



## Histoires d'hommes-machines

Cassou-Noguès Pierre

### Pour citer cet article

Cassou-Noguès Pierre, « Histoires d'hommes-machines », *Cycnos*, vol. 22.1 (La science-fiction dans l'histoire, l'histoire dans la science-fiction), 2005, mis en ligne en octobre 2006.

<http://epi-revel.univ-cotedazur.fr/publication/item/644>

Lien vers la notice <http://epi-revel.univ-cotedazur.fr/publication/item/644>

Lien du document <http://epi-revel.univ-cotedazur.fr/cycnos/644.pdf>

### *Cycnos, études anglophones*

*revue électronique éditée sur épi-Revel à Nice*

ISSN 1765-3118      ISSN papier 0992-1893

### AVERTISSEMENT

*Les publications déposées sur la plate-forme épi-revel sont protégées par les dispositions générales du Code de la propriété intellectuelle. Conditions d'utilisation : respect du droit d'auteur et de la propriété intellectuelle.*

*L'accès aux références bibliographiques, au texte intégral, aux outils de recherche, au feuilletage de l'ensemble des revues est libre, cependant article, recension et autre contribution sont couvertes par le droit d'auteur et sont la propriété de leurs auteurs. Les utilisateurs doivent toujours associer à toute unité documentaire les éléments bibliographiques permettant de l'identifier correctement, notamment toujours faire mention du nom de l'auteur, du titre de l'article, de la revue et du site épi-revel. Ces mentions apparaissent sur la page de garde des documents sauvegardés ou imprimés par les utilisateurs. L'université Côte d'Azur est l'éditeur du portail épi-revel et à ce titre détient la propriété intellectuelle et les droits d'exploitation du site. L'exploitation du site à des fins commerciales ou publicitaires est interdite ainsi que toute diffusion massive du contenu ou modification des données sans l'accord des auteurs et de l'équipe d'épi-revel.*

# EPI-REVEL

Revue électronique de l'Université Côte d'Azur

## Histoires d'hommes-machines

Pierre Cassou-Noguès

**Pierre Cassou-Noguès** est chercheur au CNRS (UMR 8519) et enseigne à l'université Lille III. Il a travaillé en philosophie des sciences et s'intéresse actuellement aux rapports entre science et littérature autour de thèmes-phares, tel que celui de la machine. Il a notamment publié *De l'expérience mathématique* (Paris : Vrin, 2001) et *Gödel* (Paris : Les belles lettres, 2004)

Le but de cet article est d'identifier et de distinguer les deux images de l'homme-machine : l'une que nous rencontrons au XVII<sup>ème</sup> siècle dans les textes de Descartes en particulier, l'autre qui s'élabore au XIX<sup>ème</sup> siècle et que l'on retrouve dans la science-fiction. Il s'agit de confronter à Descartes, le texte de Philip K. Dick, *Do Androids Dream of Electric Ships ?* La science-fiction a contribué à déplacer la question de la différence entre l'homme et la machine.

robot, androïde, Doyle, Dick, Descartes, émotion, langage, parole

Le but de cet article est de distinguer deux images de l'homme-machine, l'une qui joue dans l'imaginaire classique et qui est fixée dans les textes de Descartes, l'autre qui est au centre de notre imaginaire depuis la fin du XIX<sup>ème</sup> siècle et que l'on rencontre en particulier dans la science-fiction.<sup>1</sup> Par homme-machine, j'entends l'homme représenté par une machine ou une machine qui imite autant que possible le comportement humain. Dans le vocabulaire de la science-fiction, l'homme-machine est un robot, un androïde. Or je voudrais montrer que le texte cartésien donne à l'homme-machine des caractères exactement contraires à ceux que prend le robot dans notre propre imaginaire. Avant tout, le paradigme de la machine, pour Descartes et dans l'imaginaire classique, est une horloge, c'est-à-dire un automate qui, une fois mis en branle possède en lui-même le principe de son mouvement. En revanche, le modèle de machine auquel nous pensons avec le robot est celui de l'ordinateur, de la machine à calculer. Sans doute, un robot n'est pas seulement un ordinateur mais son fonctionnement dépend de façon essentielle du dispositif calculatoire, de l'ordinateur, qui constitue son « cerveau ». Les progrès de la robotique, dans la science-fiction, sont presque toujours liés à une transformation de l'ordinateur central, du cerveau de la machine. On peut s'intéresser - c'est vrai - mais seulement de façon accessoire à l'appareil digestif du robot.<sup>2</sup> Le robot, pour nous, est d'abord un cerveau. D'autre part et surtout, les machines classiques, les machines telles que les conçoit Descartes, sont susceptibles de manifester toutes sortes d'émotions, la joie, la tristesse, la colère ou l'amour. En revanche, Descartes leur refuse la parole. C'est à cela qu'on les distingue des êtres humains : les robots, pour Descartes, ne peuvent pas user du langage. Les robots de la science-fiction sont alors à l'opposé des robots cartésiens. Nos robots imaginaires sont capables de parler et, du point de vue de la parole, rien ne les distingue d'un être humain. Mais ils ne peuvent pas exprimer d'émotion. Ils sont impassibles et, comme le formule Ph. K. Dick, dépourvus d'empathie.

---

<sup>1</sup> Je développe cette thèse dans *Une histoire de machines, de vampires et de fous*, Paris : Vrin, à paraître (2006), et, sous un autre angle, dans « Sherlock Holmes, l'homme des preuves », dans V. Ando (éd.), *La prova*, Palerme : Carroci, à paraître (automne 2005).

<sup>2</sup> Notamment, Asimov, dans *The Caves of Steel* (1954), qui insiste sur la capacité digestive de son robot, Daniel R. Olivaw. Mais les quelques exemples que l'on peut trouver ne contrebalancent pas l'intérêt et l'abondante littérature que suscite au XVIII<sup>ème</sup> siècle le canard de Vaucanson avec son mécanisme digestif (cf. Seris, *Langages et machines à l'âge classique*, Paris : Hachette, 1995).

Nous passons donc d'une image, où la machine est une horloge muette (ou une horloge qui parle peu) mais dont le cadran peut exprimer l'amour, à l'image d'une machine à calculer, froide et impassible, mais qui raisonne, parle comme un être humain.

De l'homme-machine, du robot, je ne prends en compte ici que la machine, une machine qui, autant que possible, imite l'homme et constitue un individu. Il y a dans l'imaginaire de la science-fiction, des machines qui, comme celles de Descartes, ont part à quelque chose comme une émotion. Ce sont les machines qui forment le cyborg. Le cyborg, on le sait, est un être mixte, qui comporte une partie organique, biologique, et une partie mécanique. Or, dans tout un corpus de texte, qui présente donc une autre image de la machine, ce couplage du mécanique et du biologique donne lieu à un désir, que l'on ne peut attribuer en propre ni aux machines ni au noyau biologique du cyborg. La machine, dans le cyborg, est liée au désir, qu'elle en soit l'instrument ou, sinon la source, le moteur. En tout cas, elle n'a plus le visage impassible, froid ou indifférent, du robot. Le cyborg offre donc une autre image de la machine, différente du robot, et à laquelle on pourrait sans doute confronter le concept de machine désirante qu'introduisent Deleuze et Guattari en le référant du reste au récit de S. Butler *Erewhon*, presque de la science-fiction.

Par ailleurs, lorsque j'oppose les machines cartésiennes, qui restent muettes, à nos robots, qui ont gagné la parole mais ont perdu l'émotion, je ne veux pas dire, évidemment, que l'on a réussi depuis le XVII<sup>e</sup> siècle à construire des machines qui parlent ou que l'on a prouvé qu'une machine, comme un ordinateur, peut en principe utiliser le langage naturel. Je ne m'intéresse qu'aux images de la machine. Par image, j'entends une description de la machine qui prend une portée, qui, si l'on veut, nous convainc ou nous intrigue, sans forcément que les caractéristiques qui lui sont données aient été justifiées par un discours rationnel ou aient fait l'objet d'une preuve. La portée d'une image est indépendante de sa rationalité. Ainsi, dans notre imaginaire et en particulier dans la science-fiction, les robots parlent ou peuvent parler, alors que, du point de vue de la science ou du savoir au sens large, du point de vue de ce que l'on veut prouver, il n'est nullement clair qu'une machine puisse utiliser le langage naturel, et la philosophie, les sciences cognitives, l'informatique continuent d'en discuter. Notre image, le robot qui parle, s'est donc fixée d'elle-même sans avoir de base scientifique. En fait, il faut reconnaître une indépendance relative, ou une indépendance interne, de l'imaginaire par rapport aux sciences et aux savoirs (j'entends savoir au sens large d'un discours supposé faire l'objet d'une justification rationnelle). Le passage du paradigme de l'horloge à celui de la machine à calculer, ou de l'ordinateur, est bien lié au progrès des sciences et des techniques. Mais la thèse selon laquelle les machines parlent, thèse qui est implicite dans notre image de la machine, est posée et vaut d'elle-même, indépendamment de toute justification scientifique ou malgré une possible réfutation par les sciences. Par conséquent, notre imaginaire dépend d'un certain contexte, scientifique et technique, mais reste indépendant des thèses, isolées, locales en quelque sorte, qu'une science, ou un savoir, peut prononcer.

Je commencerai par évoquer les machines de Descartes pour les confronter ensuite à l'image de la machine que l'on rencontre dans la science-fiction. Il me faudra faire un détour par Baker Street où l'on verra comment le Docteur Watson observant son ami Holmes fixe cette image de la machine. Puis je suivrai plus longuement un chasseur d'androïdes dans un roman de Ph. K. Dick, qui met en scène nos machines parlantes mais a-pathiques tout en conservant un contexte cartésien.

## La machine cartésienne

Par souci de brièveté, je me contenterai de rappeler quelques thèses et une expérience cartésiennes.

- *Thèse 1.* Les animaux sont des machines.

- *Thèse 2.* Descartes ne nie pas que les animaux puissent nous signifier quelque chose comme des émotions : ce qu'il appelle des appétits naturels, comme la faim ou la soif, et des passions, comme la joie, la tristesse, la peur, la colère ou l'amour. Descartes se fait fort de rendre compte mécaniquement de l'expression de ses émotions. Les machines cartésiennes sont donc capables de donner les signes extérieurs de la sensibilité, c'est-à-dire de l'émotion : « Je ne refuse la vie à aucun animal, car je crois qu'elle consiste dans la seule chaleur du cœur; je ne lui refuse même pas la sensibilité, dans la mesure où elle dépend d'un organe corporel. »<sup>3</sup>
- *Thèse 3.* Descartes admet également que le corps humain fonctionne comme une machine. L'homme est alors l'union d'une âme, substance pensante et d'un corps-machine. Cette union est, reconnaît Descartes, impensable.
- *Thèse 4.* Si l'homme a un corps de machine, Descartes se demande comment distinguer un homme véritable, avec un corps de machine et une âme, d'une simple machine qui se présenterait devant nous. Le critère de cette distinction est la parole. Les machines, pour Descartes, ne pourront jamais utiliser le langage humain. La parole fait le propre de l'homme.

Mais Descartes sait bien que les perroquets (qui pour lui sont des machines) peuvent dire quelques mots. Il faut donc préciser cette thèse. La parole proprement humaine, celle que les machines ne pourront pas reproduire, a deux caractéristiques négatives. D'une part, elle ne doit pas être le signe extérieur d'une émotion immédiate (comme le cri de joie ou le « J'ai faim », auquel, pour Descartes, se réduit le langage du perroquet). D'autre part, elle peut ne pas être raisonnable. Cela inclut dans la parole humaine le discours de l'insensé et en exclut le raisonnement logique dont serait peut-être capable la machine à calculer:

Il n'y a aucune de nos actions extérieures, qui puisse assurer ceux qui les examinent, que notre corps n'est pas seulement une machine qui se remue de soi-même, mais qu'il y a aussi en lui une âme qui a des pensées, excepté les paroles, ou autres signes faits à propos des sujets qui se présentent, sans se rapporter à aucune passion. Je dis [...] que ces signes soient à propos, pour exclure le parler des perroquets, sans exclure celui des fous, qui ne laisse pas d'être à propos des sujets qui se présentent bien qu'il ne suive pas la raison; et j'ajoute que ces paroles ou signes ne se doivent rapporter à aucune passion, pour exclure non seulement les cris de joies ou de tristesse, et semblables, mais aussi tout ce qui peut être enseigné par artifice aux animaux.<sup>4</sup>

*Expérience.* Descartes, en effet, est conduit à des expériences de pensée qui font déjà figure de science-fiction. Dans une lettre à un correspondant anonyme, le philosophe imagine par exemple un enfant élevé en un lieu où il n'a jamais vu d'animaux. Il n'y a là avec cet enfant que des hommes et des automates, des automates qui imitent les animaux terrestres mais aussi des robots qui sont de l'extérieur presque humains. Toutefois, notre enfant connaît le critère cartésien, qui lui permet de distinguer les robots des hommes, la parole. Le visage parle, c'est un homme. Il reste muet, c'est un robot. L'enfant est ensuite reconduit dans nos pays, il voit alors, pour la première fois, des véritables animaux et, tout naturellement, il leur applique le critère qu'il a toujours utilisé pour reconnaître les créatures mécaniques. Ces animaux terrestres ne parlent pas, ce sont donc des machines - se dit l'enfant.

Avec cette histoire, Descartes entend montrer que notre réticence à penser les animaux comme des machines n'est qu'un préjugé qui dépend de notre éducation et qu'une autre éducation, dans d'autres circonstances, pourrait éliminer. La situation qu'imagine Descartes ressemble une histoire de science-fiction. Il suffirait d'ajouter que l'enfant est né dans une station orbitale, disons XR18, où il n'a connu que ses parents, humains, et les robots qui travaillaient pour eux. On lui offrait parfois quelques automates, comme des jouets. Mais l'enfant revient sur terre. Il voit un animal, qu'en pense-t-il ? Est-ce un autre automate, plus

<sup>3</sup> Descartes à Morus, 5 février 1649, Paris : Gallimard (Pléiade), p. 1320.

<sup>4</sup> Descartes au marquis de Newcastle, 23 novembre 1646, *id.*, p. 1255.

perfectionné que ceux avec lesquels il jouait ? Ou bien est-ce que l'enfant reconnaît dans l'animal un être vivant, comme lui-même et hétérogène à la mécanique.

Il me semble que l'image de la machine qui parcourt la science-fiction conduirait à une autre conclusion que celle de Descartes. Il sera beaucoup plus facile, je crois, de poser que les automates que l'on offre à l'enfant dans la station orbitale parlent et que, pourtant, l'enfant, en rencontrant des animaux terrestres, y décèle quelque chose dont l'automate reste dépourvu, quelque chose comme l'émotion. Nous verrons plus exactement comment Dick reprend l'histoire de Descartes. Mais disons dès maintenant que c'est à contre-emploi. La machine, dans l'imaginaire de la science-fiction, peut utiliser le langage mais elle reste impassible, elle est a-pathique.

L'image cartésienne de l'homme-machine, comme horloge muette, reste prégnante au XVII<sup>ème</sup> et XVIII<sup>ème</sup> siècles. On en trouve un écho dans des textes littéraires et, pour prendre l'exemple le plus célèbre, dans *Les Voyages de Gulliver*. Lorsque Gulliver est amené à la cour de *Brobdingnag*, chez les géants, le roi commence à considérer Gulliver comme « une pièce d'horlogerie qu'aurait conçu un artiste ingénieux ». Et c'est seulement au moment où Gulliver parle, d'une parole sensée, que le roi est pris d'un doute. Le roi de *Brobdingnag*, dont Gulliver précise qu'il a étudié la philosophie et les mathématiques, utilise bien le critère cartésien, la parole, pour distinguer l'homme et la machine.<sup>5</sup>

## Sherlock Holmes, un nouvel homme-machine

Je voudrais maintenant considérer l'image de la machine dans notre propre imaginaire. L'une de ses premières incarnations se trouve dans les textes de Conan Doyle. Sherlock Holmes, dans ses moments de réflexions du moins, est en effet l'un des premiers robots modernes. C'est dans leur deuxième aventure, *Le Signe des quatre*, que Watson découvre l'être véritable de son ami. Miss Morstan, qui deviendra l'épouse de Watson, vient de quitter le 221 B Baker Street après avoir exposé son problème au détective. Watson, qui assistait à la scène, ne l'a pas écoutée. Il est resté fasciné par la beauté de la jeune femme, beauté, que Holmes, lui, n'a pas « remarquée » (« *observed* » dit l'anglais). Watson comprend alors la vérité : « Vous êtes vraiment un automate, une machine à calculer [...] il y a parfois quelque chose de positivement inhumain en vous. »<sup>6</sup>

Cette remarque de Watson n'est pas isolée. Le médecin décrit Holmes « comme la plus parfaite machine à raisonner et à observer que le monde ait connue. »<sup>7</sup> Il note de façon précise les moments où Holmes, commençant à réfléchir à un problème, prend « cette contenance [impassible] qui faisait que beaucoup le considéraient comme une machine plutôt qu'un homme. »<sup>8</sup> Et les moments où Holmes redevient humain, lorsqu'il se laisse gagner par un sentiment, le plus souvent la vanité : « C'était en effet à ces moments qu'il cessait d'être une machine à raisonner et trahissait son amour tout humain pour l'admiration et les applaudissements. »<sup>9</sup>

---

<sup>5</sup> « The King, although he be as learned a person as any in his dominions, and had been educated in the study of philosophy, and particularly mathematics; yet when he observed my shape exactly, and saw me walk erect, before I began to speak, conceived I might be a piece of clockwork, (which is in that country arrived to a very great perfection) contrived by some ingenious artist. But when he heard my voice, and found what I delivered to be regular and rational, he could not conceal his astonishment » (Swift, *Gulliver's Travels*, Londres : Penguin Popular Classics, 1994, p. 106). Cf. aussi Seris, *op. cit.*; Rosenfield, *From Beast-Machine to Man-Machine*, New-York : Oxford Univ. Press, 1941.

<sup>6</sup> *The Sign of Four*, in Doyle, *The complete Sherlock Holmes*, Londres : Penguin, 1981, p. 96.

<sup>7</sup> *The Adventures of Sherlock Holmes*, « A scandal in Bohemia » in Doyle, *op. cit.*, p. 161.

<sup>8</sup> *Memoirs of Sherlock Holmes*, « The crooked man », in Doyle, *op. cit.*, p. 412

<sup>9</sup> *The return of Sherlock Holmes*, « The adventures of the six napoleons », in Doyle, *op. cit.*, p. 593.

On voit d'abord dans ces remarques de Watson que le paradigme de la machine a changé. L'homme-machine n'est plus l'horloge mais déjà cette machine à calculer, que nous appellerons bientôt un ordinateur. C'est l'une des premières fois qu'un être humain est identifié à une machine à calculer. D'autre part, ce à quoi Watson reconnaît la machine en Holmes est son incapacité à manifester une émotion et, plus largement même, son incapacité à saisir ce qui échappe à la description objective, ce qui ne se laisse pas « observer ». Watson de fait caractérise la machine par son caractère impassible. La différence entre l'homme et la machine, le propre de l'homme, est dans l'émotion. Une machine peut parler, raisonner comme un homme : elle ne donne aucun signe d'émotion.

On pourrait ici parler d'une thèse de Watson par analogie avec la thèse et le test de Turing (qu'il est impossible de discuter ici). Les machines, si elles peuvent parler, sont incapables d'exprimer l'émotion. On reconnaîtra donc la machine à son caractère a-pathique. Mais il faut ajouter que cette thèse de Watson n'est pas sans conséquence. Elle suppose d'abord que les signes de l'émotion n'ont pas de description objective. Par exemple, que l'on ne peut pas identifier la honte à une certaine rougeur avec une définition exacte (c'est-à-dire la peau sous une lumière de telle intensité rendra une couleur de telle longueur d'onde). Si les signes de l'émotion se laissaient définir de cette façon, avec exactitude, il serait en principe possible de les introduire sur des machines. En principe, dans une technologie suffisamment avancée, on pourrait imaginer une machine avec un visage presque humain et qui serait capable, dans certaines circonstances, de le colorer de ce rouge de la honte. A la limite, la machine pourrait être incapable d'interpréter les situations humaines et se tromper dans l'évaluation des circonstances qui appellent la honte. Elle colorerait son visage à contre-temps mais, si la honte est strictement signifiée par un certain rouge, cette couleur sur le visage de la machine serait lue, avec surprise peut-être mais sans ambiguïté, comme de la honte. La thèse de Watson implique que les signes de l'émotion, que la machine ne peut pas reproduire, n'ont pas de caractérisation exacte. Les émotions, sur le visage humain, ne se reconnaissent donc pas à des signes objectifs mais à une sorte de « je ne sais quoi ». Ce « je ne sais quoi » qui exprime l'émotion, s'il ne connaît pas de description exacte, ne pourra pas être imité par la machine mais ne pourra pas non plus être reconnu par la machine. La machine ne peut pas plus exprimer des émotions qu'elle ne peut les reconnaître sur le visage d'autrui. La conséquence, enfin, est que, si l'émotion fait le propre de l'homme, et le distingue de la machine, la machine, qui ne reconnaît pas l'émotion, ne pourra pas se distinguer ou distinguer une autre machine d'un être humain. La machine ne saura donc pas d'elle-même ou ne saura qu'abstraitement qu'il y a des êtres qui ne sont pas des machines. J'y reviendrai.

On pourrait discuter de la thèse de Watson et de ses conséquences. Mais il ne s'agit pas ici d'évaluer leur vérité. La thèse de Watson est peut-être fautive. Le point qui m'intéresse, est que, même fautive, cette thèse joue dans notre imaginaire: elle détermine notre image de la machine. Je voudrais, dans cette perspective, suivre le roman de Ph. K. Dick, de 1968, *Do Androids Dream of Electric Sheep ?*

## Les machines de Dick

Nous sommes donc à San Francisco, au milieu du XXI<sup>ème</sup> siècle. Les hommes construisent des androïdes pour les aider dans la colonisation des planètes du système solaire. Mais il arrive que les androïdes s'échappent et reviennent sur Terre. Ils sont alors recherchés, poursuivis et, lorsqu'ils sont découverts, systématiquement détruits, « mis à la retraite », dit-on, plutôt que « tués ». Le roman raconte la journée d'un chasseur d'androïdes, Rick Deckard.

Rick Deckard, René Descartes. Xavier Mauméjean l'a déjà remarqué, ce n'est pas une coïncidence.<sup>10</sup> Le chasseur d'androïde est-il l'héritier du philosophe ? Du moins, Deckard doit bien être un peu philosophe. Il lui faut d'abord repérer les androïdes et, pour cela, apprendre à les distinguer des humains. Deckard est donc confronté à ce problème de savoir ce qui distingue l'homme de la machine et, par conséquent, ce qui fait le propre de l'homme. Seulement, pour Deckard et à la différence de Descartes, le propre de l'homme n'est plus dans la parole mais dans l'émotion ou, plus exactement, dans l'empathie.

Dick présente une image de la machine opposée à celle de Descartes tout en conservant un contexte cartésien. Les androïdes, et leur capacité à imiter l'être humain, sont déterminés par le type de leur cerveau. Leurs fonctions corporelles, la digestion, la circulation du sang, ne sont pratiquement pas évoquées. L'androïde est donc essentiellement son cerveau, un ordinateur. D'autre part et surtout, à l'opposé des machines cartésiennes, les androïdes de Dick ont gagné la parole mais perdu l'émotion. Les androïdes ont une intelligence souvent supérieure à celle des hommes.<sup>11</sup> Ils peuvent user du langage, et non seulement d'un langage rationnel mais de l'éventail entier de la parole humaine. Comme le voulait déjà Turing, les androïdes savent mentir et, en réalité, passent leur vie à mentir puisque, comme dans le test de Turing, ils doivent cacher leur identité (lorsqu'ils connaissent leur identité d'androïde). Les androïdes peuvent également plaisanter, faire de l'esprit. La télévision américaine utilise même des androïdes pour présenter ses émissions de variété. L'esprit de leur conversation est inégalable. Buster Friendly et ses invités « ne se répètent jamais... Leurs remarques toujours spirituelles (« *witty* »), toujours nouvelles, ne sont pas préparées. »<sup>12</sup>

Cette parole, cet esprit, est précisément celui que Descartes refusait à la machine et dont il faisait le propre de l'homme. On peut voir dans ces remarques « spirituelles », dans ces mots d'esprit l'exemple même d'une parole, qui, comme celle qu'évoquait Descartes, prend un sens alors qu'elle échappe aussi bien à l'émotion qu'à la logique. Le mot d'esprit est une parole qui n'est ni le simple signe d'une émotion immédiate, comme le cri de joie, ni ne respecte le raisonnement logique ou, pour reprendre l'expression de Descartes, les règles de la raison. Or, ici, les androïdes, avec leurs remarques qui ne sont pas préparées, semblent bien être capables de développer cet esprit à l'infini.

Les androïdes, dans le roman de Dick, ont donc appris à user de la parole humaine. En revanche, ils ont perdu toute capacité à manifester une émotion. C'est le thème sur lequel joue le roman. Les androïdes sont « froids » :

Il émane d'eux quelque chose d'autre. Quelque chose d'étrange. Et, pensait-il, de déplorable. Une froideur. Comme, se disait-il, le souffle du vide entre les mondes habités, un souffle de nulle part en fait.<sup>13</sup>

Ce thème de la froideur de la machine est omniprésent dans les textes de Dick. La froideur, véritablement, définit ce que c'est qu'une machine :

Dans notre univers, il existe des choses froides, cruelles, auxquelles je donne le nom de machines. Leur comportement m'effraie surtout lorsqu'il imite le comportement humain [...]. Je les appelle alors des androïdes. [...] Leur poignée de main est une étreinte avec la mort, et leur sourire a la froideur de la tombe.<sup>14</sup>

Dans ce roman de 1968, l'accent est mis sur une émotion particulière, « l'empathie », c'est-à-dire la sympathie avec les être vivants, la possibilité de souffrir de leur souffrance ou de participer à leur joie. C'est à l'absence d'empathie, l'a-pathie si l'on veut, que les androïdes

---

<sup>10</sup> Xavier Mauméjean, « Je pense donc je flippe », *Science-fiction Magazine*, hors série n° 8, septembre 2002.

<sup>11</sup> Dick, *Do Androids Dream of Electric Sheep?*, New-York : Millennium, 1999, p. 143

<sup>12</sup> *ibid* p.64: « They never repeated themselves ... Their remarks, always witty, always new, weren't rehearsed. »

<sup>13</sup> *ibid* p.58: « Something else had begun to emerge from her. Something more strange. And, he thought, deplorable. A coldness. Like, he thought, a breath from the vacuum between inhabited worlds, in fact from nowhere ».

<sup>14</sup> Dick, *The Shifting Realities of Ph. K. Dick*, New-York : Vintage Books, 1995, p. 211.

sont détectés. Leur corps est de l'extérieur identique à celui des humains, et il faut pour identifier les androïdes les soumettre au test de « Voigt-Kampff ». Ce test aurait pu porter le nom de celui qui reste le plus célèbre médecin britannique. Bien que Dick ne l'indique pas, le test de Voigt-Kampff repose sur la thèse du docteur Watson : il est fondé sur l'émotion. Il consiste à analyser les réactions d'un patient (la rougeur du visage, la dilatation des pupilles), en même temps qu'on lui propose une série de questions mettant en scène différents comportements, supposés cruels, vis-à-vis des animaux. Un être humain sera tantôt effrayé, tantôt révolté, alors qu'un androïde, bien qu'essayant de feindre, ne manifesterait pas d'émotion véritable. Comme la parole pour Descartes, l'empathie, telle qu'elle est révélée par le test de « Voigt-Kampff », permet donc de distinguer les hommes-machines des êtres humains :

Un androïde, quelque capacité intellectuelle qu'il eut, ne pouvait rien comprendre à cette fusion [...] à cette expérience que lui [Deckard] et à peu près tout le monde y compris les idiots sous-normaux [« subnormal chickenheads »] faisaient sans difficulté.<sup>15</sup>

La phrase de Dick, avec la référence à ces « idiots sous-normaux », répond encore parfaitement aux énoncés de Descartes, avec cette différence que l'empathie ou, de façon plus générale, la capacité à l'émotion a remplacé la parole. L'empathie caractérise, par opposition aux hommes-machines, le genre humain dans sa totalité, avec « les hommes les plus hébétés », pour reprendre l'expression de Descartes.

Maintenant, je voudrais discuter du contexte cartésien, du décor si l'on veut, dans lequel Dick présente cette image de la machine. Le premier élément cartésien tient à l'omniprésence des animaux-machines. En effet, avec les nuages radioactifs qui balayent le continent américain, les animaux véritables sont aussi rares que prisés dans le monde que nous décrit Dick. Ils coûtent cher, et Deckard, qui n'en a pas les moyens, a acheté un mouton électrique. De loin, on ne voit pas la différence, les voisins peuvent le prendre pour un véritable ovine mais Deckard lui-même n'en connaît que trop le caractère factice.

La fabrication, la promotion, la vente des animaux artificiels représentent tout un secteur de l'économie, que Dick compare à l'industrie automobile du XX<sup>ème</sup> siècle. Le roman de Dick est, comme les textes de Descartes, peuplé d'animaux artificiels. Seulement, dans le roman de Dick, ceux-ci, qui peuvent être à peu près « convaincants », ne remplacent jamais les animaux véritables. La thèse, que l'émotion est propre au vivant et que la mécanique ne peut pas en rendre les signes extérieurs, implique une différence, non seulement de l'androïde à l'humain, mais de l'artificiel au vivant. Les animaux ne sont pas des machines, et les animaux-machines ne font pas illusion<sup>16</sup>. L'être humain qui les approche, les reconnaît pour ce qu'ils sont. Seuls les androïdes, qui, n'ayant pas eux-mêmes d'émotion, réussissent mal à en saisir les signes extérieurs, s'y laissent prendre.

Cette insensibilité des androïdes aux animaux donne lieu à un épisode directement inspiré de Descartes. En effet, lorsque Deckard fait subir à Rachel Rosen un test d'empathie, celle-ci échoue, et son oncle Eldon Rosen explique l'insensibilité de la jeune femme par son enfance.<sup>17</sup> Rachel est née et a grandi sur le *Salander 3*. Elle n'a donc jamais eu l'expérience directe des animaux terrestres. Elle ne connaissait notre planète que par la bibliothèque du vaisseau et la conversation des membres de l'équipage. Rachel est bien cet enfant qu'imaginait Descartes, grandissant au milieu de machines, ramené ensuite sur Terre et confronté au règne animal. Cet enfant, disait Descartes, considérera les animaux comme des machines, de même nature que les automates avec lesquels il avait l'habitude de jouer. Et Eldon Rosen, qui explique de cette façon l'insensibilité de Rachel, semble bien avoir lu

---

<sup>15</sup> *ibid.* p. 27, p. 28.

<sup>16</sup> *ibid.*, p. 70.

<sup>17</sup> *ibid.*, p. 46.

Descartes. Seulement, dans le roman de Dick, l'argument cartésien n'est pas valide. L'enfance de Rachel ne suffit pas à expliquer son apathie : Rachel est une androïde. La sympathie avec une vie animale irréductible au règne mécanique n'est pas un préjugé qu'une autre éducation pourrait écarter mais une expérience, immédiate pour l'être humain et dont les androïdes sont exclus.

Il faut dire toutefois que l'empathie, si elle est immédiate, est toujours ambiguë. Il y a toujours une incertitude sur la nature de l'être que l'on a devant soi. L'androïde, comme l'animal artificiel, est de l'extérieur et, à l'observer, froidement indiscernable de son analogue naturel. L'être humain en ressent de façon immédiate la froideur. Mais cette impression n'est jamais une preuve objective. Elle ne peut pourtant pas être écartée. C'est un soupçon qui revient, sans fondement rationnel. L'incertitude, à laquelle sont constamment confrontés les personnages, tient à une tension entre ce qu'ils pensent ou ce que le contexte, des obligations de toutes sortes, les amènent à penser et ce qu'ils ressentent. Une tension, entre une ressemblance objective du visage androïde avec le visage humain, et l'impression inarticulée d'une froideur, d'une inhumanité radicale. Il n'y a que les androïdes qui ignorent cette incertitude, cette inquiétude. Ils ne saisissent pas eux-mêmes la froideur des êtres artificiels.<sup>18</sup> Ils ne se reconnaissent pas entre eux et ne distinguent pas, ou difficilement, les animaux artificiels.

Le test de Voigt-Kampff fait, du reste, appel à l'humanité de l'examineur et ne pourrait pas être appliqué de façon purement mécanique, ni réalisé par un androïde. Les réactions du patient, comme la dilatation de pupilles, sont magnifiées, mesurées par l'appareil mais le résultat doit être interprété par un être humain, qui, seul, est capable de sentir la présence de l'émotion. C'est pourquoi également le test n'est pas absolument fiable. On retrouve là ces conséquences de la thèse de Watson : si les machines par principe ne peuvent pas exprimer d'émotion, c'est que les signes de l'émotion ne se laissent pas décrire de façon objective et que les machines elles-mêmes sont incapables de les repérer. Les signes de l'émotion se ressentent mais ne se décrivent pas ou ne se pensent pas.

Il faut aussi noter à propos de cette ambiguïté de l'émotion une divergence entre Dick et Asimov. Asimov a d'abord une image de la machine comparable à celle de Dick.<sup>19</sup> Le robot, pour Asimov, est une machine qui parle mais reste froide, sans émotion et insensible à l'amour, au beau, au bien ou au religieux. Son comportement, absolument amoral, serait donc potentiellement dangereux s'il n'était pas réglé par les trois lois de la robotique. Ces lois sont implémentées dans le cerveau des robots. Ceux-ci sont construits de telle façon qu'ils ne peuvent en aucun cas transgresser ces lois.

La première loi stipule qu'un robot ne peut en aucun cas blesser ou tuer un être humain. Or les mésaventures de Deckard mettent bien en évidence la difficulté qui y est contenue. Pour que la loi soit applicable, il faut que le robot puisse reconnaître les êtres humains et les différencier des machines. Mais, imaginons que les progrès de la robotique en viennent à donner au robot un corps, qui de l'extérieur ne se distingue pas du corps humain. Comme dans le roman de Dick, la différence entre l'humain et la machine tiendra alors seulement à l'émotion. Et il n'est nullement clair que les robots, qui, eux-mêmes, sont dépourvus d'émotions, puissent en repérer les signes et reconnaître l'être humain qu'ils doivent protéger. Il me semble que, ici, Dick est plus cohérent qu'Asimov. Si l'expression de l'émotion ne se laisse pas reproduire de façon mécanique, c'est qu'elle n'a pas d'objectivité ou ne consiste pas

---

<sup>18</sup> *ibid*, p. 100. Phil. Resch, un autre chasseur d'androïdes, a bien décelé cette froideur caractéristique chez Polokov (« Polokov struck me as cold. Extremely cerebral and calculating; detached »), alors que son supérieur Garland, qui est lui-même un androïde, s'est laissé prendre à jeu de celui-ci.

<sup>19</sup> Ainsi, Daniel R. Olivaw, le robot-détective, « n'est gêné ni par la maladie, la panique, l'amour ou la culpabilité ». La « curiosité » est pour lui la recherche vaine d'une connaissance inutile, et il lui manquera toujours le sens de « la beauté, le bien, l'art, l'amour et Dieu » (*The Caves of Steel*, *op. cit.*, p. 221).

en des signes qui se définiraient de façon univoque. L'émotion de l'autre ne sera donc sensible que dans l'ambiguïté et à un être qui est lui-même capable d'émotions. De ce point de vue, la première loi d'Asimov est sans objet.

Maintenant, la question qui se pose est de savoir d'où viennent ces émotions qui, pour Dick, font le propre de l'homme. Il faut à ce propos évoquer le rôle que joue la « boîte Penfield » ou « l'orgue à humeur » (« mood organ »). Il s'agit d'une console sur laquelle on peut programmer à l'avance son humeur, avec toutes les émotions qui l'accompagnent : un repos bien mérité ; une dépression sans fond ; avoir envie de regarder la télévision quel que soit ce qui passe. A la différence d'instruments analogues dans d'autres textes de Dick, la boîte Penfield semble fonctionner de façon purement physique, induisant une humeur chez l'utilisateur par de légères décharges électriques.<sup>20</sup> C'est donc dire que les humeurs, les émotions, peuvent être produites dans le corps humain de façon mécanique. Il y a du reste un autre indice de ce soubassement physique des émotions humaines. Deckard, en effet, est convaincu que l'existence de l'émotion chez l'homme est liée à l'évolution de l'espèce, l'empathie en particulier apparaissant comme un élément central pour la survie de cet animal grégaire qu'est l'être humain.<sup>21</sup> C'est l'évolution naturelle qui aurait donné à l'espèce humaine un système d'émotions augmentant ses chances. L'émotion, avec ses manifestations extérieures, apparaît donc comme un mécanisme physiologique. Cette hypothèse est tout à fait cartésienne. Descartes pensait rendre compte des passions par le mouvement des esprits animaux et leur influence sur cette glande pinéale, qu'il avait découverte au milieu du cerveau. Cependant, cette perspective, d'une production physiologique des émotions, est difficilement conciliable avec la thèse de Dick, que les émotions font le propre de l'homme et que les machines sont incapables d'en manifester même les signes extérieurs. Si les émotions et leurs expressions extérieures sont déterminées par un mécanisme physiologique, il devrait être possible de reproduire ce mécanisme chez l'androïde (d'autant plus qu'on lui a donné un corps presque identique au corps humain).

Il est vrai que Deckard évoque au détour d'une phrase un dispositif, un défaut en réalité, placé à dessein dans le cerveau des androïdes et qui leur interdit de participer au « Mercerisme », la religion en vogue dans ce XXI<sup>ème</sup> siècle.<sup>22</sup> Le Mercerisme est fondé sur l'empathie. Il est donc possible de penser que la froideur des androïdes, l'absence d'empathie est voulue par le constructeur et n'est due qu'à un module spécifique dans leur cerveau et qui les exclut de l'émotion. C'est la lecture qui gouverne le film, *Blade Runner*, que Ridley Scott a tiré du roman de Dick. Ici, les hommes tiennent (pour une raison qui reste assez obscure) à empêcher les androïdes de développer leurs propres émotions. Le moyen pour cela est radical. Il suffit de faire en sorte que ces machines « s'éteignent » d'elles-mêmes au bout de quelques années, avant d'avoir eu le temps de se construire un système d'émotions et, en quelque sorte, une personnalité.

Pourtant, c'est, je crois, une lecture partielle du roman de Dick. Elle amène à réduire à un simple fait, une simple insuffisance de fait, l'insensibilité, l'apathie, qui, pour Dick, définit les machines. Par ailleurs, cette incise, sur le « défaut » des androïdes intervient très tard dans le roman, de façon très rapide et ne concerne spécifiquement que l'expérience du « Mercerisme ». Il me semble donc qu'il faut prendre au sérieux, plutôt que de l'écarter en invoquant ce « défaut délibéré » laissé dans le cerveau des androïdes, la tension entre les deux thèses : d'un côté, l'émotion est le propre de l'homme, les robots étant par principe incapables d'en imiter ou d'en déceler les manifestations extérieures ; de l'autre côté, l'émotion chez l'homme est occasionnée par des phénomènes physiques. Les décharges électriques de la

---

<sup>20</sup> Dick, *op. cit.*, p. 3.

<sup>21</sup> *ibid*, p. 27.

<sup>22</sup> *ibid*, p. 158.

boîte Penfield ne font que prendre la place des esprits animaux de la physiologie cartésienne. Ces deux thèses, qui semblent contradictoires, sont au centre de l'idée de l'homme, de l'image de l'homme et de la machine que véhicule la science-fiction. Je n'ai trouvé dans le roman de Dick aucun élément qui permette de les concilier.

Dans ses textes ultérieurs, Dick semble renoncer à la thèse cartésienne, selon laquelle l'émotion a un soubassement physiologique. L'empathie définit l'humain, et il faudrait dire qu'une machine qui en est capable est « humaine », aussi humaine que les êtres humains. Mais Dick pose maintenant que l'empathie est inexplicable par la physiologie de son agent. Sa source, notre humanité donc, n'est pas plus dans le corps de la machine que l'âme cartésienne. En fait, la source de l'émotion est comme l'âme cartésienne, quelque chose qui est attaché au corps d'une machine mais en est distinct et que rien ne montre dans ce mécanisme. C'est en ce sens, je crois, qu'il faut comprendre ce passage assez obscur d'un essai de 1976 :

Si un dispositif mécanique interrompt ses opérations pour venir à votre aide, vous lui reconnaissez, avec gratitude, une humanité qu'aucune analyse de ses transistors ou de ses circuits de transmission ne permettra d'élucider. Un scientifique, qui analyserait les circuits électriques de cette machine pour localiser son humanité serait comme notre savant qui essaye en vain de localiser l'âme dans l'homme et, faute de trouver un organe particulier placé en un point précis, préfère refuser d'admettre que nous avons une âme. Ce que l'âme est à l'homme, l'homme l'est à la machine.<sup>23</sup>

Partant de l'idée que l'émotion est le propre de l'homme, Dick a le choix entre deux options : ou bien accepter que les manifestations de l'émotion sont le produit naturel d'un mécanisme physiologique et jouer avec deux thèses apparemment inconciliables ; ou bien distinguer la source des émotions des mécanismes corporels pour la réaliser dans une sorte d'âme et retrouver alors le problème cartésien de l'union de l'âme et du corps. Si l'on suit cette deuxième voie, il faut bien considérer le support de l'émotion, cela qui en l'homme donne lieu à l'émotion et lui permet de l'exprimer, comme une sorte d'âme, distincte du corps. L'expression de l'émotion, à laquelle on reconnaît les humains et que ne peuvent pas imiter les androïdes, devient alors l'exact analogue de la parole cartésienne : le signe d'une âme que possède l'homme et dont est dépourvue la machine. Le dualisme cartésien est simplement déplacé, avec un support de l'émotion à la place de la substance pensante. Dans ses Essais ou dans la cosmologie de *Valis*, Dick emprunte du reste le registre de la philosophie, sans forcément réussir à résoudre ces nouveaux problèmes. En fait, le roman de 1968 semblait déjà indiquer, en jouant sur des hypothèses apparemment inconciliables, que, comme l'union de l'âme et du corps et la possibilité de la parole, l'émotion proprement humaine se vit, ou se raconte dans des « conversations ordinaires », mais ne se pense pas.<sup>24</sup>

---

<sup>23</sup> Dick, *The Shifting Realities of Ph. K. Dick*, op. cit., p. 212.

<sup>24</sup> Descartes à Elisabeth, 28 juin 1643 op.cit., p. 1158.