp[-\sqrt{2}\pi(K^*+1)/\kappa]\Big)^2 (1-\exp[-\pi/\kappa\sqrt{2}])}\end{aligned}$$

où $K^* = min(K, K')$.

Dans le cas où κ est petit comparé à $\tau^{\frac{1}{2}}$, nous pouvons facilement obtenir une estimation grossière du nombre de termes requis pour une précision donnée. Car dans ce cas K, K' sont tels que $p_{-K}/\mu - 1$ et $p_{K'}/\mu - 1$ sont petits et u_K est approximativement égal à $\exp[-2\pi(K+1)^2/\kappa^2]$. Si les restes R et R^* sont du même ordre de grandeur, alors on a approximativement $2\pi(K+1)^2/\kappa^2 = \dfrac{1}{2}\pi\kappa^2$, i.e., $K+1 = \dfrac{1}{2}\kappa^2$. Le nombre de termes T que nous prenons est $2K+1$, i.e., approximativement κ^2, et l'erreur totale est de l'ordre de grandeur de $e^{-\frac{1}{2}\pi T}$. Si cette assertion doit être mise sous une forme exacte, nous pouvons dire que si μ et κ sont convenablement choisis comme des fonctions de t, η, et σ et sont dans l'intervalle $0 \leqslant \sigma \leqslant 1$, et si η tend vers 0 et si t tend vers l'infini de telle façon que $\kappa^{-1}, \kappa\tau^{-\frac{1}{2}}$ tende aussi vers 0, alors l'erreur n'excèdera pas η et $T^{-1}\log\ \eta^{-1}$ tendra vers $\frac{1}{2}\pi$; cela n'est pas vérifié pour n'importe quel nombre supérieur à $\frac{1}{2}\pi$.

Quand $\kappa\tau^{-\frac{1}{2}}$ est de l'ordre de grandeur de 1, nous ne pouvons obtenir une estimation si simple du nombre de termes nécessité, mais nous pouvons obtenir une estimation dans le cas limite $\kappa\tau^{-\frac{1}{2}} \to \infty$. Nous pouvons alors négliger tous les facteurs en u_h, excepté $\Big|e^{i\pi}p_k^2\Big|$. En posant $\mu + \epsilon v = p_{-K}$, nous avons approximativement

$$\Big|e^{i\pi(\mu+\epsilon v)^2}\Big| = e^{-\frac{1}{2}\pi\kappa^2}$$

si R et R^* sont du même ordre de grandeur ; i.e.,

$$\sqrt{2}\pi\mu v + \pi v^2 = \frac{1}{2}\pi\kappa^2.$$

Mais pour de grandes valeurs de μ, nous avons approximativement $\mu = \kappa/2\sqrt{2}$, et par conséquent approximativement

$$2v^2 + \kappa v - \kappa^2 = 0.$$

Les deux racines de cette équation sont $\bar{\epsilon}(P_{-K}-\mu)$ et $\bar{\epsilon}(P_{K'}-\mu)$ (approximativement) ; la différence entre les racines est $\frac{3}{2}\kappa$ et par conséquent $T=\frac{3}{2}\kappa^2$ approximativement, et l'erreur est de l'ordre de grandeur de $e^{-\frac{1}{2}\pi T}$.

Il est possible que cela soit amélioré en prenant μ différent de μ_0 ; car si nous prenons μ plus proche de $\tau^{\frac{1}{2}}$ le reste R^* est rendu plus petit pour toute valeur donnée de T. Une telle amélioration serait nécessairement obtenue aux dépens du reste R ; je ne pense pas qu'une amélioration appréciable puisse réellement être obtenue dans cette direction.

Dans le cas où $T=3$, nous pouvons poser $\kappa=1.6\sqrt{2}$, et alors, si $\sigma=\frac{1}{2}, \mu=\tau^{\frac{1}{2}}$ et $t>350$, et si le facteur $(1-\exp[2\pi\kappa\epsilon(r-\mu)])^{-1}$ est remplacé par 0 ou 1 sauf pour deux termes de la série principale, l'erreur de toutes les sources n'excède pas $0.0044\tau^{-\frac{1}{2}}$.

7. Une méthode similaire.

Il y a une alternative, et mieux connue, une représentation intégrale de la fonction zeta sur laquelle nous pouvons baser nos calculs, viz.,

$$\zeta(s)=\sum_{r=1}^{m} r^{-s}+2(2\pi)^{s-1}\sin\frac{1}{2}\pi s\Gamma(1-s)\sum_{r=1}^{m'} r^{-s}+\frac{1}{1+e^{i\pi s}}\int_{Q_m}\frac{e^{2m'i\pi z}z^{-s}dz}{1-e^{-2\pi iz}}.$$

Ici Q_m est une courbe venant de l'infini dans le premier quadrant, traversant l'axe réel entre m et $m+1$ et à nouveau entre $-m$ et $-m-1$, et repartant à l'infini dans le second quadrant ; z^{-s} est définie comme dans le § 2. Si nous choisissons $m=m'=[\tau^{\frac{1}{2}}]$, et laissons la partie de Q_m au voisinage de l'axe réel positif être une ligne droite coupant l'axe réel dirigé négativement en μ et à angle de $+\frac{1}{4}\pi$, alors, pour de grandes valeurs de t, la seule contribution appréciable à l'intégrale provient du voisinage de l'axe réel positif. Nous pouvons approximer cette intégrale de la même manière que précédemment, la valeur approximée résultante pour $\zeta(s)$ étant

$$\sum_{r=1}^{\infty} r^{-s}\left\{1-(1+e^{i\pi s})^{-1}(1-\exp[-2\pi\kappa\epsilon(r-\mu)])^{-1}\right\}$$

$$+2(2\pi)^{s-1}\sin\frac{1}{2}\pi s\Gamma(1-s)\sum_{r=1}^{\infty} r^{-s}$$

$$-\frac{\epsilon}{\kappa(1+e^{i\pi s})}\sum_{k=-K}^{K}\frac{(\mu+\epsilon k/\kappa)^{-s}\exp[2\pi im(\mu+\epsilon k/\kappa)]}{1-\exp[-2\pi i(\mu+\epsilon k/\kappa)]}.$$

L'entier K ne doit pas être choisi trop grand ; $K<\frac{1}{2}\tau^{\frac{1}{2}}\kappa$ est habituellement suffisamment petit. Cette méthode a l'avantage que pour les points qui ne sont pas sur la droite critique, seuls deux intégrales réelles doivent être évaluées, et non quatre. Cela peut présenter un intérêt pour le calcul des zéros qui ne sont pas sur la droite critique. Dans ce but, cela n'importe pas que la méthode ne soit applicable que pour de grandes valeurs de t ; il est, pourtant, possible d'enlever cette restriction en intégrant le long d'une parabole, e.g., la parabole

$$x^2=2\mu y+\mu^2([\tau^{\frac{1}{2}}]\leqslant\mu\leqslant[\tau^{\frac{1}{2}}]+1).$$

L'application conforme $u^2=z$ transforme cette parabole en une ligne droite, telle que

$$\int_{Q_m}\frac{e^{2\pi imz}z^{-s}dz}{1-e^{-2\pi iz}}=-2\int_{m^{\frac{1}{2}}\nearrow(m+1)^{\frac{1}{2}}}\frac{e^{2\pi imu^2}u^{-2s+1}du}{1-e^{-2\pi iu^2}}.$$

La ligne d'intégration coupe l'axe imaginaire entre $-im^{\frac{1}{2}}$ et $-i(m+1)^{\frac{1}{2}}$.

King's College,
Cambridge.

Extrait de la biographie *Alan Turing ou l'énigme de l'intelligence* d'Andrew Hodges (DC 30/12/13)

(p. 218) Ce n'était pas le seul parallèle entre les travaux de Claude Shannon et ceux d'Alan Turing. Il existait entre eux comme une sorte de réciprocité. Alan, dont le point fort était plutôt la logique des machines, s'était néanmoins plongé dans l'étude de l'information.

Shannon, de son côté, s'était également intéressé au concept de machine logique. Alan lui fit lire ses *Nombres calculables* et ils parlèrent d'une idée très présente dans l'article de Turing, à savoir la reproduction mécanique du cerveau. De 1936 à 1938, Shannon avait travaillé sur l'analyseur différentiel du Massachussetts Institute of Technology et, ayant étudié la neurologie au même titre que les mathématiques et la logique, il avait vu dans cette recherche un premier pas vers la machine pensante. Ils se rendirent compte qu'ils partageaient une même conception des choses : le cerveau n'avait rien de sacré, et si une machine parvenait un jour à faire aussi bien qu'un cerveau, alors elle serait effectivement douée de la faculté de *penser*. Ni l'un ni l'autre ne proposait cependant de moyen d'y arriver.

C'était là, au moins, un sujet dont ils pouvaient parler librement. Alan s'étonna un jour : "Shannon ne veut pas entrer seulement des *données* dans un cerveau, il veut lui donner de la *culture* ! Il veut lui faire écouter de la *musique* !". Une autre fois, à la cantine, alors qu'il dissertait sur les possibilités d'une "machine pensante", sa voix haut perchée commença à dominer le brouhaha général des jeunes cadres dynamiques en quête de promotion au sein des Bell Labs. Tous l'entendirent bientôt affirmer : "Non, ce qui m'intéresse, ce n'est pas de mettre au point un cerveau puissant. Je ne cherche rien d'autre qu'un cerveau médiocre, dans le genre de celui du président de l'American Telephone and Telegraph Company.". La salle entière fut pétrifiée, mais Alan continua nonchalamment à exposer son idée : fournir à la machine toutes les données concernant les cours de la bourse et les matières premières puis lui poser simplement la question : "J'achète ou je vends ?". Le téléphone sonna ensuite tout l'après-midi dans son laboratoire et l'on ne cessa de lui demander qui diable il pouvait bien être.

Note : Alan Turing avait inventé bien avant l'heure le trading haute-fréquence...

Machines informatiques et intelligence

A. M. Turing

1 Le jeu de l'imitation

Je propose de considérer la question, "Les machines peuvent-elles penser ?". On devrait commencer par définir les termes "machine" et "pensée." Les définitions devraient être choisies de manière à refléter aussi bien que possible l'usage courant de ces mots, mais cette attitude est dangereuse. Si les significations des mots "machine" et "pensée" doivent être utilisées de la manière dont elles le sont habituellement, il est difficile d'échapper à la conclusion que le sens de la question "Les machines peuvent-elles penser ?" et la réponse à cette question doivent être recherchés de façon statistique, comme par sondage. Mais cela est absurde. Plutôt que de tenter une telle définition, je remplacerai la question par une autre, qui lui est intimement reliée et qui s'exprime en termes relativement non-ambigus.

La nouvelle forme du problème peut être décrite en termes d'un jeu que nous appelons le "jeu de l'imitation". Il se joue à trois, un homme (A), une femme (B), et un interrogateur (C) qui peut être de l'un ou l'autre sexe. L'interrogateur reste dans une pièce et n'est pas vu par les deux autres. L'objectif du jeu pour l'interrogateur est de déterminer qui est l'homme et qui est la femme des deux autres. Il les connaît par leur étiquette (X et Y), et à la fin du jeu, il dit soit "X est A et Y est B" soit "X est B et Y est A." L'interrogateur a le droit de poser des questions à A et B comme :

C : X peut-il ou elle me dire la longueur de ses cheveux ?

Maintenant supposons que X est vraiment A, alors A doit répondre. L'objectif de A pour ce jeu est d'essayer de faire que C se trompe dans son identification. Sa réponse pourrait donc être :

"J'ai les cheveux attachés, et ils sont longs de 20 cm."

Pour que les hauteurs des voix ne puissent pas aider l'interrogateur, les réponses seront écrites ou mieux, tapées à la machine. Le meilleur dispositif consiste à avoir un téléscripteur de communication entre les deux pièces. Sinon, les questions et les réponses

http://www.turingarchive.org/browse.php/B/9

peuvent être répétées par un intermédiaire. L'objectif du jeu pour le troisième joueur (B) est d'aider l'interrogateur. La meilleure stratégie pour cette personne est probablement de donner les vraies réponses. Elle peut ajouter des choses comme "Je suis la femme, ne l'écoutez pas !" à ses réponses, mais ça ne servira à rien parce que l'homme pourra faire des remarques similaires.

Maintenant posons la question, "qu'arrivera-t-il si une machine prend la place de A dans ce jeu ?". L'interrogateur(-machine) se trompera-t-il aussi souvent que lorsque le jeu est joué par des hommes et des femmes ? Ces questions remplacent notre question originale, "Les machines peuvent-elles penser ?"

2 Critique du nouveau problème

De la même façon qu'on peut se demander "Quelle est la réponse à la question sous sa nouvelle forme", on peut aussi se demander "Cette nouvelle question est-elle digne de réflexion ?". On réfléchira à cette dernière question sans tergiverser davantage, coupant là une régression infinie.

Le nouveau problème présente l'avantage de dessiner une frontière assez nette entre les capacités physiques et intellectuelles d'un homme. Aucun ingénieur ou chimiste ne se targue de pouvoir produire un matériau qui ne soit pas distinguable de la peau humaine. Il est possible qu'un jour cela soit réalisé, mais même en supposant que cette invention ait un jour été faite, nous pouvons sentir combien il y a peu en commun entre le fait d'essayer de rendre une "machine pensante" plus humaine en la recouvrant de cette peau artificielle. La forme dans laquelle nous avons spécifié le problème montre que les conditions empêchent l'interrogateur de voir ou toucher les autres personnages, ou d'entendre leurs voix. D'autres avantages du critère proposé peuvent se voir dans des questions et réponses specimen. Par exemple :

Q : S'il vous plaît, écrivez-moi un sonnet avec comme sujet le quatrième pont.

A : Ne comptez pas sur moi. Je ne pourrai jamais écrire de poèmes.

Q : Ajoutez 34957 à 70764.

A : (Pause d'environ 30 secondes et réponse donnée ensuite) 105721.

Q : Jouez-vous aux échecs ?

A : Oui.

Q : J'ai K en K1, et pas d'autres pièces. Vous avez K en K6 et R en R1. C'est votre tour. Que jouez-vous ?

A : (Après une pause de 15 secondes) R-R8 mat.

La méthode des questions et réponses semble adaptée pour introduire presque tous les champs d'application que nous souhaiterions inclure. Nous ne voulons pas pénaliser la machine pour son incapacité à briller dans des concours de beauté, ni pénaliser un humain parce qu'il perd une course contre un avion. Les conditions de notre jeu rendent ces incompétences non pertinentes. Les "témoins" peuvent se vanter, s'il pensent que c'est judicieux, autant qu'ils le souhaitent sur leurs charmes, leur force, ou leur héroïsme, mais l'interrogateur ne peut pas leur demander de démonstrations pratiques.

Le jeu peut peut-être être critiqué sur la base que les chances sont trop désavantageuses contre la machine. Si l'homme voulait faire semblant d'être une machine, il se montrerait vraisemblablement très médiocre. Il serait mis en échec du premier coup par sa lenteur et ses erreurs en arithmétique. Les machines ne sont-elles pas capables de faire quelque-chose qu'on a l'habitude de nommer *penser* mais qui est très différent de ce qu'un humain fait ? Cette objection est une objection très forte, mais au moins, nous pouvons dire que si, néanmoins, une machine peut être construite pour jouer de façon satisfaisante au jeu de l'imitation, il ne faudrait pas être troublé par cette objection.

On doit souligner qu'en jouant au "jeu de l'imitation", la meilleure stratégie pour la machine peut possiblement être quelque-chose de différent de l'imitation du comportement humain. C'est possible, mais je pense qu'il est peu probable que cela ait un grand effet. Dans tous les cas, on n'a pas d'intention ici de faire des recherches en théorie des jeux, et on supposera que la meilleure stratégie est d'essayer de fournir des réponses qui seraient naturellement données par un humain.

3 Machines concernées par le jeu

La question que nous avons posée en 1 ne sera pas bien définie tant que nous n'aurons pas précisé ce que signifie le mot "machine." Il est naturel de vouloir autoriser toutes sortes de techniques d'ingénierie dans nos machines. Nous pouvons aussi souhaiter la possibilité que des ingénieurs puissent construire une machine qui fonctionne, mais qui ne peut être considérée comme satisfaisante car ses constructeurs ont utilisé une méthode qui est très expérimentale. Finalement, nous souhaitons exclure des machines les humains nés de la façon (biologique) habituelle. Il est difficile de cadrer les définitions de manière à satisfaire les trois conditions. On peut par exemple insister pour que l'équipe d'ingénieurs soient tous du même sexe, mais cela ne serait pas vraiment satisfaisant, car il est possible de construire un individu complet à partir d'une seule cellule de peau (par exemple) d'un humain. Réaliser cela serait un exploit de technique biologique méritant les plus grandes louanges, mais nous ne saurions considérer cela comme un cas de "construction d'une machine pensante". Cela nous enjoint à abandonner la nécessité que tout type de technique puisse être permis. Nous sommes d'autant plus prêts à cela étant donné que l'intérêt actuel pour les "machines pensantes" a été motivé par une sorte particulière de machine, habituellement appelées "ordinateurs électroniques" ou "ordinateurs digitaux". Selon cette suggestion, nous n'autorisons que les ordinateurs digitaux à prendre part à notre jeu.

Cette restriction apparaît au premier abord comme étant très drastique. Je vais essayer de montrer qu'elle ne l'est pas en réalité. Faire cela nécessite une légère prise en compte de la nature et des propriétés de ces ordinateurs.

On peut aussi dire que cette identification des machines aux ordinateurs digitaux, selon notre critère qualifiant la "pensée" ne sera pas satisfaisant si (contrairement à mon sentiment), il s'avère que les ordinateurs ne se montrent pas bons dans le jeu.

Il y a déjà un certain nombre d'ordinateurs digitaux capable de travailler, et on peut se demander "Pourquoi ne pas tenter cette expérience ? Il serait facile de satisfaire les conditions du jeu. Un certain nombre d'interrogateurs pourraient être utilisés, et des statistiques pourraient être calculées qui compteraient le nombre de fois où l'identification aurait été correcte". La réponse rapide à cela est que nous ne nous demandons pas si tous les ordinateurs digitaux seraient capables de jouer à ce jeu, ni si les ordinateurs actuels pourraient le faire, mais si on peut imaginer des ordinateurs capables de le faire. Mais c'est seulement une réponse rapide. Nous verrons cette question sous un autre angle ultérieurement.

4 Ordinateurs digitaux

L'idée derrière les ordinateurs digitaux peut être expliquée en disant que ces machines sont destinées à prendre en charge toutes les opérations qui pourraient être effectuées par des calculateurs humains. Le calculateur humain est supposé suivre des règles fixées ; il n'a pas autorité à dévier d'elles dans le moindre détail. On peut supposer que ces règles sont fournies dans un livre, qui est modifié à chaque fois qu'une nouvelle tâche est à réaliser. Il a aussi une quantité de papier illimitée sur laquelle il peut faire ses calculs. Il peut aussi faire ses multiplications et additions sur une "calculatrice de bureau", mais cela n'a pas d'importance.

Si nous utilisons l'explication ci-dessus comme une définition, nous pouvons risquer d'être confronté à un argument circulaire. On évite cela en donnant un aperçu des moyens par lesquels l'effet désiré peut être obtenu. Un ordinateur digital peut habituellement être vu comme constitué de 3 parties :

(i) la mémoire.

(ii) l'unité d'exécution.

(iii) le contrôle.

La mémoire stocke l'information, et correspond au papier du calculateur humain, que ce soit le papier sur lequel il fait ses calculs ou celui sur lequel son livre de règles est imprimé. Puisque l'humain peut effectuer une partie de ses calculs de tête, une partie de la mémoire de la machine correspond à cette mémoire du calculateur humain.

L'unité d'exécution est la partie qui effectue les opérations individuelles impliquées dans un calcul. Ce que sont ces opérations variera d'une machine à l'autre. Habituellement des opérations assez longues peuvent être effectuées comme "Multiplier 3540675445 par 7076345687" mais pour certaines machines, seules des opérations très simples comme "Ecris 0" sont envisageables.

Nous avons mentionné que le "livre de règles" fourni à l'ordinateur est stocké dans la machine dans une partie de sa mémoire. On appelle cette partie de la mémoire la "table d'instructions". C'est la tâche du contrôle de voir que ces instructions sont exécutées correctement et dans le bon ordre. Le contrôle vérifie que ces contraintes sont respectées.

L'information en mémoire est habituellement découpée en paquets de taille modérément petite. Dans une machine, par exemple, un paquet peut consister en dix unités décimales. Des nombres sont assignés aux parties de la mémoire dans lesquelles les différents paquets d'information sont stockés, de manière systématique. Une instruction typique peut dire :

"Ajoute le nombre stocké à la position 6809 à celui en 4302 et met le résultat obtenu dans cette dernière unité de mémoire utilisée."

Il est inutile de dire que cela ne sera pas stocké en machine en anglais courant. Ca a plus de chances d'être codé dans une forme comme 6809430217. Ici 17 dit quelle opération doit être faite sur les deux nombres. Dans ce cas, l'opération est celle décrite plus haut, i.e. "Ajouter les nombres...". On notera que l'instruction prend 10 chiffres et constitue ainsi un paquet d'information, de façon très pratique. Le contrôle prendra normalement les instructions devant être effectuées dans l'ordre dans lequel elles ont été stockées, mais occasionnellement, une instruction comme "N'exécute pas l'instruction à la position 5606, et continue à partir de là" peut être rencontrée, ou à nouveau "Si à la position 4505, il y a un 0, exécute ensuite l'instruction stockée en 6707, sinon continue séquentiellement".

Les instructions de ces dernières sortes sont très importantes parce qu'elles rendent possible le remplacement d'une séquence d'instructions par une autre plusieurs fois de suite jusqu'à ce qu'une certaine condition soit remplie, et ce faisant, d'exécuter non pas les nouvelles instructions à chaque répétition, mais la même instruction plusieurs fois successives. Pour prendre une analogie domestique, supposons que Maman veuille que Tommy passe chez le cordonnier chaque matin lorsqu'il va à l'école pour demander si ses chaussures sont prêtes, elle peut le lui rappeler chaque matin. Alternativement, elle peut coller un papier une fois pour toutes dans le hall et il le verra quand il part à l'école et cela lui rappellera de demander pour les chaussures, et le post-it sera détruit quand Tommy reviendra avec les chaussures réparées.

Le lecteur doit accepter cela comme un fait que les ordinateurs digitaux peuvent être construits, et par exemple ont été construits, selon les principes que nous avons décrits, et qu'ils peuvent quasiment simuler les actions d'un calculateur humain.

Le livre de règles qu'utilise le calculateur humain dont nous avons parlé est bien sûr une fiction pratique. Les véritables calculateurs humains se rappellent vraiment ce qu'ils ont à faire. Si l'on veut qu'une machine simule le comportement d'un calculateur humain pour des tâches complexes, on doit lui demander comment il fait et ensuite traduire sa réponse en utilisant une table d'instructions. Construire des tables d'instructions est habituellement appelé "programmer". "Programmer une machine pour qu'elle fasse l'opération A" signifie mettre l'instruction appropriée dans la machine de manière à ce qu'elle exécute A.

Une variante intéressante à l'idée d'ordinateur digital est celle d'"ordinateur digital contenant un composant aléatoire". Ces ordinateurs ont des instructions impliquant un lancer de dé ou un processus électronique équivalent ; une telle instruction par exemple peut être "lancer le dé et mettre le nombre résultant dans la case mémoire 1000". Parfois une telle machine est décrite comme possédant un libre-arbitre (même si je n'utiliserai pas cette expression moi-même). Il n'est normalement pas possible de déterminer, en observant la machine, si elle contient un composant basé sur l'aléa, car un effet similaire peut être produit par les composants non aléatoires en rendant les choix dépendant des chiffres des formes décimales des nombres en jeu.

La plupart des ordinateurs digitaux actuels ont seulement une mémoire limitée. Il n'y a pas de difficulté théorique dans l'idée d'un ordinateur avec une mémoire illimitée. Bien sûr que seule une partie limitée de la mémoire peut être utilisée à tout instant, puisque seulement une quantité finie de mémoire a pu être fabriquée, mais on peut imaginer qu'on en rajoutera de plus en plus lorsque ça sera nécessité. De tels ordinateurs ont un intérêt théorique spécifique et nous les appellerons ordinateurs de capacité infinie.

L'idée d'ordinateur digital est une idée ancienne. Charles Babbage, Professeur Lucasien de Mathématique à Cambridge de 1828 à 1839, avait conçu une telle machine, appelée le moteur analytique, mais il ne l'a jamais terminé. Bien que Babbage ait eu les idées principales, sa machine n'était pas à l'époque un projet très attractif. La vitesse qui aurait été atteinte par cette machine aurait définitivement été plus rapide que celle d'un calculateur humain mais quelque-chose comme 100 fois plus lente que la machine de Manchester, elle-même l'une des plus lentes des machines modernes. La mémoire devait être purement mécanique, et utiliser des roues et des cartes.

Le fait que le moteur analytique de Babbage soit complètement mécanique nous aidera à nous débarrasser d'une superstition. On attache souvent de l'importance au fait que les ordinateurs digitaux modernes sont électriques, et que le système nerveux est également électrique. Puisque la machine de Babbage n'était pas électrique, et puisque tous les ordinateurs digitaux sont en quelque sorte équivalents, nous voyons que cette utilisation de l'électricité ne peut pas avoir d'importance théorique. Bien sûr, l'électricité intervient lorsqu'on doit traiter des signaux rapidement, et ce n'est donc pas surprenant que nous la trouvions dans ces deux sortes de connexions. Dans le système nerveux, les phénomènes chimiques sont au moins aussi importants que les phénomènes électriques. Dans certains ordinateurs, le système de mémoire est principalement acoustique. Le fait d'utiliser l'électricité est ainsi vu comme une simple similarité superficielle. Si nous souhaitons trouver de telles similarités, nous devrions plutôt chercher des analogies mathématiques au niveau des fonctions.

5 L'universalité des ordinateurs digitaux

Les ordinateurs digitaux considérés dans la section précédente peuvent être classés parmi les “machines à états discrets”. Ce sont des machines qui bougent par sauts ou clics d'un état parfaitement défini à un autre. Ces états sont suffisamment différents pour que la possibilité de confusion entre eux soit ignorée. De façon stricte, il n'y a pas de sous-machines. Tout va continûement dans la réalité. Mais il y a de nombreuses sortes de machines pour lesquelles il est profitable de les penser comme des machines à états discrets. Par exemple, pour les états d'un système lumineux, c'est une fiction pratique que d'imaginer le système soit comme complètement allumé, soit comme complètement éteint. Il doit y avoir des positions intermédiaires, mais dans de nombreux contextes, nous pouvons les oublier. Comme exemple de machine à états discrets, nous pouvons considérer une roue qui tourne 120 fois par seconde, mais qui peut être arrêtée par un levier de l'extérieur ; supposons en plus qu'une lampe doive être allumée lorsque la roue est dans une certaine position. Cette machine pourrait être décrite abstraitement comme suit. L'état interne de la machine (qui est décrit par la position de la roue) peut être q_1, q_2 ou q_3. Il y a un signal en entrée i_0 ou i_1 (position du levier). L'état interne à tout moment est déterminé par l'état final et l'état d'entrée selon la table

		Etat final		
		q_1	q_2	q_3
Signal en entrée	i_0	q_2	q_3	q_1
	i_1	q_2	q_3	q_1

Les signaux de sortie, les seules indications visibles de l'extérieur de l'état interne (les lumières) sont décrites par la table

Etat interne	q_1	q_2	q_3
Sortie	o_0	o_0	o_1

Cet exemple est typique des machines à états discrets. Elles peuvent être décrites par de telles tables en supposant qu'elles n'ont qu'un nombre fini d'états possibles.

Il semble que selon un état initial donné de la machine et selon les signaux d'entrée, il soit toujours possible de prédire tous les états futurs. Cela rappelle le point de vue de Laplace selon lequel si l'on connaît l'état complet de l'univers à un instant temporel donné, comme décrit par les positions et vitesses de toutes les particules, on devrait pouvoir prédire tous ses états futurs. La prédiction que nous sommes en train de considérer ici, pourtant, est, cependant, bien plus proche de la prédictabilité que celle considérée par Laplace. Le système de l'"univers comme un tout" est tel que des erreurs plutôt petites dans les conditions initiales peuvent avoir un gros effet plus tard. Le déplacement d'un seul électron d'un billionième de centimètre à un moment peut faire toute la différence entre un homme tué par une avalanche un an après, ou bien en réchappant. C'est une propriété essentielle des systèmes mécaniques que nous avons appelées "machines à états discrets" que ce phénomène ne se produise pas. Même quand nous considérons de vraies machines physiques plutôt que des machines idéelles, une connaissance raisonnable de l'état à un moment donné amène une connaissance raisonnable de l'état quelques étapes plus tard.

Comme nous l'avons mentionné, les ordinateurs digitaux font partie de la classe des machines à états discrets. Mais le nombre d'états d'une telle machine peut être vraiment très grand. Par exemple, le nombre d'états pour une machine de Manchester est environ de $2\ 165,00$, i.e., environ $10 \times 50\ 000$. Comparons cela à notre exemple de la roue à clics décrite précédemment, qui a trois états. Il n'est pas difficile de voir pourquoi le nombre d'états pourrait être aussi grand. L'ordinateur contient une mémoire correspondant au papier utilisé par un calculateur humain. Il doit être possible d'écrire dans la mémoire n'importe quelle combinaison de symboles qui pourrait avoir été écrite sur le papier. Pour des raisons de simplicité, supposons que l'on n'utilise comme symboles que les chiffres de 0 à 9. Les variations d'écriture sont ignorées. Supposons que l'ordinateur puisse avoir 100 feuilles de papier, chacune contenant 50 lignes de 30 chiffres chacune. Alors le nombre d'états des trois machines mises ensemble est 10 $100 \times 50 \times 30$ i.e., 10 150 000. C'est à peu près le nombre d'états de trois machines de Manchester mises ensemble. Le logarithme en base 2 de ce nombre d'états est habituellement appelé la "capacité mémoire" de la machine. Ainsi, la machine de Manchester a une capacité mémoire d'environ 165 000 et la machine à roues de notre exemple a une mémoire de taille 1.6. Si deux machines sont mises ensemble, leurs capacités doivent être ajoutées pour obtenir la capacité de la machine résultante. Cela amène à la possibilité d'assertion comme "La machine de Manchester contient 64 barrettes magnétiques chacune d'une capacité de 2560, huit tubes électroniques avec une capacité de 1280. Des petits ajouts accessoires de mémoires s'élèvent à 300, ce qui fait un total de 174 380."

La table correspondant à la machine à états discrets étant donnée, il est possible de prédire ce qu'elle fera. Il n'y a pas de raison que ce calcul ne puisse pas être effectué par un ordinateur digital. Si l'on suppose qu'il peut l'effectuer assez rapidement, l'ordinateur digital devrait pouvoir simuler le comportement d'une machine à états discrets quelconque. On pourrait alors imaginer jouer au jeu de l'imitation avec la machine en question (comme B) et l'ordinateur digital simulateur (comme A) et l'interrogateur ne pourrait pas les distinguer l'un de l'autre. Bien sûr, l'ordinateur digital doit avoir une capacité mémoire adéquate et doit être en mesure d'exécuter ses instructions suffisamment rapidement. De plus, il doit être programmé à nouveau à chaque fois qu'on souhaite le faire simuler une nouvelle machine.

Cette propriété spéciale des ordinateurs digitaux, le fait qu'ils puissent simuler n'importe quelle machine à états discrets, est décrit en les désignant par l'expression "machines universelles". L'existence de machines avec cette propriété a comme conséquence importante que, les considérations de vitesse étant mises à part, il n'est pas nécessaire de concevoir de nouvelles machines variées pour réaliser des processus de calcul variés. Cela peut aussi être fait avec un ordinateur digital, chacun adéquatement programmé pour chaque cas. On verra qu'une conséquence de cela est que tous les ordinateurs digitaux sont équivalents dans un certain sens.

Nous pouvons maintenant considérer à nouveau le point mis en avant à la fin du §3. On avait suggéré conjecturalement que la question "Les machines peuvent-elles penser" soit remplacée par "Les ordinateurs digitaux imaginables pourraient-ils se débrouiller dans le jeu de l'imitation ?". Si nous le souhaitons, nous pouvons rendre cela plus général de façon artificielle et demander "Y a-t-il des machines à états discrets qui seraient capables de faire cela ?". Mais en regardant la propriété universelle, nous voyons que l'une ou l'autre de ces questions est équivalente à celle-ci : "Fixons notre attention sur un ordinateur digital C. Est-il vrai qu'en modifiant cet ordinateur pour avoir une mémoire adéquate, et en augmentant adéquatement sa rapidité d'exécution, et en lui fournissant un programme adéquat, C peut être fabriqué de façon à jouer de façon satisfaisante la partie A du jeu de l'imitation, la partie B étant tenue par un humain ?".

6 Des vues différentes de la même question

Nous pouvons considérer maintenant que le socle a été clarifié et nous sommes prêts à débattre de notre question "Les machines peuvent-elles penser ?" et de la variante qui a été citée à la fin de la section précédente. Nous ne pouvons pas abandonner la forme originale du problème, car les opinions diffèreront sur le caractère approprié de la substitution des problèmes et nous devons au moins entendre ce qui doit être dit au sujet de cette relation.

Cela simplifiera les choses pour le lecteur si j'explique d'abord mes propres convictions sur le sujet. Considérons d'abord la forme la plus précise de la question. Je crois que dans cinquante ans environ, il sera possible de programmer des ordinateurs d'une capacité d'envison 109, pour les faire jouer au jeu de l'imitation de façon à ce qu'un interrogateur n'ait plus que 70 pour cent de chances de faire l'identification correcte après cinq minutes de questions/réponses. La question originale, "Les machines peuvent-elles penser ?", je la crois trop privée de sens pour mériter une discussion. Pourtant, je crois qu'à la fin du siècle, l'usage des mots et l'opinion éduquée en général aura tellement changé qu'on sera capable de parler de pensée des machines sans s'attendre à être contredit. Je crois même qu'on ne sert aucun but utile en cachant de telles convictions. L'idée populaire que les scientifiques avancent inexorablement d'un fait établi à un fait établi, en n'étant jamais influencés par aucune conjecture améliorée, est légèrement erronée. A partir du moment où l'on sait clairement ce qui constitue les faits et ce qui constitue les conjectures, il n'en résulte aucun préjudice. Les conjectures sont d'une grande importance puisqu'elles suggèrent les orientations utiles de la recherche.

Je vais maintenant considérer les opinions opposées à la mienne.

(1) L'objection théologique

Penser est une fonction de l'âme immortelle humaine. Dieu a donné une âme immortelle à chaque homme et à chaque femme, mais non pas aux animaux ou aux machines. De ce fait, aucun animal et aucune machine ne peut penser.

Je ne peux accepter aucun élément de l'argument ci-dessus, mais vais essayer d'y répondre en termes théologiques. Je trouverais l'argument plus convaincant si les animaux étaient classés du côté des humains, parce qu'à mon sens, la différence entre l'inanimé et l'animé est plus grande que celle entre l'humain et les autres animaux. Le caractère arbitraire d'un tel point de vue orthodoxe devient clair si nous considérons la manière dont cet argument peut être perçu par un membre d'une autre communauté religieuse. Comment les Chrétiens regardent-ils l'opinion musulmane selon laquelle les femmes n'ont pas d'âme ? Mais laissons ce point de côté et retournons à l'argument principal. Il me semble que l'argument cité ci-dessus entraîne une sérieuse restriction à l'omnipotence du Tout-Puissant. Il est admis qu'il y a certaines choses qu'Il ne peut pas faire comme par exemple faire que un soit égal à deux, mais ne devrions-nous pas croire qu'Il a la liberté de donner une âme à un éléphant s'Il trouve cela approprié ? Nous pourrions nous attendre à ce qu'Il puisse exercer son pouvoir conjointement à une mutation qui pourvoierait l'éléphant d'un cerveau adéquatement amélioré pour gérer des besoins de ce type. Un argument du même type peut être utilisé dans le cas des machines. Il peut sembler différent parce qu'il est plus difficile à "avaler". Mais cela signifie seulement que nous pensons qu'il serait moins vraisemblable qu'Il considère les circonstances appropriées pour leur conférer une âme. Les circonstances en question sont discutées dans le reste de cet article. En essayant de créer de telles machines, nous ne serions pas irrespectueux en usurpant Son pouvoir de créer des âmes, de même que nous ne le sommes pas quand nous procréons et avons des enfants : nous sommes plutôt, dans les deux cas, des instruments de Sa volonté pour créer les réceptacles des âmes qu'Il crée.

Pourtant, ceci, c'est de la pure spéculation. Je ne suis pas très impressionné par les arguments théologiques, quel que soit ce qu'ils sont destinés à expliquer. De tels arguments se sont souvent avérés insatisfaisants par le passé. A l'époque de Galilée, il avait été dit que les textes, "Et le soleil dura alors... et ne descendit pas pendant une journée complète" (Josué 10 :13) et "Il a posé les fondations de la Terre, de manière à ce qu'elle ne bouge jamais" (Psaume 104) étaient une réfutation de la théorie de Copernic. Avec nos connaissances actuelles, un tel argument semble futile. Quand cette connaissance n'était pas encore acquise, cela a fait une impression plutôt différente.

(2) L'objection "la tête dans le sable"

"Les conséquences du fait que des machines pensent seraient trop horribles. Espérons et croyons qu'elles ne pourront jamais le faire."

Cet argument est rarement exprimé de manière si claire qu'il ne l'est ci-dessus. Mais il affecte beaucoup d'entre nous qui le pensent complètement. Nous aimons penser que l'Homme est en quelque sorte supérieur au reste de la création. Il est mieux qu'il puisse être vu comme nécessairement supérieur, car alors il n'y a pas de danger qu'il perde sa position dominante. La popularité de l'argument théologique est clairement reliée à ce sentiment. Il est vraisemblable qu'un tel avis soit très partagé par les intellectuels, parce qu'ils considèrent le pouvoir de la pensée comme plus important que ne le font d'autres personnes, et ils sont donc plus enclins à baser leurs convictions sur la supériorité de l'Homme concernant cette possibilité.

Je ne crois pas que cet argument soit suffisamment substantiel pour nécessiter une réfutation. La consolation serait plus appropriée : peut-être qu'elle pourrait être recherchée dans la métempsychose.

(3) L'objection mathématique

Il y a un certain nombre de résultats de logique mathématique qui peuvent être utilisés pour montrer qu'il y a des limitations à ce que peuvent les machines à états discrets. Le plus connu de ces résultats est le théorème de Gödel (1931) qui montre que dans tout système logique suffisamment puissant, des assertions peuvent être formulées qui ne peuvent ni être prouvées ni être réfutées dans ce système, à moins que le système lui-même dans son ensemble ne soit démontré comme étant inconsistant. Il y a d'autres résultats, similaires à celui-là en quelque sorte, dus à Church (1936), Kleene (1935), Rosser, et Turing (1937). Le dernier résultat est le plus pratique à considérer, puisqu'il fait directement référence aux machines, alors que les autres peuvent seulement être utilisés dans un argument indirect : par exemple, si le théorème de Gödel doit être utilisé, on a besoin d'avoir en plus des moyens de décrire les systèmes logiques par rapport aux machines, et les machines par rapport aux systèmes logiques. Le résultat en question [1] fait référence à un type de machine qui consiste essentiellement en un ordinateur digital avec une capacité infinie. Il établit qu'une machine ne peut effectuer certaines tâches. Si une telle machine doit donner des réponses à des questions comme dans le jeu de l'imitation, il y aura des questions auxquelles soit elle donnera une réponse fausse, soit elle échouera à donner une quelconque réponse quel que soit le temps qui lui est alloué pour y répondre. Il peut, bien sûr, exister de nombreuses telles questions, et les questions auxquelles une certaine machine ne pourra pas répondre pourront cependant recevoir une réponse satisfaisante de la part d'une autre machine. Nous supposons là que les questions sont de type fermé, i.e. elles attendent comme réponse appropriée une réponse “Oui” ou “Non”, plutôt que des questions ouvertes comme “Que pensez-vous de Picasso ?”. Les questions dont nous savons que les machines doivent échouer sont de ce type. “Considérons les machines spécifiées comme suit... Cette machine répondra-t-elle “oui” à toute question ?”. Les points de suspension doivent être remplacés par une description d'une machine de forme standard, qui pourrait être du type de celles qui ont été envisagées au §5. Quand la machine décrite partage une certaine relation comparativement simple avec la machine que l'on interroge, on peut montrer que la réponse est soit fausse soit ne vient jamais. C'est cela le résultat mathématique : on soutient qu'il prouve l'incapacité des machines qui n'ont pas un intellect humain.

1. *ndt* : le résultat le plus pratique

La réponse courte à cet argument est que même s'il est établi qu'il y a des limitations à la possibilité de penser pour une machine particulière, il a seulement été affirmé, sans aucune sorte de preuve, que de telles limitations ne s'appliquent pas à l'intellect humain. Mais je ne pense pas qu'un tel point de vue doive être rejeté si clairement. A chaque fois que l'on pose à une telle machine une question critique appropriée, et qu'elle fournit une réponse, et que nous savons que cette réponse est fausse, cela nous donne un certain sentiment de supériorité. Ce sentiment est-il illusoire ? C'est sans doute vrai, mais je ne pense pas que trop d'importance doive être accordée à cela. Nous donnons nous-mêmes trop souvent des mauvaises réponses à des questions pour être satisfaits que cela soit utilisé comme évidence de la faillibilité des machines. De plus, notre supériorité peut seulement se ressentir dans certaines occasions précises, en relation avec une machine spécifique contre laquelle nous avons enregistré un petit triomphe. Il ne saurait être question de triompher simultanément de toutes les machines. En bref, de ce fait, il se pourrait qu'il existe un humain qui soit plus intelligent que n'importe quelle machine, mais alors à nouveau, il y aura des machines plus intelligentes que lui, et etc.

Ceux qui croient en cet argument mathématique devraient, je pense, accepter la plupart du temps le jeu de l'imitation comme base de discussion. Ceux qui sont convaincus des deux objections précédentes ne seraient probablement intéressés par aucun critère quel qu'il soit.

(4) L'argument de la conscience

Cet argument est très bien exprimé par le Professeur Jefferson dans son discours d'obtention de la médaille Lister, dont je cite : "Pas avant qu'une machine n'ait écrit un sonnet ou composé une symphonie à cause d'émotions et pensées ressenties, et pas par le hasard de concordance de symboles, nous ne pourrons être d'accord sur le fait qu'une machine égale un cerveau humain, c'est-à-dire que non seulement cette machine écrit mais qu'en plus, elle sait ce qu'elle a écrit. Aucun mécanisme ne pourrait ressentir de plaisir lorsqu'il réussit (et pas seulement des signaux artificiels, des stratagèmes faciles), ne pourrait ressentir de difficulté quand ses vannes fusionnent, être réconforté par des flatteries, ou rendu misérable par ses erreurs, charmé par le sexe, en colère ou déprimé parce qu'il n'arrive pas obtenir ce qu'il veut.".

Cet argument semble être un déni de la validité de notre test. Selon une forme extrême de ce point de vue selon lequel le seul moyen d'être sûr qu'une machine pense est d'être une machine et de se sentir penser. On pourrait alors décrire ce qu'une machine ressent au monde, mais bien sûr personne n'aurait la possibilité de donner son avis. Selon ce point de vue également, le seul moyen de savoir si un homme pense est d'être cet homme particulier. C'est en fait un point de vue solipsiste. Il peut être le point de vue le plus logique mais il rend la communication des idées difficile. A est susceptible de croire "A pense mais B ne pense pas" tandis que B croit "B pense mais pas A.". Plutôt que de continuer à tergiverser éternellement sur ce point, il est habituel d'avoir la convention polie que tout le monde pense.

Je suis sûr que le Professeur Jefferson ne souhaite pas adopter le point de vue extrême et solipsiste. Il serait vraisemblablement prêt à accepter le jeu de l'imitation comme test. Le jeu (sans le joueur B) est fréquemment utilisé en pratique sous le nom de *viva voce* pour découvrir si quelqu'un comprend effectivement quelque-chose ou bien "répète cette chose comme un perroquet". Ecoutons un tel extrait du jeu *viva voce* :

Interrogateur : Dans la première ligne de votre sonnet "Je te comparerai à une journée d'été", est-ce qu'"un jour de printemps" serait aussi bien ou mieux ?

Témoin : Ca n'irait pas.

Interrogateur : Et "un jour d'hiver", ça sonnerait bien.

Témoin : Oui, mais personne n'a envie d'être comparé à un jour d'hiver.

Interrogateur : Diriez-vous que M. Pickwick vous rappelle Noël ?

Témoin : D'une certaine manière, oui.

Interrogateur : Bien, Noël est un jour d'hiver, et je ne pense que M. Pickwick soit gêné par la comparaison.

Témoin : Je ne pense pas que vous ayez raison. Par un jour d'hiver, on entend un jour d'hiver basique, plutôt qu'un jour aussi spécial que le jour de Noël.

Et le tout à l'avenant... Que dirait le Professeur Jefferson si la machine à écrire des sonnets était capable de répondre ainsi au jeu de *viva voce* ? Je ne sais pas s'il considèrerait la machine comme "fournissant plutôt artificiellement" ces réponses, mais si les réponses étaient aussi satisfaisantes et soutenues que dans l'extrait ci-dessus, je ne pense pas qu'il la décrirait comme "un stratagème facile". Cette phrase est destinée, je pense, à couvrir des mécanismes tels que l'inclusion dans la machine d'un enregistrement de quelqu'un lisant un sonnet, avec la possibilité appropriée de la mettre en marche de temps en temps.

En résumé donc, je pense que la plupart de ceux qui sont en faveur de l'argument de la conscience devraient être persuadés d'abandonner un tel point de vue plutôt que d'être forcés à prendre une position solipsiste. Ils souhaiteraient alors probablement accepter notre test.

Je ne veux pas donner l'impression que je pense qu'il n'y a aucun mystère à propos de la conscience. Il y a, par exemple, quelque-chose de paradoxal lié au fait de souhaiter la localiser. Mais je ne pense pas que ces mystères aient nécessairement besoin d'être résolus avant que nous puissions répondre à la question à laquelle nous nous intéressons dans le présent article.

(5) Des arguments d'impossibilités diverses

Ces arguments sont de la forme "Je vous accorde le fait que vous puissiez fabriquer des machines qui font toutes les choses que vous avez mentionnées mais vous ne serez jamais capable d'en fabriquer une qui puisse faire X.". De nombreuses possibilités sont suggérées pour X dont je fournis une sélection :

Etre gentil, ingénieux, beau, amical, prendre des initiatives, avoir le sens de l'humour, dire ce qui est vrai ou pas, faire des erreurs, tomber amoureux, aimer les fraises à la chantilly, rendre quelqu'un amoureux de soi, apprendre de l'expérience, utiliser les mots à bon escient, être le sujet de ses propres pensées, présenter autant de variété dans son comportement qu'un être humain, faire quelque-chose de vraiment nouveau.

Aucun étayage pour soutenir ces assertions n'est en général fourni. Je crois que ces arguments sont la plupart du temps fondés sur le principe de l'induction scientifique. Un humain a vu des milliers de machines dans sa vie. De ce qu'il a vu d'elles, il tire un certain nombre de conclusions. Elles sont laides, chacune d'elle est conçue pour atteindre un objectif spécifique, si on a un autre objectif que celui-là, elles ne nous sont d'aucune utilité, la variété de comportement de toutes ces machines est très faible, etc., etc. Naturellement, il conclut que ces propriétés sont des propriétés nécessaires des machines en général. Beaucoup de telles limitations sont associées à la très faible capacité mémoire de la plupart des machines (je suppose que l'idée de capacité mémoire est étendue de façon à couvrir les machines autres que celles à états discrets. La définition exacte n'importe pas puisqu'aucune précision mathématique n'est revendiquée dans la présente discussion). Il y a quelques années, quand on avait encore peu entendu parlé des ordinateurs digitaux, il était possible d'éluder une telle incrédulité, si l'on mentionnait leurs propriétés sans décrire leur construction. Cela était sûrement dû à une application similaire du principe d'induction scientifique. Ces applications du principe sont bien sûr largement inconscientes. Quand un enfant qui s'est brûlé craint le feu et montre qu'il le craint en l'évitant, on pourrait dire qu'il applique une induction scientifique (je pourrais décrire son comportement de nombreuses autres façons). Les travaux et les habitudes des humains ne semblent pas être un matériau adéquat auquel appliquer l'induction scientifique. Une grande partie de l'espace-temps doit être étudié, si on souhaite obtenir des résultats fiables. Sinon nous pourrions (comme le font la plupart des enfants anglais) décider que tout le monde parle anglais, et qu'il est idiot d'apprendre le français.

Il y a, cependant, des remarques particulières à faire à propos d'un certain nombre d'incapacités (des machines) qui ont été mentionnées. L'impossibilité d'aimer les fraises à la chantilly pourrait avoir semblé futile au lecteur. On pourrait fabriquer une machine pour apprécier ce met délicieux, mais toute tentative de faire cela semblerait débile. Ce qui est important dans cette incapacité, c'est qu'elle contribue à d'autres incapacités, e.g. à la difficulté qu'il y ait la même sorte d'amitié entre un homme et une machine qu'entre un homme et un autre.

L'argument "Les machines ne peuvent pas se tromper." semble curieux. On est tenté de rétorquer "Sont-elles pires de ce fait ?". Mais adoptons une attitude plus sympathique, et essayons de voir ce que l'on cherche réellement à dire par là. Je pense que cette critique peut s'expliquer selon le jeu de l'imitation. L'argument dit que l'interrogateur pourrait distinguer la machine de l'humain simplement en lui posant un certain nombre de problèmes d'arithmétique. La machine serait démasquée à cause de sa piètre compétence à les résoudre. La réponse à cela est simple. La machine (programmée pour jouer au jeu) n'essaierait pas de donner des réponses correctes aux problèmes arithmétiques. Elle introduirait délibérément des erreurs de manière à tromper l'interrogateur. Une erreur mécanique lui montrerait probablement à travers une décision inadéquate quelle sorte d'erreur faire dans un problème arithmétique. Même cette interprétation de la critique n'est pas suffisamment sympathique. Mais nous ne pouvons perdre de la place à entrer plus avant dans les détails. Il me semble que cette critique dépend d'une confusion entre deux sortes d'erreurs. Nous pourrions les appeler les "erreurs de fonctionnement" et les "erreurs de raisonnement". Les erreurs de fonctionnement sont dues à des défauts électriques ou mécaniques qui empêchent la machine de se comporter comme elle a été programmée à le faire. Dans les discussions philosophiques, on aime ignorer ces possibilités d'erreurs ; ce faisant, on discute de "machines abstraites". Ces machines abstraites sont des idées mathématiques plutôt que des objets physiques. Par définition, elles sont incapables d'erreurs de fonctionnement. C'est en ce sens qu'on peut dire que "les machines ne font jamais d'erreurs". Les erreurs de raisonnement peuvent quant à elles seulement advenir quand un sens est attaché aux signaux en sortie de la machine. La machine pourrait, par exemple, écrire des équations mathématiques, ou des phrases en anglais. Quand une phrase fausse est tapée, nous disons que la machine a fait une erreur de raisonnement. Il n'y a bien sûr aucune raison de dire qu'une machine ne peut jamais faire ce genre d'erreur. Elle pourrait ne rien faire et écrire sans fin "$0 = 1$". Pour prendre un exemple moins pervers, elle pourrait avoir une méthode pour trouver ses conclusions par induction scientifique. Nous pouvons nous attendre à ce qu'une telle méthode puisse parfois amener à des résultats erronés.

On peut répondre à l'argument selon lequel une machine ne peut être le sujet de sa propre pensée seulement si l'on peut montrer qu'une machine pense à certains sujets. Néanmoins, "Le sujet des opérations d'une machine" semble ne rien vouloir dire, au moins pour les personnes qui s'intéressent à un tel sujet. Si, par exemple, la machine essaye de trouver une solution de l'équation $x^2 - 40x - 11 = 0$, on peut être tenté de décrire cette équation comme faisant partie du sujet de la pensée de la machine à ce moment-là. Dans ce sens-là, une machine peut sans aucun doute être le sujet de sa propre pensée. Elle peut être utilisée pour mettre à jour ses propres programmes, ou pour prédire les effets des altérations de sa propre structure. En observant les résultats de son propre comportement, elle peut modifier ses propres programmes pour atteindre certains buts plus efficacement. Ce sont des possibilités de l'avenir proche, plutôt que des rêves utopiques.

La critique selon laquelle une machine ne peut pas avoir un comportement aussi diversifié est juste une manière de dire qu'elle ne peut pas avoir beaucoup de capacité mémoire. Jusqu'à assez récemment, une capacité mémoire d'un millier de digits était très rare.

Les critiques que nous considérons ici sont souvent des formes déguisées de l'argument concernant la conscience. Habituellement, si on maintient qu'une machine peut faire l'une de ces choses, et si l'on décrit le genre de méthode que la machine devrait utiliser pour ce faire, cela ne fera pas plus qu'une impression. On pense qu'une telle méthode (quelle qu'elle soit, car elle peut être mécanique) est vraiment plutôt basique. Comparer cela à ce qui est entre parenthèses dans l'assertion de Jefferson vue précédemment.

(6) L'objection de Lady Lovelace

L'information en notre possession la plus détaillée concernant le moteur analytique de Babbage est un mémoire d'Ada Lovelace (1842). Dans celui-ci, elle écrit "Le moteur analytique ne prétend pas *inventer* quoi que ce soit. Il peut faire *tout ce dont nous savons comment le lui ordonner*." (caractères en italique selon l'écrit original d'Ada Lovelace). Cette assertion est citée par Hartree (1949) qui ajoute : "Cela n'implique pas qu'il ne puisse être possible de construire un équipement électronique qui "penserait par lui-même", ou dans lequel, en termes biologiques, quelqu'un pourrait incorporer un réflexe conditionné, qui servirait de base à un "apprentissage". Que cela soit ou pas possible en principe est une question stimulante et excitante, suggérée par certains développements récents. Mais il ne semble pas que les machines construites ou projetées de l'être ont cette propriété.".

Je suis en profond accord avec Hartree sur cela. Il sera noté qu'il ne dit pas que les machines en question n'avaient pas cette propriété, mais plutôt que la conviction de Lady Lovelace ne l'encourageait pas à penser que les machines avaient cette propriété.

Il est assez possible que les machines en question avaient cette propriété dans un certain sens. Car supposons qu'une machine à états discrets ait cette propriété. Le moteur analytique était un ordinateur digital universel, et donc, si sa capacité mémoire et sa vitesse étaient adéquates, il aurait pu par programmation simuler la machine en question. Il est probable que cet argument n'ait pas été trouvé par la Comtesse ou par Babbage. Dans tous les cas, ils n'avaient aucune obligation de dire tout ce qui aurait pu être dit.

L'ensemble de cette question sera considéré à nouveau dans le paragraphe concernant les machines apprenantes.

Une variante de l'objection de Lady Lovelace affirme qu'une machine ne pourra "jamais faire quelque-chose de vraiment nouveau". On peut parer à cet argument par la litanie "Rien de nouveau sous le soleil". Qui peut être certain qu'un "travail original" qu'il a effectué n'était pas simplement la récolte d'une semaille qui a été plantée en lui lorsqu'il a reçu son enseignement, ou l'effet de principes généraux bien connus. Une meilleure variante de l'objection dit qu'une machine ne peut jamais "nous prendre par surprise". Cet argument est un défi plus direct et on peut le contrer directement. Les machines m'ont surpris très souvent. Cela est largement dû au fait que je ne fais pas suffisamment de calcul pour décider de ce que je dois attendre d'elles, ou plutôt parce que, même si je fais un tel calcul, je le fais à toute vitesse, d'une façon négligeante, en prenant des risques. Peut-être que je me dis à moi-même "Je suppose que le voltage ici doit être le même que là : bon, supposons que c'est le cas.". Bien sûr, j'ai souvent tort, et le résultat m'est une surprise car quand l'expérimentation arrive à son terme, j'ai oublié les suppositions que j'ai faites. Admettre cela me laisse ouvert pour écouter des conférences au sujet de ces mauvaises manières, mais ne laisse aucun doute sur ma crédibilité quand je témoigne de ces surprises que je rencontre.

Je ne m'attends pas à ce que cette réponse fasse se taire les points de vue critiques sur mon point de vue. On dira probablement que les surprises sont dues à une action mentale créative de ma part, et ne sauraient être attribuées à la machine. Cela nous ramène à l'argument de la conscience, et loin de l'idée de surprise. C'est une ligne d'argumentation que nous devons considérer comme fermée, mais il est peut-être pire de remarquer que l'appréciation de quelque-chose comme étant surprenant nécessite autant d'"action mentale créative", que l'événement surprenant ait été provoqué par un humain, un livre, une machine ou quoi que ce soit d'autre.

Le point de vue selon lequel les machines ne peuvent pas créer de surprises est dû, je pense, à une erreur que les philosophes et les mathématiciens font particulièrement souvent. C'est la supposition selon laquelle dès qu'un fait est présenté à un esprit, toutes les conséquences de ce fait germent dans l'esprit simultanément à ce fait. Cette supposition est très utile dans de nombreuses circonsctances, mais on oublie trop souvent qu'elle est fausse. Une conséquence naturelle du fait de faire de la sorte est que l'on suppose alors qu'il n'y pas de vertu à traiter les conséquences des données et des principes généraux.

(7) L'argument au sujet de la continuité du système nerveux

Le système nerveux n'est certainement pas une machine à états discrets. Une petite erreur concernant l'information à propos de la taille de l'impulsion nerveuse en entrée d'un neurone peut engendrer une très grande différence sur l'impulsion en sortie. On peut arguer que, cela étant, on ne peut s'attendre à être capable de simuler le système nerveux par un système à états discrets.

Il est vrai qu'une machine à états discrets doit être différente d'une machine continue. Mais si nous nous plaçons dans les conditions du jeu de l'imitation, l'interrogateur ne pourra pas tirer avantage de cette différence. La situation peut être clarifiée si nous considérons une machine continue sonore la plus simple qui soit. Un analyseur différentiel sera parfait (un analyseur différentiel est une certaine sorte de machine qui n'est pas du type à états discrets et qui est utilisée pour certaines sortes de calculs). Certains d'entre eux fournissent leurs réponses sous une forme typée, et sont ainsi capables d'être utilisés dans le jeu d'imitation. Il ne serait pas possible qu'un ordinateur digital prédise exactement quelles réponses l'analyseur digital donnerait à un problème, mais il pourrait être capable de donner la bonne sorte de réponse. Par exemple, si on lui demande de donner la valeur de π (environ 3.1416), il serait raisonnable de choisir au hasard parmi les valeurs 3.12, 3.13, 3.14, 3.15, 3.16 avec les probabilités de 0.05, 0.15, 0.55, 0.19, 0.06 (disons). Dans ces circonstances, il serait très difficile pour l'interrogateur de distinguer l'analyseur digital de l'ordinateur digital.

(8) L'argument provenant du caractère informel du comportement

Il n'est pas possible de produire un ensemble de règles destinées à décrire ce qu'un homme devrait faire dans tout ensemble concevable de circonstances. On pourrait par exemple avoir une règle qui est qu'on doit s'arrêter quand on voit un feu rouge, et avancer quand on voit un feu vert mais que faire dans le cas où ils apparaissent tous les deux simultanément ? On peut peut-être décider qu'il est plus prudent de s'arrêter. Mais d'autres difficultés peuvent alors découler plus tard de cette prise de décision. Essayer de fournir des règles de conduite couvrant toutes les éventualités, même celles provenant des feux de la circulation, semble impossible. Avec tout ça, je suis d'accord.

A cause de ces raisons, on dit que nous ne pouvons pas être des machines. Je vais essayer de reproduire l'argument, mais je crains de ne pas lui rendre justice. C'est un peu quelque chose comme ça : “si tout homme avait un ensemble défini de règles de conduite par lesquelles il pouvait régenter sa vie, il ne serait rien de plus qu'une machine. Mais de telles règles n'existent pas, et donc les hommes ne peuvent pas être des machines.". Le déséquilibre est flagrant. Je ne pense pas que l'argument soit jamais exprimé en ces termes, mais je crois que c'est cependant ce genre d'argument qui est utilisé. La confusion certaine entre les "règles de conduite" et les "lois de comportement" obscurcit le problème. Par "règles de conduite", je veux parler de préceptes tels que "Arrêtez-vous si vous voyez des feux rouges", que l'on peut faire, et dont on peut être conscient. Par "règles de comportement", je veux parler des lois de la nature telles qu'elles s'appliquent à un corps humain comme "si vous le pincez, il va crier". Si vous substituez "lois de comportement qui régentent sa vie" à "lois de conduite par lesquelles il régente sa vie" dans l'argument cité, le juste équilibre devient atteignable. Car nous croyons qu'il est non seulement vrai que voir son comportement commandé par des règles de comportement implique d'être une sorte de machine (même si cela ne signifie pas nécessairement être une machine à états discrets), mais qu'inversement, être une machine implique d'être régenté par de telles lois de comportement. Pourtant, nous ne pouvons pas nous convaincre aussi simplement de l'absence de lois pour tous les comportements possibles, ou de règles de conduite dans toutes les circonstances possibles. La seule manière que nous connaissions pour trouver de telles lois est l'observation scientifique, et nous ne connaissons sûrement aucun contexte dans lequel nous pourrions dire "Nous avons assez cherché. Il n'y a pas de telles lois.".

Nous pouvons démontrer de façon plus forte encore qu'une telle assertion serait injustifiée. Car supposons que nous soyons certains de trouver de telles lois si elles existaient. Alors étant donnée une machine à états discrets, il serait certainement possible de découvrir en l'observant suffisamment comment prédire son comportement futur et cela dans un temps raisonnable, disons mille ans. Mais cela ne semble pas être le cas. J'ai écrit un petit programme sur la machine de Manchester qui n'utilise que 1000 unités mémoire, et ce programme est tel que si l'on fournit à la machine un nombre à seize chiffres en renvoie un autre en deux secondes. Je défie quiconque d'en apprendre suffisamment sur ce programme en étudiant les réponses qu'il fournit, et d'être capable de prédire la réponse qu'il renverrait pour n'importe quelle valeur non déjà testée.

(9) L'argument de la perception extra-sensorielle

Je suppose que le lecteur est familier avec l'idée de perception extra-sensorielle (ESP), et qu'il connaît la signification des quatre termes suivants : la télépathie, la clairvoyance, la préscience et la psychokinèse. Ces phénomènes troublants semblent mettre en défaut toutes nos idées scientifiques habituelles. Comme nous aimerions les discréditer ! Malheureusement l'évidence statistique, au moins pour la télépathie, est accablante. Il est très difficile de réordonner les idées de quelqu'un pour que ces faits cadrent dans sa pensée. Une fois qu'on les a admis, ça n'a pas l'air d'être une grosse étape de plus que de croire aux fantômes. L'idée que nos corps bougent simplement selon les lois connues de la physique, et de quelques autres idées non encore découvertes mais à peu près similaires, va être la première idée vers laquelle se diriger.

Cet argument est selon moi un argument assez fort. On peut dire en réponse à cela que de nombreuses théories scientifiques semblent rester applicables en pratique, malgré leur opposition aux idées de perceptions extra-sensorielles ; c'est en fait ce que l'on obtient très simplement, quand on oublie celles-ci. C'est plutôt confortable, et on craint que la pensée soit la seule sorte de phénomène où les ESP puissent être particulièrement pertinents.

Un argument plus spécifique basé sur les ESP pourrait être le suivant : “Jouons au jeu de l'imitation, en utilisant comme témoins un homme qui est un bon récepteur télépathe, et un ordinateur digital. L'interrogateur peut poser des questions comme “A quelle suite la carte dans ma main droite appartient-elle ?”. L'homme par clairvoyance télépathique donnera la bonne réponse pour 130 cartes sur 400. La machine peut seulement deviner au hasard et elle obtiendra peut-être 104 bonnes réponses, et alors l'interrogateur fera la bonne identification.”. Il y a une possibilité intéressante qui s'ouvre ici. Supposez que l'ordinateur digital contienne un générateur de nombres aléatoires. Alors il sera naturel qu'il utilise ce dispositif pour décider de la réponse à fournir. Mais alors le générateur de nombres aléatoires pourra être manipulé par les pouvoirs psychokinétiques de l'interrogateur. Peut-être que ces pouvoirs psychokinétiques pourraient permettre à la machine de répondre juste plus souvent que ce qui est attendu selon un calcul de probabilités, de telle façon que l'interrogateur ne soit plus capable de faire la bonne identification. D'un autre côté, il pourrait être capable de deviner correctement sans poser aucune question, par sa clairvoyance. Avec l'ESP, tout peut arriver.

Si on admet la télépathie, il sera nécessaire de resserrer nos tests. La situation devrait être regardée comme analogue à celle qui arriverait si l'interrogateur se parlait à lui-même et si l'un des compétiteurs écoutait avec son oreille collée au mur. Mettre les compétiteurs dans une “pièce pour preuve télépathique” satisferait toutes les contraintes.

7 Machines apprenantes

Le lecteur aura compris que je n'ai aucun argument convaincant de nature positive pour appuyer mon point de vue. Si j'en avais, je n'aurais pas fait tant d'efforts pour montrer la fausseté des points de vue contraires. Je vais exposer la conviction qui est la mienne.

Retournons un instant à l'objection de Lady Lovelace, qui exprime que la machine ne peut que faire ce qu'on lui dit de faire. On pourrait dire qu'un humain “injecte” une idée dans la machine, et qu'elle répondra en quelque sorte, et puis retournera dans le calme, comme une corde de piano frappée par un marteau. Une autre image similaire serait celle d'une pile atomique avec moins d'énergie qu'une certaine dose critique : une idée qu'on lui injecte correspond là à un neutron entrant dans la pile de l'extérieur. Un tel neutron provoquera une perturbation certaine qui éventuellement n'aura aucun effet. Si, cependant, la taille de la pile est suffisamment augmentée, la perturbation causée par un électron entrant va continuer à augmenter jusqu'à ce que la pile entière soit détruite. Y a-t-il un phénomène correspondant pour les esprits, y en a-t-il un pour les machines ? Il semble y en avoir un pour l'esprit humain. La majorité d'entre eux semblent être "sous-critiques", i.e. correspondre à cette analogie des piles de taille sous-critique. Une idée présentée à un tel esprit va en moyenne donner naissance à une idée en retour. Une petite proportion d'entre eux sont super-critiques. Une idée présentée à un tel esprit donnera naissance à une "théorie" complète consistant en une idée secondaire, une troisième et d'autres plus éloignées. Les esprits des animaux semblent être définitivement sous-critiques. En adhérant à cette analogie, on se demande "Une machine peut-elle être rendue super-critique ?".

L'analogie de la "peau d'oignon" est aussi utile. En considérant les fonctions de l'esprit ou du cerveau, nous trouvons certaines opérations que nous pouvons expliquer en termes purement mécaniques. Nous disons que cela ne correspond pas à l'esprit véritable : c'est une sorte de peau que nous devons arracher si nous voulons trouver le véritable esprit. Mais là, nous trouvons une nouvelle peau, que nous devons arracher aussi, et etc. En procédant de cette manière, nous n'arrivons jamais au "véritable" esprit, ou bien est-ce que nous finissons par arriver à une peau qui ne contient rien ? Dans ce dernier cas, tout l'esprit est mécanique (ce ne sera pas pour autant une machine à états discrets. Nous avons déjà discuté de cela.).

Ces deux derniers paragraphes ne prétendent pas être des arguments convaincants. Ils devraient plutôt être décrits comme des "récitations tendant à produire une croyance".

Le seul appui réellement satisfaisant qui peut être donné au point de vue exprimé au début du §6, sera d'attendre la fin du siècle et puis de réaliser l'expérience décrite. Mais que pouvons-nous dire en attendant ? Par quelles étapes devrions-nous passer aujourd'hui si l'expérience devait s'avérer une réussite ?

Comme je l'ai expliqué, le problème est principalement un problème de programmation. Des avancées dans l'ingénierie devront avoir être effectives également, mais il ne semble pas qu'elles ne puissent pas permettre ce que l'expérience requiert. Les estimations de la capacité mémoire du cerveau varie de 10^{10} à 10^{15} unités binaires (bits). Je penche pour les valeurs les plus basses et je crois que seule une petite fraction est utilisée pour les types de pensées les plus hautes. La plupart de ces unités sont probablement utilisées pour retenir les impressions visuelles, je ne serais pas surpris si plus de 10^9 était nécessitées pour jouer de manière satisfaisante au jeu de l'imitation, à n'importe quel niveau contre un homme aveugle (Note : la capacité de l'Encyclopaedia Britannica, 11ème édition, est de 2×10^9). Une capacité mémoire de 10^7 serait une possibilité praticable même par les techniques actuelles. Il ne serait probablement pas nécessaire du tout d'augmenter les vitesses de traitement des opérations des machines. Des parties des machines modernes qui peuvent être regardées comme analogues aux cellules nerveuses travaillent environ mille fois plus vite que ces dernières. Cela fournirait une "marge de sécurité" qui pourrait couvrir la perte de vitesse pouvant advenir de multiples façons. Notre problème alors est de trouver comment programmer ces machines pour qu'elles puissent jouer au jeu. A mon niveau de travail actuel, je produis environ un millier de digits de code par jour, ce qui fait qu'environ soixante personnes, travaillant pendant cinquante ans pourraient accomplir le travail, si aucune ligne n'est jamais mise à la poubelle. Des méthodes plus expéditives semblent souhaitables.

Dans la tentative d'imiter l'esprit d'un adulte humain, nous devons forcément répondre au défi de comprendre le processus qui l'a amené à l'état dans lequel il est. Nous pouvons considérer trois composants :

(a) L'état initial de l'esprit, disons à la naissance,
(b) L'éducation à laquelle il a été soumis,
(c) L'expérience, différente de son éducation, qu'il a vécue.

Plutôt que d'essayer de produire un programme pour simuler l'esprit adulte, pourquoi ne pas plutôt essayer d'en produire un qui simule celui de l'enfant ? Si celui-ci était soumis à une éducation adéquate, il permettrait d'obtenir un esprit adulte. On présume que le cerveau de l'enfant est une sorte de carnet de note qu'on achète chez le papetier. Plutôt peu de mécanismes, et beaucoup de papier blanc (mécanisme et écriture sont de notre point de vue presque synonymes). Notre espoir qu'il y ait peu de mécanisme dans le cerveau de l'enfant permettrait de le programmer facilement. La quantité de travail d'éducation que nous pouvons supposer pour notre système doit être environ la même que celle nécessaire pour l'éducation d'un cerveau d'enfant, en première approximation.

Nous avons ainsi divisé notre problème en deux parties : la programmation du cerveau d'enfant et le processus d'éducation. Ces deux aspects restent très intimement connectés. Nous ne pouvons espérer trouver une bonne machine-enfant à la première tentative. On doit alors expérimenter l'enseignement à une telle machine et voir si elle apprend bien. On peut alors en essayer une autre et voir si elle est meilleure ou pire. Il y a une connexion évidente entre ce processus et l'évolution, par les identifications suivantes :

Structure de la machine enfant	=	matériau héréditaire
Changements dans la machine enfant	=	mutation
Sélection naturelle	=	jugement de l'expérimentateur

On peut espérer, cependant, que ce processus sera plus rapide que l'évolution. Le fait que le test d'adaptation perdure est dû à un lent processus de mesure des avantages. L'expérimentateur, en exerçant son intelligence, devrait être capable de l'accélérer. Le fait que le processus ne soit pas restreint à des mutations hasardeuses est aussi important. Si l'on peut retrouver la trace de certaines faiblesses, l'expérimentateur pourra probablement penser aux sortes de mutations qui l'amélioreront.

Il ne sera pas possible d'appliquer exactement le même processus d'apprentissage à une machine qu'à un enfant normal. Elle n'aura pas de jambes, et on ne pourra pas lui demander de sortir et de remplir le seau à charbon. Il est possible qu'elle n'ait pas d'yeux. Mais pourtant, ces déficiences peut être surmontées par une ingénierie intelligente, on ne peut pas envoyer la créature à l'école, les autres enfants ne pourront pas trop jouer avec elle. On doit payer des frais de scolarité. On n'a pas besoin de trop se préoccuper de ces problèmes de jambes, d'yeux, etc. L'exemple de Mademoiselle Helen Keller montre que l'éducation est possible du moment que la communication est possible dans les deux sens entre l'élève et l'enseignant et qu'elle peut être assurée d'une manière ou d'une autre.

Nous associons habituellement des récompenses et des punitions au processus d'enseignement. On peut construire des machines enfants simples ou les programmer selon cette sorte de principe. La machine doit être ainsi construite que les événements qui précèdent l'arrivée d'un signal de punition ne doivent pas être répétés alors que la probabilité de répétition de l'occurrence d'événements qui ont précédé un signal de récompense augmentera. Ces définitions ne présupposent aucun sentiment de la part de la machine, j'ai fait quelques expériences avec une telle machine enfant, et j'ai réussi à lui apprendre quelques petites choses, mais la méthode d'enseignement était trop peu orthodoxe pour qu'on puisse considérer cette expérimentation comme un réel succès.

L'utilisation de punitions et récompenses peut être au mieux une partie du processus d'enseignement. Pour parler grossièrement, si l'enseignant n'a pas d'autre moyen de communiquer avec l'élève, le montant d'information qui peut l'atteindre n'excède pas le nombre total de punitions et de récompenses appliquées. Quand un enfant apprendra à répéter "Casablanca", il sera sûrement très endolori, si ce mot peut seulement être découvert par la technique des "Vingt questions", chaque lettre "o" occasionnant un coup. Il est nécessaire par conséquent d'avoir des canaux de communication "non émotionnels". Si ceux-ci existent, on peut enseigner à une machine par des punitions et des récompenses à obéir à des ordres exprimés dans un certain langage, e.g. un langage symbolique. Ces ordres doivent être transmis à travers des canaux "non émotionnels". L'utilisation de ce langage ne diminuera pas beaucoup le nombre de punitions et récompenses requis.

Les opinions peuvent varier au sujet de la complexité qui est appropriée pour la machine enfant. On pourrait essayer de la rendre aussi simple que possible de façon à satisfaire les principes généraux. De façon alternative, on pourrait avoir un système complet d'inférences logiques "pré-construites". Dans ce dernier cas, la mémoire devrait être largement occupée par les définitions et les propositions. Les propositions auraient plusieurs sortes de statuts, e.g. le statut de faits bien établis, de conjectures, de théorèmes mathématiquement prouvés, d'assertions données par une autorité, d'expressions ayant la forme logique de propositions mais qui ne sont pas des croyances. Certaines propositions peuvent être décrites comme "impératives" (les consignes). La machine pourrait être construite de telle manière que dès qu'un ordre est classé comme "bien établi", l'action appropriée serait automatiquement effectuée. Pour illustrer cela, supposons que l'enseignant dise à la machine "fais tes devoirs maintenant". Cela pourrait avoir comme conséquence l'inclusion de "L'enseignant dit "Fais tes devoirs" " dans les faits bien établis. Un autre tel fait pourrait être "Tout ce que dit l'enseignant est vrai". Combiner ces faits pourrait finalement amener la consigne "Fais tes devoirs maintenant", parmi les faits bien établis, et cela, par construction de la machine, signifierait que les devoirs seraient alors effectués, et cet effet est très satisfaisant. Les processus d'inférence utilisés par la machine n'ont pas besoin d'être tels qu'ils satisfassent les logiciens les plus sévères. Il pourrait par exemple n'y avoir aucune hiérarchie des types. Mais cela ne signifie pas pour autant que des erreurs de type auront lieu, ni que nous allons tomber dans des pièges. Les consignes adéquates (exprimées dans les systèmes, ne faisant pas partie des règles du système) telles que "N'utilisez pas une classe à moins que ce soit une sous-classe d'une classe qui a été mentionnée par l'enseignant" peuvent avoir un effet similaire à "Ne vous approchez pas du bord".

Les consignes auxquelles une machine qui n'a pas de membres peut obéir sont de nature plutôt intellectuelles, comme dans l'exemple donné ci-dessus (des devoirs). Les consignes importantes de l'ensemble des consignes seront les consignes qui déterminent l'ordre dans lequel les règles du système logique concerné doivent être appliquées. Car à chaque étape de la mise en œuvre d'un système logique, il y a un grand nombre d'étapes alternatives, chacune d'entre elles pouvant être appliquée, du moment qu'elle obéit aux règles du système logique. Ces choix font la différence entre un système qui raisonne vite et un système qui raisonne lentement, et non pas la différence entre un système qui raisonne juste, et un système qui raisonne faux. Les propositions amenant à des consignes de cette sorte peuvent être "Quand Socrate est mentionné, utilise le syllogisme de Barbara" ou bien "Si une méthode s'est avérée plus rapide qu'une autre, n'utilise pas la méthode lente.". Certaines de ces règles peuvent être "données par l'autorité" mais d'autres peuvent être produites par la machine elle-même, e.g. par induction scientifique.

L'idée d'une machine apprenante peut sembler paradoxale à certains lecteurs. Comment les règles opératoires de la machine peuvent-elles être modifiées ? Elles devraient décrire complètement comment la machine réagira quelle que soit les événements auxquels elle est soumise, quels que soient les changements qu'elle peut subir. Les règles sont ainsi assez indépendantes du temps. C'est presque vrai. L'explication du paradoxe est que les règles qui seront modifiées par le processus d'apprentissage sont plutôt d'un type moins prétentieux, et n'ont qu'une validité éphémère. Le lecteur pourrait faire un parallèle avec la Constitution des Etats-Unis.

Une propriété importante d'une machine apprenante est que son enseignant sera souvent ignorant de ce qui se passe à l'intérieur de la machine, bien qu'il soit cependant capable dans une certaine mesure de prédire le comportement de son élève. Cela devrait s'appliquer d'autant plus à l'éducation d'une machine qui était précédemment une machine enfant avec un entraînement bien conçu (bien programmé). Ceci est en contraste clair avec la procédure normale qui est que quand on utilise une machine pour faire des calculs, on a une image mentale claire de l'état de la machine à tout moment durant l'exécution du calcul. Cet objectif est difficile à atteindre. Le point de vue qu'"une machine ne peut faire que ce qu'on lui a ordonné de faire" semble étrange face à cela. La plupart des programmes que nous pouvons entrer en machine auront comme résultat qu'elle fera quelque-chose à quoi nous ne pouvons donner du sens (ou du moins, auquel nous attribuerons un sens complètement hasardeux). Se comporter intelligemment consiste justement à s'éloigner du comportement complètement discipliné tel que celui utilisé lorsqu'on effectue un calcul, et d'effectuer plutôt de légers changements, ce qui ne signifie pas pour autant de se comporter de manière aléatoire, ou de faire des répétitions sans fin. Une autre conséquence importante de la préparation de notre machine pour qu'elle puisse exécuter sa partie dans le jeu de l'imitation en la lui enseignant est que la "faillibilité humaine" sera vraisemblablement omise de façon assez naturelle, i.e. sans "coaching" particulier (le lecteur devrait réconcilier cela avec le point de vue de certaines des pages précédentes). Les processus appris ne produisent pas un résultat sûr à cent pour cent ; si c'était le cas, ils ne pourraient pas être désappris.

Il est probablement sage d'inclure un élément aléatoire dans la machine apprenante. Un composant aléatoire est assez utile quand on cherche une solution à un problème. Supposons par exemple que nous voulions trouver un nombre entre 50 et 200 qui soit égal au carré de la somme de ses chiffres. On peut commencer par 51, puis essayer 52 et continuer jusqu'à trouver un nombre qui marche. On peut alternativement choisir des nombres au hasard jusqu'à en trouver un correct. Cette méthode a comme avantage qu'il n'est pas nécessaire de garder trace des valeurs qui ont été testées, mais le désavantage c'est qu'on peut tester deux fois le même nombre, mais ça n'est pas important s'il y a plusieurs solutions possible. La méthode systématique présente le désavantage qu'il peut y avoir un énorme bloc de nombres successifs qui ne sont pas solutions dans la région qu'on testera en premier. Maintenant le processus d'apprentissage peut être vu comme la recherche d'une forme de comportement qui satisfera l'enseignant (ou bien satisfera un autre critère). Puisqu'il y a probablement un très grand nombre de solutions satisfaisantes, la méthode aléatoire semble être meilleure que la méthode systématique.

On peut noter qu'elle est utilisée dans le processus analogue de l'évolution. Mais là, la méthode systématique n'est pas possible. Comment pourrait-il être gardé trace des différentes combinaisons génétiques qui ont été essayées, de manière à éviter de les tester à nouveau ?

Nous pouvons espérer que les machines finiront par entrer en compétition avec les hommes dans des domaines purement intellectuels. Mais quels sont les meilleurs domaines par lesquels commencer ? Même si c'est une question difficile, beaucoup de personnes pensent qu'une activité très abstraite, comme le fait de jouer aux échecs, serait la plus adaptée. On peut aussi arguer qu'il serait plus judicieux de pourvoir les machines des meilleurs organes des sens que l'on peut payer, et d'alors leur apprendre à parler anglais. Ce processus pourrait être identique à celui par lequel on enseigne habituellement à un enfant. On montrerait les choses du doigt et on dirait leur nom, etc. A nouveau je ne sais pas quelle est la bonne réponse, mais je pense que les deux approches devraient être tentées.

Nous ne pouvons que regarder à une petite distance temporelle dans le futur, mais nous voyons là qu'il y a beaucoup de choses à faire.

Machines intelligentes, une théorie hérétique
A. M. Turing

"Vous ne pouvez pas faire qu'une machine pense pour vous". Ceci est un lieu commun qui est habituellement accepté sans questionnement. Ce sera l'objectif du présent article que de questionner cette phrase.

La plupart des machines fabriquées à des fins commerciales sont conçues pour effectuer une tâche très spécifique de façon sûre et extraordinairement vite. Très souvent, une telle machine fait la même série d'opérations de nombreuses fois sans aucune variété. Ce fait à propos des machines réelles est un argument puissant pour de nombreuses personnes de la phrase énoncée plus haut. Pour un mathématicien logicien, cet argument n'est pas valable, car il a été démontré qu'il est possible de fabriquer des machines qui feront quelque chose qui est très proche de la pensée. Elles pourront, par exemple, tester la validité d'une preuve formelle dans le système des *Principia Mathematica*, ou même dire d'une formule d'un tel système si elle est prouvable ou réfutable. Dans le cas où la formule n'est ni prouvable, ni réfutable,une telle machine ne se comportera vraisemblablement pas d'une manière très satisfaisante, car elle continuera à tourner indéfiniment, sans produire de résultat du tout, mais cela ne peut pas être considéré comme très différent comme attitude de la réaction des mathématiciens, qui ont par exemple travaillé des centaines d'années sur la question de savoir si le dernier théorème de Fermat est vrai ou faux. Dans le cas des machines de ce type, une sorte d'argument plus subtil est nécessaire. Par le fameux théorème de Gödel, ou par un argument similaire, on peut montrer que quelle que soit la manière dont une machine est construite, il y aura des cas où la machine échouera à donner une réponse, mais où un mathématicien pourra en donner une. D'un autre côté, la machine a certains avantages sur le mathématicien. On peut s'appuyer sur tout ce que la machine fait, si l'on suppose qu'il n'y aura pas de panne mécanique, tandis que le mathématicien commet parfois des erreurs, selon une certaine proportion. Je crois que ce danger du mathématicien faisant des erreurs est un corollaire inévitable de la possibilité qu'il utilise parfois de mettre en œuvre une méthode complètement nouvelle. Cela semble être confirmé par le fait bien connu que les personnes les plus fiables n'utilisent en général pas de méthodes vraiment nouvelles.

Ce que je prétends, c'est que des machines peuvent être construites qui simuleront la

Philosophia Mathematica (3) Vol. 4 (1996), pp. 256-260.
http://www.turingarchive.org/browse.php/B/4

Résumé. Dans cet essai posthume, Turing prétend qu'il pourrait être possible de construire une machine qui contiendrait un composant aléatoire et un analogue au principe de plaisir en psychologie, à qui on pourrait apprendre, et qui pourrait finalement devenir plus intelligente que les humains.

pensée humaine de façon très approchée. Parfois elles feront des erreurs, et parfois elles feront de nouvelles assertions très intéressantes, et globalement, les réponses qu'elles fourniront en sortie seront à première vue quasiment les mêmes réponses que celles fournies par un cerveau humain. Le contenu de mon assertion réside dans la grande fréquence attendue d'assertions vraies, et elle ne peut pas, je pense, être reçue comme une assertion vraie. Il ne pourrait pas, par exemple, être suffisant pour dire simplement qu'une machine exprimera une assertion vraie un jour ou l'autre, car un exemple d'une telle machine serait celui d'une machine qui exprime toutes les assertions un jour ou l'autre. Nous savons comment les construire, et comme elles devraient produire (probablement) des assertions vraies et des assertions fausses à peu près aussi fréquemment les unes que les autres, leurs verdicts seraient bien pire. Ce serait la réaction réelle de la machine aux circonstances qui prouverait ce que je prétends, si tant est qu'elle puisse être prouvée.

Voyons plus précisément la nature de cette argumentation. Il est clairement possible de fabriquer une machine qui fournirait un compte-rendu très précis à propos d'elle-même pour n'importe quel jeu de tests, si cette machine était suffisamment élaborée. Pourtant, ceci ne pourrait à nouveau que très difficilement être considéré comme une preuve adéquate. Une telle machine finirait par se perdre en faisant toujours la même sorte d'erreur encore et encore, et en étant quasiment incapable de se corriger elle-même, ou en étant corrigée par des arguments fournis de l'extérieur. Si la machine était capable d'une certaine manière d'"apprendre par expérience", ce serait beaucoup plus impressionnant. Si c'était le cas, il semblerait qu'il n'y ait aucune raison réelle pour que quelqu'un n'essaie pas de commencer avec une machine comparativement simple, et, en la soumettant à un certain nombre d'expériences adéquates, à la transformer en une autre machine qui serait plus élaborée, et qui serait capable de gérer un nombre plus élevé de contingences. Ce processus serait probablement accéléré par une sélection appropriée des expériences auxquelles la machine serait soumise. On pourrait appeler ce processus l'"éducation". Mais ici, nous devons être prudents. Ce serait assez facile d'arranger des expériences de telle manière qu'elles fassent que la structure de la machine se retrouve dans une forme particulière, et cela serait de façon évidente une grosse manière de tricher, presque autant que d'avoir un homme caché dans la machine. A nouveau ici, le critère exprimant ce qui devrait être considéré comme raisonnable en termes d'"éducation" ne peut pas être mis en termes mathématiques, mais je suggère que ce qui suit pourrait être considéré comme adéquat en pratique. Supposons que l'on souhaite que la machine comprenne l'anglais, et qu'elle n'ait ni mains ni pieds, et aucun besoin de se nourrir, aucun désir de cigarette, elle occupera son temps principalement à jouer à des jeux tels que les échecs ou le Go, et peut-être le Bridge. La machine est équipée d'un clavier de machine à écrire sur lequel on peut taper toute remarque qu'on souhaite lui faire, et elle écrit en retour toute chose qu'elle veut dire. Je suggère que l'éducation de la machine soit confiée à un maître d'école très compétent qui est intéressé par le projet mais à qui l'on n'a communiqué aucune connaissance détaillée du fonctionnement interne de la machine. Le mécanicien qui a construit la machine, par contre, doit la maintenir en état de marche, et s'il suspecte que la machine n'a pas effectué ses tâches correctement, il a le doit de la remettre dans un état antérieur et de demander au professeur de répéter sa leçon à partir de cet endroit-là, mais il n'a pas le droit de prendre part en aucune manière au processus d'enseignement. Puisque cette procédure est uniquement destiné à tester la *bonne foi* de la mécanique, il faut que j'insiste sur le fait qu'elle ne pourrait pas être adoptée dans les étapes expérimentales. Comme je le vois, le processus d'enseignement serait en pratique essentiel à l'obtention d'une machine raisonnablement intelligente dans un intervalle de temps raisonnablement court. L'analogie humaine suggère cela.

Je peux maintenant donner quelques indications sur la manière de fonctionner que l'on peut attendre d'une telle machine. La machine devrait incorporer une mémoire. Cela ne nécessite pas beaucoup d'explication. Elle devrait consister en une liste de toutes les assertions qui existent pour elle ou bien qu'elle a faites, et de tous les mouvements qu'elle a faits, et de toutes les cartes qu'elle a jouées dans les jeux auxquels elle a participé. Tout ça sera listé dans l'ordre chronologique. Outre cette mémoire évidente, il y aura un certain nombre d'"index d'expériences". Pour expliquer cette idée, je suggèrerai la forme qu'un tel index pourrait prendre. Cela pourrait être un classement en ordre alphabétique des mots qui ont été utilisés en donnant l'"instant" auquel ils ont été utilisés, de telle manière qu'on puisse les retrouver dans la mémoire. Un autre tel index pourrait contenir les images des humains ou des parties des gobans qui auront été rencontrées. Aux périodes plus tardives de l'éducation, la machine pourrait être étendue pour inclure d'importantes parties de la configuration de la machine à tout moment, ou en d'autres termes, elle pourrait se rappeler quelles ont été les pensées qu'elle a eues. Cela donnerait naissance à de nouvelles formes d'indexation très fructueuses. Les nouvelles formes d'indexation pourraient être introduites en tenant compte de motifs spéciaux observés dans les index déjà utilisés. Les index seraient utilisés de la façon suivante : à chaque fois qu'un choix devrait être fait sur l'action à effectuer ensuite, les motifs de la situation présente seraient observés dans les index disponibles, et le choix précédent qui aurait été fait dans des situations similaires, et la sortie, bonne ou mauvaise, qui en aurait découlé seraient retrouvés.

Le nouveau choix sera fait selon toutes ces données. Cela entraîne un certain nombre de problèmes. Si certaines indications sont favorables et d'autres non favorables, que faire ? La réponse à cela diffèrera probablement d'une machine à l'autre et variera également en fonction du degré d'éducation de la machine. Initialement, probablement que quelques règles assez succinctes suffiront, e.g. faire l'action qui a le plus de votes en sa faveur. A un stade bien plus tardif de l'enseignement, la question entière de la procédure dans de tels cas aura probablement été étudiée par la machine elle-même, par les moyens d'une sorte d'index, et cela pourra résulter en quelque chose de beaucoup plus sophistiqué et, on espère, à des formes de règles bien plus satisfaisantes. Il semble probable pourtant que les formes comparativement brutes des règles seront elles-mêmes raisonnablement satisfaisantes, de telle manière qu'un progrès global puisse être fait malgré le manque de précision du choix de règles. Cela semble être vérifié par le fait que les problèmes d'ingénierie sont parfois résolus de la manière la plus rustre par des procédures ad-hoc qui ne gèrent que les aspects les plus superficiels du problème, e.g. si une fonction croît ou décroît en l'une de ses variables. Un autre problème soulevé par cette image de la manière dont le comportement est déterminé est l'idée d'une "sortie favorable". Sans une telle idée, correspondant au "principe de plaisir" des psychologues, il est très difficile de voir comment procéder. Certainement qu'il serait plus naturel d'introduire quelques petites choses comme celles ci-après en machine. Je suggère qu'il pourrait y avoir deux clefs manipulables par le maître, et qui représente les idées de plaisir et déplaisir. A des stades ultérieurs de l'apprentissage, la machine reconnaîtrait certaines autres conditions comme étant désirables compte tenu du fait qu'elles auront par le passé constamment été associées au plaisir, et inversement, un certain nombre d'autres choses sont non désirables. Les expressions de colère de la part du maître pourraient, par exemple, être reconnues comme si déplaisantes qu'elles seraient négligées, de telle façon que le maître trouverait qu'il n'est plus nécessaire d'"employer la punition".

Faire de plus amples suggestions à propos de ces lignes serait probablement sans effets à ce niveau, dans la mesure où elles ne consisteront vraisemblablement en rien de plus qu'en une analyse des méthodes actuelles d'éducation appliquée aux enfants humains. Il y a, pourtant, une fonctionnalité dont j'aimerais suggérer qu'elle soit incorporée dans les machines, et c'est le "composant aléatoire". On devrait doter chaque machine d'une bande magnétique contenant une série aléatoire de figures, e.g., des 0 et des 1 à quantités égales, et cette série de figures devrait être utilisée dans les choix effectués par la machine. Cela aura pour conséquence que le comportement de la machine ne sera pas complètement déterminé par les expériences auxquelles elle a été sujette, et aura des utilisations précieuses quand on aura expérimenté ces nouvelles idées. En simulant les choix effectués, quelqu'un pourrait être capable de contrôler le développement de la machine dans une certaine mesure. On pourrait, par exemple, insister sur le fait que le choix effectué doit être un choix particulier, disons, en 10 endroits particuliers, et cela signifierait qu'une machine sur 1024 devrait se développer à un degré au moins aussi haut que celui qui aurait été simulé. Cela aura du mal à devenir une assertion précise à cause de la nature subjective de l'idée de "degré de développement" en ne disant même rien sur le fait que la machine qui avait été simulée aurait pu être aussi chanceuse dans ses choix non simulés.

Supposons maintenant, pour conforter l'argument, que ces machines soient une véritable possibilité, et regardons les conséquences de leur construction. Les construire effectivement devrait recontrer une grande opposition, à moins que nous n'ayions grandement avancé en terme de tolérance religieuse depuis le temps de Galilée. Il y aurait alors une grande opposition des intellectuels, qui craindraient de perdre leur travail. Il est probable pourtant que les intellectuels se tromperaient à ce propos. Il y aurait beaucoup de choses à faire pour essayer, disons, de maintenir notre intelligence au niveau standard établi par les machines, car il semble probable qu'une fois que la méthode de pensée de la machine aura démarré, elle ne mettra pas longtemps à dépasser nos faibles possibilités. Il ne sera pas question de leur mort, et elles seront capables de converser les unes avec les autres pour aiguiser leurs esprits. Nous devons ainsi nous attendre à ce qu'à un moment, elles prennent le contrôle, comme cela est mentionné dans le livre de Samuel Butler *Erewhon*.

Les ordinateurs digitaux peuvent-ils penser ? (Alan M. Turing)

Les ordinateurs digitaux ont souvent été décrits comme des cerveaux mécaniques. La plupart des scientifiques regardent probablement ces descriptions comme de simples slogans journalistiques, mais d'autres non. Un mathématicien m'a exposé le point de vue opposé plutôt violemment en ces termes "On dit communément que ces machines ne sont pas des cerveaux mais vous et moi savons que c'en sont". Dans cet exposé, j'essaierai d'expliquer les idées derrière les différents points de vue possibles, mais je ne le ferai pas de façon impartiale. J'accorderai davantage d'attention au point de vue qui est le mien, qui est qu'il n'est pas déraisonnable de décrire les ordinateurs digitaux comme des cerveaux. Un point de vue différent a déjà été défendu par le Professeur Hartree.

D'abord, nous pouvons considérer le point de vue naïf de l'homme de la rue. Il entend des compte-rendus surprenants à propos de ce que ces machines peuvent faire : la plupart semblent avoir des capacités intellectuelles qu'il ne possède lui-même pas. Il ne peut l'expliquer qu'en supposant que la machine est une sorte de cerveau, même s'il préfère plutôt ne pas croire ce qu'il a entendu.

La majorité des scientifiques méprisent ces attitudes quasiment superstitieuse. Ils savent quelque-chose des principes sur lesquels ces machines sont construites et de la manière dont on les utilise. Leur organisation générale a été résumée par Lady Lovelace il y a une centaine d'années environ, lorsqu'elle a décrit le moteur analytique de Babbage. Elle dit, comme Hartree l'a citée "Le moteur analytique n'a aucune prétention à initier quoi que ce soit. Il peut faire tout ce qu'on lui ordonne de faire.". Ceci décrit très bien la manière dont les ordinateurs digitaux sont utilisés au jour d'aujourd'hui, et la manière dont ils seront principalement utilisés dans de nombreuses années à venir. Pour le moindre calcul, la totalité de la procédure que la machine va utiliser est planifiée à l'avance par un mathématicien. Moins il y a de doute sur ce qui va se produire, plus le mathématicien est content. C'est comme planifier une opération militaire. Dans ces conditions, on peut dire que la machine n'initie rien.

Il y a pourtant un troisième point de vue, qui est le mien. Je suis d'accord dans la mesure du possible avec l'énoncé de Lady Lovelace, mais je crois que sa validité dépend du fait de considérer la manière dont les ordinateurs digitaux sont utilisés plutôt que celle dont ils pourraient être utilisés. En fait, je crois qu'ils pourraient être utilisés de telle manière qu'on les décrirait adéquatement par le terme cerveaux. Je dirais également que "si une machine peut être décrite de manière appropriée comme un cerveau, alors tout ordinateur digital peut également être décrit ainsi".

---

15 Mai 1951.

Cette dernière phrase nécessite d'être expliquée. Cela peut sembler surprenant, mais avec quelques réserves, cela semble inévitable. On peut montrer que cela découle d'une propriété caractéristique des ordinateurs digitaux, que j'appellerai leur universalité. Un ordinateur digital est une machine universelle au sens où elle peut remplacer toute machine d'un ensemble très grand. Elle ne remplacera pas un bulldozer ou une machine à vapeur ou un télescope, mais elle remplacera toute autre machine à calculer, c'est-à-dire toute machine dans laquelle on entre des données et qui plus tard renvoie des résultats. Pour que notre ordinateur imite une machine donnée, il est seulement nécessaire de le programmer pour qu'il simule ce que la machine en question aurait fait dans telles circonstances, et en particulier quelles données en sortie elle aurait fournies. L'ordinateur peut ainsi être programmé pour fournir les mêmes réponses.

Si maintenant une machine particulière peut être décrite comme un cerveau, nous n'avons qu'à programmer notre ordinateur pour l'imiter et ce sera aussi un cerveau. Si l'on accepte que les cerveaux réels, comme ceux des animaux, et en particulier ceux des humains, sont des sortes de machines, il en découlera que nos ordinateurs digitaux convenablement programmés se comporteront comme des cerveaux.

Cet argument implique un certain nombre de suppositions qui peuvent être raisonnablement défiées. J'ai déjà expliqué que la machine à imiter doit ressembler plutôt à une calculatrice qu'à un bulldozer. C'est seulement une réflexion au sens où nous sommes en train de parler d'analogues mécaniques des cerveaux, plutôt que des jambes ou des mâchoires. Il est aussi nécessaire que cette machine soit d'un genre dont le comportement peut être prédit par le calcul. Nous ne savons certainement pas comment un tel calcul devrait être fait, et il a même été expliqué par Sir Arthur Eddington que du fait du principe d'incertitude de la mécanique quantique, une telle prédiction n'est même pas possible théoriquement.

Une autre supposition était que la capacité de la mémoire de l'ordinateur devrait être suffisante pour prédire le comportement de la machine à imiter. Il faudrait aussi disposer d'une vitesse de calcul suffisante. Nos ordinateurs actuels n'ont probablement pas assez d'espace mémoire, bien qu'il soit possible qu'ils aient la vitesse de traitement appropriée. Cela signifie en effet que si nous souhaitons imiter quelque-chose d'aussi compliqué que le cerveau humain, nous avons besoin d'une bien plus grande machine qu'aucun des ordinateurs dont nous disposons actuellement. Nous avons vraisemblablement besoin de quelque-chose qui soit au moins cent fois plus grand que l'ordinateur de Manchester. Alternativement bien sûr, une machine de taille égale ou plus petite pourrait aller si des progrès suffisants étaient faits en termes de stockage de l'information.

Il faudrait noter qu'il n'est pas nécessaire d'augmenter la complexité des ordinateurs utilisés. Si nous essayons d'imiter des machines encore plus compliquées ou des cerveaux, nous devons utiliser des ordinateurs de plus en plus gros pour le faire. Nous n'avons pas besoin d'en utiliser qui soient de plus en plus compliqués. Cela peut sembler paradoxal, mais l'explication n'en est pas difficile. L'imitation d'une machine par un ordinateur nécessite non seulement que nous ayons fabriqué l'ordinateur, mais également que nous l'ayons programmé de la façon appropriée. Plus la machine à imiter est compliquée, plus le programme doit l'être. Peut-être cela pourra-t-il être rendu plus clair par une analogie. Supposons que deux hommes veuillent écrire leur auto-biographie, et que l'un ait eu une vie pleine d'événements mais que très peu d'événements se soient produits dans la vie de l'autre. Il y aurait deux difficultés troublant l'homme avec une vie pleine, et qui le gêneraient plus que l'autre. Il lui faudrait plus de papier et il aurait plus de difficulté à décider de ce qu'il va écrire. Le fait de le pourvoir en papier ne semble pas être une difficulté sérieuse, à moins par exemple qu'il soit sur une île déserte, et dans tous les cas, ce ne serait qu'un problème technique et financier. L'autre difficulté serait plus fondamentale et deviendrait plus sérieuse encore si plutôt que d'écrire sa vie, il s'agissait d'effectuer un travail sur un sujet auquel il ne connaît rien, disons sur la vie sur Mars. Notre problème pour programmer un ordinateur pour qu'il se comporte comme un cerveau est quelque-chose qui ressemble à l'écriture de ce traité, et dans tous les cas, nous ne savons pas ce que nous devrions écrire si nous l'avions. C'est un piètre état de choses, mais, pour poursuivre l'analogie, il faudrait savoir quoi écrire, et apprécier le fait que la plupart des connaissances peuvent être incarnées dans les livres.

Au vu de cela, il semble que le plus sage étayage sur lequel critiquer la description des ordinateurs digitaux en tant que "cerveaux mécaniques" ou "cerveaux électroniques" est que, bien qu'ils puissent être programmés pour se comporter comme des cerveaux, nous ne savons pas à présent comment cela pourrait être fait. Avec cet argument, je suis en total accord. Il laisse ouverte la question de savoir si nous finirons par réussir ou pas à trouver un tel programme. Je pense par exemple probable qu'à la fin du siècle, il sera possible de programmer une machine qui répondra à des questions de telle manière qu'il sera extrêmement difficile de deviner si les réponses en sont données par un homme ou par une machine. J'imagine quelque-chose comme un examen de viva voce, mais avec des questions et réponses toutes tapées à la machine de manière à ce que nous n'ayons pas à considérer des éléments non pertinents comme la qualité avec laquelle la voix humaine peut être imitée. C'est seulement mon opinion ; il y a de la place pour de nombreuses autres opinions.

Il reste des difficultés. Se comporter comme un cerveau nécessite le libre-arbitre, mais le comportement d'un ordinateur digital, lorsqu'il a été programmé, est complètement déterministe. Ces deux faits devraient être réconciliés en quelque sorte, mais faire cela semble nous ramener à la controverse d'un autre âge du "libre-arbitre-et-déterminisme". On ne peut pas s'en sortir. Il est possible que l'impression de libre-arbitre que nous partageons tous ne soit qu'une illusion. Ou bien il est possible que nous ayons effectivement un libre-arbitre, mais qu'il soit impossible de dire qu'il en est ainsi uniquement en observant notre comportement de l'extérieur. Dans ce dernier cas, aussi réaliste que soit la manière dont une machine pourra imiter le comportement humain, elle ne pourra être considérée que comme un simulacre. Je ne sais pas comment nous pourrons jamais décider entre ces deux alternatives mais quelle que soit l'alternative correcte, il est sûr qu'une machine sensée imiter un cerveau doit sembler se comporter comme si elle avait un libre-arbitre, et c'est aussi bien de se demander comment on pourrait faire ça. Une des possibilités est que son comportement dépende de quelque-chose comme une roulette ou un atome de radium. Le comportement de ces dispositifs pourrait sembler pouvoir peut-être être prédit, mais si tel est le cas, nous ne savons pas comment faire cette prédiction.

Il n'est, cependant, pas vraiment nécessaire de faire ça. Il n'est pas difficile de concevoir des machines dont le comportement semble assez aléatoire à quiconque ne connaît aucun détail de leur construction. Naturellement, l'inclusion de cet élément aléatoire, quelle que soit la technique utilisée, ne résoud pas notre problème principal, qui est de savoir comment programmer une machine pour imiter un cerveau, ou, comme nous pourrions le dire plus brièvement, même si c'est moins précis, pour *penser*. Mais cela nous donne une indication sur ce à quoi pourrait ressembler le processus. Nous ne devons pas nous attendre à toujours savoir ce que l'ordinateur va faire. Nous devrions être content si la machine nous surprend, de la même façon que nous sommes content lorsqu'un élève fait quelque chose que nous ne lui avons pas explicitement demandé de faire.

Reconsidérons maintenant l'énoncé de Lady Lovelace : "La machine peut faire tout ce dont on sait comment lui apprendre à le faire". Le sens du reste du passage est que l'on est tenté de dire que la machine ne peut faire que ce dont on sait lui expliquer comment le faire. Mais je pense que cela peut ne pas être vrai. Certainement que la machine ne peut faire que ce que nous lui ordonnons, si elle faisait autre chose, cela proviendrait d'un problème mécanique. Mais il n'est pas nécessaire de supposer cela, quand nous lui donnons ses ordres, nous savons ce que nous faisons, ce que les conséquences de ces ordres vont être. On n'a pas besoin de comprendre comment ces ordres vont amener la machine à avoir le comportement associé, ni de comprendre le mécanisme de germination quand on plante des graines dans le sol. La plante poussera qu'on le comprenne ou pas. Si nous donnons à la machine un programme qui résultera dans le fait qu'elle fera une chose intéressante que nous n'avions pas anticipée, je serais enclin à dire que la machine a initié quelque-chose, plutôt que de dire que son comportement était implicite dans le programme, et qu'ainsi l'originalité n'est que de notre fait.

Je n'essaierai pas d'en dire beaucoup sur la manière dont ce processus consistant à "programmer une machine à penser" doit être fait. Le fait est que nous n'en savons que très peu à ce sujet, et que très peu de recherche a déjà été faite. Il y a de nombreuses idées, mais nous ne savons pas encore lesquelles sont importantes. Comme dans une histoire de détective, au début de l'enquête, tout détail peut être important pour l'enquêteur. Quand le problème a été résolu, seuls les faits essentiels doivent être racontés au jury. Mais à présent, nous n'avons rien à montrer devant un jury. Je dirai seulement cela, je crois que le processus sera très lié à celui de l'enseignement.

J'ai essayé d'expliquer quels sont les arguments rationnels principaux pour et contre la théorie selon laquelle on pourrait fabriquer des machines pensantes, mais il faut dire quelque chose à propos des arguments irrationnels. Beaucoup de personnes sont extrêmement opposées à l'idée qu'une machine puisse penser, mais je ne crois pas que ce soit pour aucune des raisons que j'ai mentionnées, ou pour une quelconque autre raison rationnelle, mais simplement parce qu'ils n'aiment pas cette idée. On peut voir de nombreuses caractéristiques qui rendent cette idée désagréable. Si une machine pense, elle pourrait penser plus intelligemment que nous ne le faisons, et alors où serions-nous ? Même si nous pouvons maintenir les machines dans une position d'esclaves, par exemple, en coupant leur alimentation à des moments stratégiques, nous pourrions nous sentir grandement humiliés en tant qu'espèce. Un danger et une humiliation similaire nous menacent lorsque nous envisageons la possibilité que nous puissions être remplacés par des cochons ou des rats. C'est une possibilité théorique qui est très controversée, mais nous avons vécu avec des cochons et des rats depuis si longtemps sans que leur intelligence n'augmente beaucoup, que nous ne sommes pas davantage troublés par cette possibilité. Nous ressentons que si ça devait arriver un jour, cela n'adviendrait pas avant quelques millions d'années. Alors que le nouveau danger semble plus proche. S'il advient, ce sera vraisemblablement au prochain millénaire. C'est dans un avenir lointain mais pas dans un avenir astronomiquement lointain, et c'est certainement quelque-chose qui peut nous rendre anxieux.

Il est coutumier, dans un exposé sur ce sujet, de fournir un peu de réconfort, en disant que quelques caractéristiques humaines particulières ne seront jamais imitées par une machine. On pourrait par exemple dire qu'aucune machine n'écrira jamais très bien l'anglais, ou qu'elle ne sera jamais attirée par le sexe ou ne fumera la pipe. Je ne peux pas fournir un tel argument réconfortant, car je crois qu'aucune telle limite ne peut être fixée. Mais j'espère certainement et je crois qu'aucun grand effort ne sera fait pour mettre en machine les caractéristiques les plus distinctives des humains, mais des caractéristiques non intellectuelles comme la forme du corps humain. Cela me semble assez stupide de faire de telles tentatives et leurs résultats ont quelque chose d'aussi déplaisant que lorsqu'on pense aux fleurs artificielles. Les tentatives de faire penser les machines me semblent d'un tout autre ordre. Le processus complet de la pensée humaine est encore plutôt mystérieux pour moi, mais je crois que la tentative de créer une machine pensante nous aidera grandement à trouver comment nous pensons nous-mêmes.