



HAL
open science

Des robots "sociaux" ? : enquête philosophique sur l'altérité des artefacts

Julien de Sanctis

► **To cite this version:**

Julien de Sanctis. Des robots "sociaux" ? : enquête philosophique sur l'altérité des artefacts. Philosophie. Université de Technologie de Compiègne, 2022. Français. NNT : 2022COMP2724 . tel-04453902

HAL Id: tel-04453902

<https://theses.hal.science/tel-04453902>

Submitted on 12 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Par Julien DE SANCTIS

*Des robots « sociaux » ?
Enquête philosophique sur l'altérité des artefacts*

Thèse présentée
pour l'obtention du grade
de Docteur de l'UT



Sou tenue le 12 décembre 2022

Spécialité : Philosophie, Épistémologie, Histoire des Sciences et
des Techniques : Unité de recherche COSTECH (EA-2223)

D2724

Julien DE SANCTIS

Des robots « sociaux » ?
Enquête philosophique sur l'altérité des artefacts



THÈSE

présentée pour l'obtention du grade de Docteur en philosophie
de l'Université de Technologie de Compiègne

12 décembre 2022

sous la direction du Pr. François-Xavier GUCHET

Des robots « sociaux » ?

Enquête philosophique sur l'altérité des artefacts

THÈSE

en vue de l'obtention du grade de Docteur en philosophie
de l'Université de Technologie de Compiègne

Spécialité : Philosophie, Épistémologie, Histoire des Sciences et des Techniques

Présentée et soutenue publiquement par

Julien DE SANCTIS

Le 12 décembre 2022

Sous la direction du

Pr. François-Xavier GUCHET

Membres du Jury

Véronique AUBERGÉ

Chargée de recherche, CNRS

Emmanuel GRIMAUD

Directeur de recherche, CNRS

François-Xavier GUCHET,

Professeur, Université de Technologie de Compiègne

Charles LENAY,

Professeur, Université de Technologie de Compiègne

Thierry MÉNISSIER

Professeur, Université Grenoble Alpes

François-David SEBBAH

Professeur, Université Paris-Nanterre



Des robots « sociaux » ?

Enquête philosophique sur l'altérité des artefacts

Julien DE SANCTIS

Thèse de **philosophie**

réalisée dans le cadre du dispositif **CIFRE**

Convention Industrielle de Formation par la Recherche

Cette thèse a été élaborée

au sein du laboratoire **COSTECH** de l'**Université de Technologie de Compiègne**

« Connaissance, Organisation et Systèmes Techniques », EA2223

Centre Pierre Guillaumat, 60319 Compiègne

et

au sein de **Spoon**

Société par Action Simplifiée

14 rue Mathurin Régnier, 75015 Paris

REMERCIEMENTS

Je tiens tout d'abord à remercier mon directeur de thèse, Xavier Guchet, de m'avoir accompagné durant ces cinq années de recherche. Merci pour ta disponibilité et pour la bienveillance de tes remarques et encouragements. Merci également d'avoir permis à la singularité de mon travail d'émerger en ne cherchant jamais à l'orienter.

Merci aux trois co-fondateurs de Spoon, Jérôme Monceaux, Aymeric Masurelle et Thibault Hervier d'avoir accepté l'idée de cette thèse CIFRE en philosophie et donc de l'avoir rendue possible. A cet égard, je me dois de mettre en évidence une causalité aussi courte qu'essentielle : merci à Bertrand Boudin de m'avoir mis en contact avec Frédéric Delaunay qui, lui-même, m'a mis en contact avec Spoon. Frédéric, je t'en suis des plus reconnaissants.

Je remercie vivement les membres du jury, Véronique Aubergé, Emmanuel Grimaud, Charles Lenay, Thierry Ménissier et François-David Sebbah d'avoir accepté d'examiner et de discuter ce travail.

Merci également à Vincenzo Raimondi et à Hugues Choplin pour leur participation à mes CSI, ainsi qu'à Serge Bouchardons et, une nouvelle fois, à Hugues d'avoir accepté de lire certains fragments de cette recherche pour en discuter.

Ma plus profonde reconnaissance va à ma mère, Catherine. Aucun mot, aucun remerciement ne saurait te faire honneur. Un mot tout de même : infini. C'est ce que tu m'as toujours donné et ce que je te dois. Je te dédie ce travail.

Cette immense reconnaissance va aussi à ma femme, Marion. Merci de m'avoir supporté dans les deux sens du terme durant ces longues et parfois très douloureuses années de recherche. Merci pour ta patience. Ce petit mot condense, lui aussi, ce que je te dois. En latin, *patientia* renvoie à l'« action de supporter, d'endurer » et dérive de *pati*, « éprouver, souffrir ». Si la patience désigne à la fois l'arme et l'expérience du thésard en philosophie, elle concerne aussi son entourage qui se retrouve indirectement embarqué dans les vicissitudes de la recherche. Au terme de ce voyage, je vois encore plus clairement le soutien que tu m'as offert. Je te dédie également ce travail, même si la philo « c'est pas ton truc ».

Merci à mon père, Rosario, d'avoir toujours cru en moi avec fierté et enthousiasme. Je ne saurais dire ce qui relève de l'inné et de l'acquis, mais je partage ta grande curiosité, ta ténacité et ton exigence, parfois inquisitrice, envers toi-même. Si ce travail a quelque valeur, c'est notamment grâce à ce fond commun.

J'adresse un remerciement très spécial à Manu, ami et compagnon de maïeutique, mais aussi, d'une certaine façon, « cause motrice » de cette thèse. Tu as été le principe du mouvement qui m'a fait revenir à la philosophie. Aujourd'hui, après tant d'années de doutes, de plaisirs intellectuels et, parfois, de souffrance morale, je sais que ma joie te doit beaucoup. Permetts-moi de te dédier cette recherche.

Un immense merci à Valentine Levacque d'avoir accepté de se livrer à l'exercice fastidieux de la relecture. Tu t'es plongée dans cette tâche aride avec autant d'enthousiasme que de rapidité et de sérieux. Je suis ton obligé.

Alors que s'achève cette aventure doctorale, j'ai une pensée profondément reconnaissante pour Brice Herrmann, mon professeur particulier de français lorsque j'étais en première. Ton encadrement, ta rigueur et tes enseignements furent pour moi aussi formateurs que décisifs. Ils ont sans conteste façonné le socle sur lequel je me suis appuyé pour mener à bien ce long et périlleux travail. Je t'en remercie de tout cœur.

Je remercie toutes celles et ceux qui, d'une façon ou d'une autre, m'ont soutenu au moment le plus sombre de ces cinq années de recherche. Je pense tout particulièrement à Antoine, autre compagnon des profondeurs. Ton soutien était peut-être « normal », comme tu dis, mais cette normalité n'en fut pas moins curatrice. Normal rime ici avec attention et considération. Je pense tout aussi particulièrement à ma sœur, Julie, qui m'accompagne dans l'existence avec amour et délicatesse depuis tant d'années. Tes mots sont des abris et ta présence un perpétuel réconfort. Je remercie également l'équipe de Spoon pour son soutien et notamment à Lionel qui m'a longuement écouté lorsque je me voyais tout arrêter. Merci, enfin, à quatre camarades de recherche, Gaëlle Garibaldi, Renaud Mignerey, Amaury Bazalgette et Aurore Dandoy pour certains mots et certains moments partagés qui me furent d'un grand secours. Aurore, je te remercie également de m'avoir accompagné dans l'écriture naissante du manuscrit et, plus largement, de m'avoir témoigné une précieuse sollicitude.

Merci à Amélie Cordier pour nos échanges stimulants et pour ses mises en contact. Je remercie également Véronique Aubergé (à nouveau) et Sophie Sakka de m'avoir chaleureusement accueilli à Grenoble et à Nantes dans leurs laboratoires respectifs pour découvrir leurs travaux et leur approche des robots sociaux.

Je remercie très chaleureusement Marion Genaivre et Flora Bernard pour leur estime et leur soutien enthousiaste tout au long de cette aventure. Merci pour votre compréhensive patience alors que, engagé dans une fin qui n'en finissait plus de finir, je tardais à rejoindre Thaé. Marion, merci également pour ton écoute et nos discussions dédiées à cette thèse.

Au moment de conclure ces remerciements, je pense à mes grands-parents, Gabriella, Guido, Joseph et Ginette, que j'aime profondément et qui m'ont tant donné. Entre la Bretagne et l'Italie, entre Paris et Argenteuil, vous m'avez offert la chaleur, l'attention et la magie qui rendent l'enfance aussi douce qu'inoubliable. Merci de m'avoir élevé. Nonna, toi qui es encore là aujourd'hui, j'ai été pleinement heureux de pouvoir t'annoncer l'accomplissement de ce travail.

Mes dernières pensées vont à celle qui, en réalité, les occupent toutes : ma fille, Giulia. Si tu lis un jour ces lignes et les suivantes, sache qu'elles ont partagé leur naissance avec la tienne et qu'elles te doivent d'avoir été écrites dans le bonheur.

RÉSUMÉ

Cette recherche doctorale menée en CIFRE trouve son origine dans l'embarras philosophique que suscite le nom même de son objet de recherche : le robot social. En quoi un robot peut-il être qualifié de social ? A première vue, une telle notion semble relever de l'oxymore. Le robot n'incarne-t-il pas originellement la figure de l'esclave artificiel, c'est-à-dire de l'être purement instrumental ? A l'inverse, la socialité n'est-elle pas extérieure à toute définition ou délimitation instrumentale en ceci qu'elle ne sert *a priori* à rien et *a posteriori* à tout ? Ne doit-on pas considérer le social comme l'horizon, c'est-à-dire comme la condition de possibilité de toute utilité et, plus largement de tout intérêt individuel et collectif ? Parler de robot *social* laisse également entendre que l'artefact possède le social comme une propriété. Mais qu'est-ce que cela signifie ? De quelle propriété s'agit-il ? Enfin, le social n'est-il pas aussi un concept moral en ceci qu'il renvoie à nos façons d'être ensemble et, plus spécifiquement, d'être avec l'autre dans le face-à-face ? Le robot serait alors social en tant qu'autre face auquel nous nous tenons. Dans cette hypothèse, la socialité de l'artefact apparaît comme ce qui, *phénoménologiquement*, nous pousse à l'envisager non pas comme une chose, un « cela », mais comme un « tu » manifestant une altérité. Mais comment un artefact pourrait-il être un autre ? Comment le concept d'altérité, ordinairement réservé aux êtres qui comptent moralement (humains, animaux et, de plus en plus, végétaux et écosystèmes), pourrait-il nous aider à comprendre la socialité des robots ? Faut-il, pour cheminer vers cette compréhension, considérer cette socialité comme une forme de moralité et, si oui, en quel sens ? Autrement dit, un robot social n'est-il pas aussi et nécessairement un robot moral ? C'est la thèse que nous cherchons à défendre en examinant le sens d'être de l'altérité artefactuelle, non pas comme subjectivation du robot social, mais comme effectivité morale interpellant notre responsabilité.

Mots clés : altérité, robotique sociale, robots sociaux, social, socialité, artefacts, éthique, postphénoménologie, médiation technique, effectivité, responsabilité

ABSTRACT

This doctoral research conducted under the CIFRE program originates from the philosophical embarrassment caused by the very name of its research object: the social robot. How can a robot be qualified as social? At first sight, such a notion seems to be an oxymoron. Doesn't the robot originally embody the figure of the artificial slave, i.e. the purely instrumental being? On the other hand, isn't sociality external to any instrumental definition or delimitation? Should we not consider it as the horizon, that is, as the condition of possibility of all utility and, more broadly, of all individual and collective interest? To speak of a social robot also suggests that the artifact possesses the social as a property. But what does this mean? What kind of property is it? Finally, isn't the social also a moral concept in that it refers to our ways of being together and, more specifically, of being with the other in the face-to-face? The robot would then be social as this other in front of whom we stand. In this hypothesis, the sociality of the artifact appears as what, phenomenologically, pushes us to consider it not as a thing, a "that", but as a "you" manifesting an alterity. But how could an artifact be an other? How could the concept of otherness, usually reserved for beings that count morally help us understand the sociality of robots? In order to move towards this understanding, should we consider this sociality as a form of morality and, if so, in what sense? In other words, is a social robot also and necessarily a moral robot? This is the thesis we seek to defend by examining the sense of being of the artefactual alterity, not as a subjectivation of the social robot, but as a moral effectivity questioning our responsibility.

Keywords: otherness, social robotics, social robots, sociality, artifacts, ethics, postphenomenology, technical mediation, effectivity, responsibility

LISTE DES ABRÉVIATIONS

AMA : agents moraux artificiels

CDO : *chief digital officer*

DNTA : *dual nature of technical artifacts*

DUDH : Déclaration universelle des droits de l'Homme

FEI : fonction extrinsèque à l'interaction

FII : fonction intrinsèque à l'interaction

IHM : interaction humain-machine

ISHM : interaction sociale humain-machine

IHR : interaction humain-robot

ISHR : interaction sociale humain-robot

IVG : interruption volontaire de grossesse

NdA : niveau d'abstraction

PDG : président directeur général

TMT : théorie des médiations technologiques

TSA : troubles du spectre autistique

UX : *user experience* (expérience utilisateur)

UXD : *user experience design* (design d'expérience Utilisateur)

VSD : *Value Sensitive Design*

TABLE DES MATIÈRES

INTRODUCTION – LE PROBLÈME PHILOSOPHIQUE DE LA SOCIALITE DES ROBOTS.....	16
0.1 Le robot social : un oxymore ?	16
0.2 L’altérité : point de suture entre socialité et moralité des robots ?	18
0.3 La robotique sociale et ses domaines d’application	19
0.3.1 <u>La robotique sociale scientifique</u>	20
0.3.2 <u>La robotique sociale de services</u>	20
0.3.3 <u>La robotique sociale dite « personnelle » ou « domotique »</u>	22
0.3.4 <u>La robotique sociale thérapeutique</u>	23
0.4 La robotique sociale dans le contexte de la thèse CIFRE	25
0.5 Positionnement scientifique	27
0.5.1 <u>Le tournant chosique de la philosophie des artefacts techniques</u>	27
0.5.2 <u>La métaphysique analytique du Dual Nature of Technical Artifacts</u>	28
0.5.3 <u>L’approche postphénoménologique des médiations technologiques</u>	29
0.5.4 <u>La philosophie des techniques au défi de la robotique sociale</u>	31
PARTIE I – DU ROBOT AU ROBOT SOCIAL : ETAT DE L’ART CRITIQUE DE LA SOCIOBOTIQUE.....	35
Chapitre I - Qu’est-ce qu’un robot ?	36
1.1 Le robot est un mythe	37
1.1.1 <u>Le robot n’existe pas</u>	37
1.1.2 <u>Le robot est une figure mythique</u>	39
1.2 Le robot est un automate	43
1.3 Le robot est un faisceau de normes	47
1.4 Définitions, taxonomie et exemples	55
1.4.1 <u>Définitions technico-fonctionnelles</u>	55
1.4.2 <u>Approches conceptuelles du robot : la question du « rapport au monde »</u>	59
1.5 Taxonomie robotique	61

1.5.1	<u>Robotique industrielle</u>	61
1.5.2	<u>Robotique domestique et servicielle</u>	63
1.5.4	<u>Robotique médicale</u>	64
1.5.4	<u>Robotique militaire et d'intervention</u>	66
1.5.5	<u>Robotique scientifique</u>	68
1.5.6	<u>Robotique de transport</u>	70
1.5.7	<u>Robotique éducative</u>	73
Chapitre 2 - Qu'est-ce qu'un robot social ?		75
2.1	Portrait-robot du sociobot : terminologie et définitions en vigueur	76
2.1.1	<u>Les approches généralistes de la robotique sociale</u>	76
2.1.1.1	Robot social <i>versus</i> Robot sociétal	77
2.1.1.2	Robot socialement interactif	79
2.1.1.3	Robot sociable	80
2.1.1.4	Normes et expériences sociales : le sociobot de l'UX design	83
2.1.2	<u>Les approches spécifiques de la sociobotique</u>	87
2.1.2.1	L'« objet à comportement » ou l'influence du mouvement	87
2.1.2.2	L'« artefact relationnel » : émotions et empathie artificielles	92
2.1.2.3	L'autonomie sociale du « substitut » : autorité, présence et coordination	98
2.1.2.4	Le « robot existentiel » comme vivant post-biologique	101
2.2	Le social comme « propriété » de l'artefact	107
2.2.1	<u>Quatre grands types de propriété</u>	107
2.2.2	<u>Le social comme propriété dispositionnelle : l'interaction sociale</u>	108
2.2.2.1	Le robot comme sujet de socialité	108
2.2.2.1.1	<i>Première critique : la dimension « interobjective » des interactions</i>	109
2.2.2.1.2	<i>Seconde critique : la coordination affective ou l'origine infrasubjective de la socialité</i> ...	112
2.2.2.2	Paro et Lovot : des robots sociaux ou sociables ?	115
2.2.3	<u>Le social comme propriété systémique : la fonction sociale</u>	119
2.2.3.1	Le sociofonctionnalisme : interaction et fonction sociales	119
2.2.3.2	Fonction sociale <i>versus</i> autonomie sociale	120
2.2.4	<u>Le social comme propriété relationnelle</u>	122
2.2.4.1	Robot social ou robot socialement interactif ?	122
2.2.4.2	De l'interaction à l'interactance : une typologie de l'agentivité des artefacts	123

2.2.4.3	Le statut protéiforme du sociobot	127
Chapitre 3 - La roboéthique : des éthiques « envers », « autour », « dans » et « avec » les robots sociaux		
		129
3.1	L'éthique <i>envers</i> les robots ou la question de leur statut moral	130
3.1.1	<u>Le sociobot comme instrument</u>	130
3.1.2	<u>Le sociobot comme patient moral</u>	134
3.1.2.1	La patientalité morale intrinsèque	134
3.1.2.2	La patientalité morale extrinsèque	138
3.1.3	<u>Le sociobot comme agent moral</u>	143
3.1.3.1	L'agentivité morale intrinsèque : conscience, intentionnalité et responsabilité	144
3.1.3.1	L'agentivité morale extrinsèque : les positions relativiste et relationnelle	146
3.2	L'éthique <i>autour</i> du robot : conception, commercialisation et usages des sociobots	152
3.2.1	<u>Éthique instrumentale et conformité axiologique des sociobots</u>	152
3.2.2	<u>La question (délaissée) de la communication</u>	161
3.3	L'éthique <i>dans</i> le robot : la moralité « implicitement » et « explicitement » programmée.....	167
3.4	L'éthique <i>avec</i> le robot : une éthique immanente à la relation	171
3.4.1	<u>« The machine question » : une approche lévinassienne</u>	171
3.4.2	<u>Une approche phénoménologique et socio-relationnelle de la roboéthique</u>	177
3.4.3	<u>De la critique de l'empathie artificielle à l'anthropomorphisme comme fondement d'une « éthique synthétique »</u>	179
Transition - Du robot « social » au robot « moral »		187
 <u>PARTIE II – L'ALTÉRITÉ : UN CONCEPT POUR LA PHILOSOPHIE DES TECHNIQUES ?</u>		
		189
Chapitre 4 - Présentation de Spoon et discussion des fondements conceptuels de l'activité de conception des robots sociaux		
		190
4.1	L'entreprise Spoon : objet social, organisation, produits et valeurs	190
4.2	Genèse et répertoire conceptuel de Spoon	193
4.2.1	<u>La « récurrence d'intérêt » des sociobots</u>	193

4.2.2	<u>Une appréhension phénoménologique du sociobot comme « reflet d'existence de l'utilisateur »</u>	194
4.2.3	<u>Du « robot social » au « créatures merveilleuses » en passant par les « créatures artificielles » : nommer pour comprendre, comprendre pour nommer</u>	196
4.2.3.1	Le « vrai robot social »	196
4.2.3.2	Le robot comme « créature merveilleuse »	208
4.3	Spoon et le marché de la robotique sociale	211
4.3.1	<u>Un marché immature ?</u>	211
4.3.2	<u>Identifier la valeur des sociobots</u>	214
4.4	Spoon et l'altérité du sociobot	218
4.4.1	<u>Être regardé par un robot : une thèse sartrienne ?</u>	219
4.4.2	<u>L'« animalité transpèce » : le vivant au-delà de l'espèce et sa question éthique</u>	222
	Chapitre 5 - L'altérité sociobotique	225
5.1	L'altérité dans la philosophie des techniques et dans la littérature sociobotique	225
5.1.1	<u>« L'autre robotique » : évacuer la socialité par l'altérité ?</u>	225
5.1.2	<u>Sociobotique et « quasi-altérité »</u>	226
5.1.3	<u>L'« autruiphane » ou la dimension phénoménotechnique de l'altérité</u>	229
5.1.4	<u>Les « alterity relations » de la postphénoménologie</u>	232
5.2	Penser l'effectivité sociobotique : l'altérité comme effet	235
5.2.1	<u>Ustensilité versus altérité : le sociobot comme phénomène persistant</u>	235
5.2.2	<u>La robotique sociale comme « allotechnique »</u>	236
5.2.3	<u>Ontophanie sociobotique et effets d'altérité</u>	240
5.3	Médiation sociobotique et subjectivation morale : l'effet d'altérité comme spécularité responsabilisante	247
5.3.1	<u>D'un phénomène ontophanique plus global de responsabilité : l'effet d'altérité est une expérience morale</u>	247
5.3.2	<u>Est-il immoral de brutaliser un robot ? : le point de vue des trois grandes éthiques normatives</u>	248
5.3.3	<u>La médiation sociobotique comme spécularité moralement subjectivante</u>	255
5.4	La socialité comme relation de responsabilité	264

5.4.1	<u>La socialité comme relation éthique originaire : la question du « face-à-face »</u>	265
5.4.2	<u>La responsabilité comme lien ancré dans l'empathie</u>	267
5.4.3	<u>Penser l'au-delà du robot à partir du robot : une herméneutique matérielle de la responsabilité envers le vivant</u>	270
	Conclusion et ouverture - Et la politique ?	275

« L'éthique est la mise en question de ma spontanéité par la présence d'autrui »

Emmanuel Levinas, *Totalité et infini*.

INTRODUCTION

Le problème philosophique de la socialité des robots

Cette thèse de philosophie résulte de cinq années de recherche dont trois effectuées dans le cadre du dispositif CIFRE¹. Elle répond à l'exigence de penser philosophiquement les « robots sociaux » à partir du terrain, dans la perspective du tournant empirique et pragmatiste pris par la philosophie des techniques ces dernières décennies. Plus particulièrement, elle cherche à articuler le terrain de recherche qu'est la robotique sociale aux problématiques spécifiques rencontrées par le partenaire industriel de la CIFRE, à savoir la société Spoon.

0.1 Le robot social : un oxymore ?

« A quoi ça sert ? » Telle est la question qui, bien souvent, gouverne notre rapport instrumental aux techniques. Malgré les multiples éclairages que les sciences sociales, l'histoire et la philosophie des techniques ont jetés sur les artefacts et leur « agentivité »² depuis 50 ans, la conception de la technique comme simple moyen au service de fins prédéfinies (conception dite « instrumentale » de la technique) demeure vivace. Pourtant, l'expérience quotidienne ne pointe-t-elle pas vers un au-delà de la pure instrumentalité de nos artefacts ? N'indique-t-elle pas que les techniques représentent plus que de simples moyens dont le mode d'existence s'épuiserait dans l'objectivité instrumentale ? Au-delà de la littérature académique et des études de terrain, l'intuition ne nous dit-elle pas que le *faire* des artefacts est aussi un *nous-faire* ? Ne sommes-nous pas, par exemple, sincèrement *attachés* à bon nombre d'entre eux ? N'avons-nous pas l'impression d'être plus ou moins nous-mêmes selon que nous en sommes « équipés » ou non ? N'appréhende-t-on pas le monde différemment selon que nous sommes dans la position du cycliste ou de l'automobiliste ? Mais tous les artefacts, il est vrai, ne semblent pas provoquer ce sentiment avec la même intensité.

¹ Convention industrielle de formation par la recherche.

² C'est-à-dire sur leur capacité à produire des effets bien au-delà de ce que nous en faisons. Il s'agit là d'une idée fondamentale sur laquelle nous reviendrons longuement.

Les robots sociaux, auxquels cette recherche est dédiée, comptent probablement parmi les êtres artificiels dont le pouvoir d'arrachement à l'instrumentalité est le plus évident. Cette « libération machinale », qu'il s'agira d'expliquer en détails, génère une perplexité qui se pose et même s'impose comme embarras philosophique. Loin des mythes et des fantasmes intéressés au « soulèvement des machines »³, le robot social « se libère » d'abord et surtout conceptuellement par son invitation quasi explicite à réviser nos catégories de pensée. Quel est cet être que nous pensons comme un objet, mais que nous ressentons, bien souvent, comme un sujet ? S'agit-il d'une machine sensible ? D'un artefact de compagnie ? D'un assistant artificiel ? La façon dont nous appréhendons le robot social, jusque dans le langage, comme objet-sujet l'institue comme entité liminaire, engluée dans un entre-deux qui n'en dit rien, ou trop peu. L'expérience qu'on en a est donc dérangeante au sens propre comme au sens figuré. Elle nous invite à quitter le confort d'une familiarité devenue étrange pour l'effervescence du doute et de l'incertitude. Autrement dit, le robot social nous invite à penser autrement, à « habiter le trouble »⁴ selon la belle formule de Donna Haraway. En cela, l'artefact se présente comme un objet philosophique certes non-identifié, mais non moins manifeste.

Notre enquête doctorale s'ouvre donc sur l'étonnement que suscite, à la fois théoriquement et empiriquement, son objet de recherche. La notion même de robot social se révèle très surprenante en ceci qu'aucun des deux termes qui la compose ne renvoie à un contenu conceptuel clairement identifié. Examiner cette notion permet donc d'interroger à nouveaux frais le robot d'un côté et le social de l'autre. Notre problème de départ peut ainsi se résumer très simplement à la question suivante : en quoi un robot peut-il être qualifié de social ? A première vue, parler de robot social peut paraître relever de l'oxymore. Le robot n'incarne-t-il pas originellement la figure de l'esclave artificiel, c'est-à-dire de l'être purement instrumental ? A l'inverse, la socialité – définie comme « fait d'être une entité sociale » - n'est-elle pas *a priori* extérieure à toute définition ou délimitation instrumentale en ceci qu'elle ne sert *a priori* à rien et *a posteriori* à tout ? Ne doit-on pas considérer le social comme l'horizon, c'est-à-dire comme la condition de possibilité de toute utilité et, plus largement de tout intérêt individuel et collectif ? Chacun peut faire l'expérience de cette idée en se demandant

³ Tel était le sous-titre, par exemple, du troisième opus de la saga Terminator.

⁴ Caeymaex, F., Despret, V., & Piéron, J. (2019). Habiter le trouble avec Donna Haraway. Éditions du Dehors.

« à quoi sert la socialité ? » Au mieux arrive-t-on à lister un certain nombre d'éléments sans jamais parvenir à faire le tour de la question. C'est que le social semble échapper par principe à toute tentative de réduction instrumentale.

0.2 L'altérité : point de suture entre socialité et moralité des robots ?

Comment, dès lors, un robot peut-il être social ? Que faut-il dire du social, mais aussi du robot pour dissiper leur contradiction apparente ? Que faut-il que le robot soit pour accéder à la socialité ? Ou, inversement, que faut-il que la socialité soit pour être accessible au robot ? S'agit-il d'une propriété que le robot posséderait comme il possède telle ou telle caractéristique de nature matérielle ou fonctionnelle ? Une telle conception incite à adopter le point de vue, classique dans la littérature, selon lequel la socialité robotique procède des capacités plus ou moins grandes de l'artefacts à interagir socialement avec nous. Le robot social serait donc synonyme de robot socialement interactif. Mais que faire, dans ce cas, du versant moral de la socialité ? Conformément à son étymologie latine, *socius*, « uni », « joint », « associé », le social ne renvoie-t-il pas également au(x) principe(s) directeur(s) de notre façon d'être ensemble et, plus spécifiquement, d'être avec l'autre dans le face-à-face ? Car au-delà du thème de l'interaction, il en va du principe des robots sociaux de

« donner à leur partenaires humains le "sentiment", l'impression "suffisamment convaincante et durable... d'être en présence de *quelqu'un*". Ils doivent provoquer chez l'interlocuteur l'impression d'être face à un "autre" » (Dumouchel & Damiano, 2016, p.117-118).

« Quelqu'un », « autre », voilà des termes que nous n'associons pas spontanément aux artefacts et qui, pourtant, semblent être au cœur de ce qui se joue avec la robotique sociale. Une telle « caractéristique définitionnelle » (*ibid*) situe très clairement l'analyse du sens d'être du sociobot dans le champ phénoménologique. La socialité du robot serait alors ce qui, *phénoménologiquement*, nous pousse à l'envisager non pas comme une chose, un « cela », dirait Martin Buber (1923), mais comme un « tu » manifestant une altérité. Mais comment un artefact pourrait-il être un autre ? Comment le concept d'altérité, ordinairement réservé aux êtres qui comptent moralement (humains, animaux et, de plus en plus, végétaux), pourrait-il nous aider à comprendre la socialité des robots ? Faut-il, pour cheminer vers cette

compréhension, considérer la socialité robotique comme une forme de moralité et, si oui, en quel sens ? Autrement dit, un robot social n'est-il pas aussi et nécessairement un robot moral ? C'est la thèse que nous souhaitons défendre. L'importance d'une telle question vient de ce qu'elle invite à conjoindre deux champs de réflexion qui, le plus souvent ne sont pas articulés : l'étude de leur socialité d'une part, et celle de leur moralité, de l'autre. Dans cette perspective, nous verrons qu'une idée d'altérité artificielle permet de conceptualiser la socialité des robots comme *effectivité morale* et, ce faisant, de les penser au-delà des catégories de l'objectivité et de la subjectivité. L'apport central de ce travail de thèse consistera ainsi à montrer comment le concept d'altérité, élaboré dans le cadre d'une philosophie des techniques, permet de faire le lien entre ces deux littératures tout en enrichissant le regard des concepteurs sur leur activité. Car si notre enquête se présente avant tout comme un travail de conceptualisation du robot social, précisons que la dimension immanquablement théorique de cet effort ne diminue en rien l'ambition pratique de sa destination, à savoir éclairer la conception technique par la conceptualisation philosophique. Nous nous inscrivons en cela dans le sillage du tournant empirique et pragmatiste pris par la philosophie des techniques contemporaine dont la vocation interventionniste s'exprime bien dans l'idée d'« accompagnement technologique » (*technology accompaniment*) proposée par Peter-Paul Verbeek (2010). Cet accompagnement assigne au philosophe la charge d'aider au design de « bonnes médiations techniques » et institue la philosophie comme pratique du « design par d'autres moyens »⁵. L'idée d'accompagnement technologique résume très bien l'ambition d'une thèse CIFRE en philosophie et, plus particulièrement, de notre propre recherche doctorale.

0.3 La robotique sociale et ses domaines d'application

La robotique sociale - ou « sociobotique »⁶ - est une notion d'apparition récente suscitant aujourd'hui un intérêt croissant, à la croisée de plusieurs disciplines relevant de l'ingénierie aussi bien que des sciences humaines et sociales. Au-delà de la pluralité des approches, des finalités et des acteurs concernés, la robotique sociale désigne un ensemble de recherches de

⁵ Nous formulons cette idée en clin d'œil à Verbeek qui, dans son ouvrage *What things do* (2005), présente l'ingénierie et le design comme de l'« éthique par d'autres moyens » (*ethics by other means*).

⁶ Nous proposons ce néologisme en guise de synonyme de « robotique sociale ».

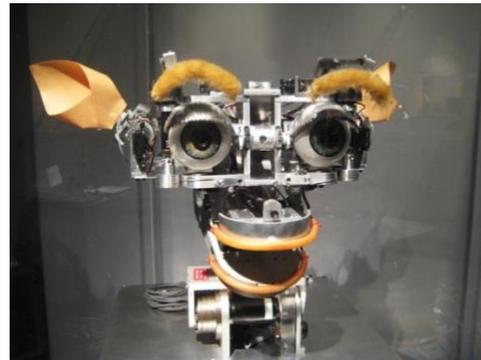
nature à la fois technologique et épistémologique. Technologique, d'une part, via la conception de robots capables d'interagir socialement avec les êtres humains dans des environnements variés ; épistémologique, de l'autre, par l'étude de ces interactions humain-robot (IHR). Concrètement, ces interactions peuvent être mises en au service d'objectifs scientifiques, thérapeutiques, ludiques ou serviciels au sens large.

0.3.1 La robotique sociale scientifique

Évoquons, en premier lieu, les applications proprement scientifiques. Les robots sociaux sont utilisés pour examiner, valider et affiner des théories concernant le développement social et biologique, la psychologie, la neurobiologie, la communication émotionnelle et non-verbale, les interactions sociales ou encore la robotique elle-même. Par exemple, *Infanoid*, développé en 2001 par la firme japonaise *Vital Corporation*, était un robot humanoïde utilisé pour enquêter sur les mécanismes sous-jacents au développement social chez les jeunes enfants.



Infanoid – Pr. Hideki Kozima©



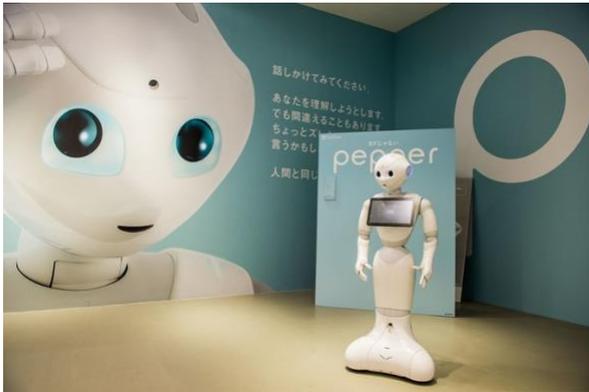
Kismet – arcticpenguin©

Kismet (MIT, 1997), quant à lui, a été utilisé pour étudier les interactions sociales enfant-soignant (tour de parole, échange émotionnel de basse intensité, développement d'actes communicationnels significatifs). Un autre exemple, beaucoup plus récent, est celui du célèbre *NAO*, de *Softbank Robotics*, dont la portée épistémologique s'étend de la recherche sur l'équilibre aux ISHR en passant par la motricité, la vision ou encore le langage.

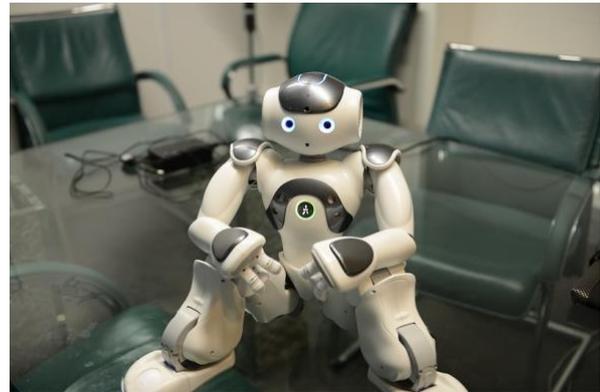
0.3.2 La robotique sociale de services

Outre la fonction épistémologique des robots sociaux, le domaine de l'assistance et du service traditionnellement associé à la robotique représente un champ de développement souvent

présenté comme prometteur pour la sociobotique. L'exemple le plus connu est celui de Pepper développé lui aussi par SoftBank Robotics après NAO. Bien que sa production soit aujourd'hui arrêtée, Pepper reste toujours présents dans de nombreux magasins SoftBank Mobile au Japon et dans d'autres boutiques à travers le monde pour accueillir, informer et divertir les clients.

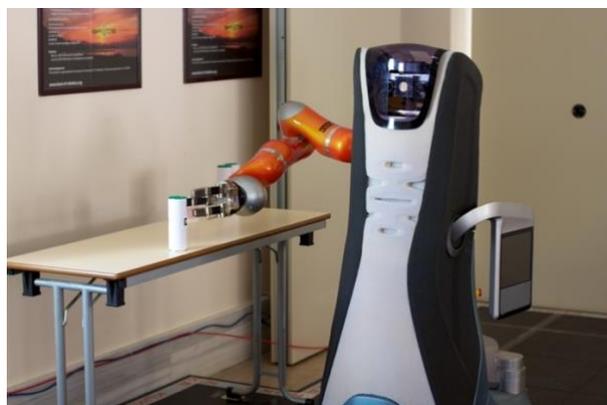


Pepper – Héctor García©



NAO – Stephen Chin©

En France, la SNCF a tenté l'expérience en implantant Pepper de décembre 2015 à mars 2016 dans trois gares des Pays-de-la-Loire. Il avait pour mission d'accueillir les voyageurs, de les aider dans leur recherche d'informations sur la circulation des trains (horaires et itinéraires) et d'effectuer des enquêtes de satisfaction. Il était également chargé d'animer la gare et de proposer de jouer avec lui en cas d'attente au guichet. Ce test eut également lieu dans un supermarché Carrefour de Seine-et-Marne où, de façon similaire, l'objectif était d'« enquêter sur la satisfaction client, [de] divertir les enfants et [d']offrir des renseignements sur la carte de fidélité » (Brigand, 2015).



Care-O-Bot 4 – Jiuguang Wang©

Évoquons également, comme second exemple, la 4^e version du Care-O-bot, conçu par l'institut allemand Fraunhofer. Ce robot « gentleman » modulaire et polyvalent est pensé pour intervenir en milieu domestique, commercial, hospitalier ou industriel. Tour à tour serveur, point d'information mobile, assistant médical, livreur ou porteur de charges (usines, hôtels), le Care-O-bot répond au concept d'« aidant universel » (*universal helper*)⁷.

0.3.3 La robotique sociale dite « personnelle » ou « domotique »

Dans cette catégorie, on trouve des robots de divertissements et de compagnie destinés à intégrer le milieu domestique ou tout environnement apparenté. Buddy, de la société française Blue Frog Robotics est un bon exemple de ce genre de robot présentés tantôt comme des compagnons, tantôt comme des assistants. A en croire sa présentation sur le site de la société, « [Buddy] ne se contente pas d'apporter une présence dans la maison, il est un véritable couteau suisse qui possède de nombreuses autres fonctionnalités vraiment utiles.⁸ »



Buddy – Pierre Metivier©

⁷ Voir, sur ce point, le communiqué de presse de l'institut Fraunhofer daté du 15 janvier 2015 : « Robots as multifunctional gentlemen ».

https://www.ipa.fraunhofer.de/en/press-media/press_releases/2015-01-15_Robots-as-multifunctional-gentlemen.html

⁸ Voir <https://buddytherobot.com/fr/buddy-le-robot-emotionnel-famille/> (site consulté le 28/09/2021)

En tant que « robot compagnon émotionnel », l'artefact peut simuler un large éventail d'émotions évoluant tout au long de la journée au gré de nos interactions avec lui ou, au contraire, de l'absence d'interaction. Ses fonctionnalités peuvent être réparties en quatre catégories : le divertissement éducatif (jeux interactifs, exercices de calculs et de mémoire, initiation à la programmation), l'assistance pratique (rappel d'agenda, minuteur, météo, dictionnaire etc.) ou thérapeutique au sens large (rappel de prise de médicament aux séniors, surveillance) et, pour finir, le lien social (divertissement en famille, téléprésence etc.). Enfin, Buddy est présenté comme une « plateforme modulaire » pouvant être indéfiniment enrichie d'accessoires et d'applications diverses.

Dans un registre différent, citons le projet AIDA (Affective, Intelligent Driving Agent) lancé en 2009 par la branche américaine de Volkswagen avec le MIT. Il s'agit d'un robot « compagnon de conduite » (*driving companion*) installé sur le tableau de bord d'un véhicule. AIDA reconnaît les expressions faciales, les humeurs, les gestes et le ton du conducteur. Il est ainsi capable de lui répondre intelligemment, de façon contextualisée. Le robot est doté d'une série d'expressions faciales, peut tourner la tête, sourire et faire des clins d'œil. D'un point de vue pratique, AIDA est pensé pour transmettre des informations sur le trafic en temps réel et les conditions climatiques, mais aussi sur l'infrastructure routière elle-même. A l'instar d'un GPS classique, il peut également indiquer quel est le chemin le plus court ou le moins embouteillé pour arriver à destination.

0.3.4 La robotique sociale thérapeutique

Nous l'avons dit, les robots sociaux sont particulièrement présents dans le champ thérapeutique. L'un des plus connus est le blanchon Paro, un « robot d'aide mentale » selon son concepteur Takanori Shibata (Dumouchel, Damiano, 2016). Doté d'un duvet très doux et équipé de différents capteurs, Paro réagit par des mouvements et des petits cris aux caresses, à la lumière ou encore aux voix. Le blanchon artificiel a pour objectif de favoriser la motivation et la relaxation des patients, de diminuer leur stress et d'aider à développer les interactions patient-soignant (rôle médiateur). Il peut également être utilisé à des fins purement récréatives par des enfants ou des adultes.



Paro – Paul Allais©

De nombreuses études dépeignent Paro comme un allié bénéfique pour la santé mentale et physique des personnes âgées dépendantes (amélioration des capacités cognitives et émotionnelles, gestion du stress, socialisation). Citons également KASPAR, conçu par Kerstin Dautenhahn et ses collaborateurs à l'université du Hertfordshire. Il s'agit d'un robot téléopéré humanoïde de la taille d'un jeune enfant, utilisé à la fois comme médiateur social entre les enfants atteints de troubles du spectre autistique (TSA) et leurs différents partenaires d'interaction (thérapeutes, parents, autres enfants autistes ou non) et comme outil éducatif de stimulation du développement social chez ces mêmes enfants.



KASPAR – Loz Pycok©

KASPAR a vocation à enseigner des aptitudes sociales qui ne se développent pas ou mal chez les enfants atteints de TSA (compréhension des émotions d'autrui, réaction adéquate à leurs manifestations affectives, expression de leurs propres émotions, respect du tour de rôle en jouant ou en interagissant verbalement, coopération etc.) Le choix de l'apparence physique de KASPAR n'est, bien entendu, pas anodin. Le robot ressemble à un enfant, mais son visage en silicone offre une expressivité minimale afin de ne pas complexifier l'apprentissage émotionnel des jeunes enfants. Cette simplicité leur laisse, par exemple, une grande liberté d'interprétation pour ne pas surcharger cognitivement les interactions avec KASPAR, ce qui

défavoriserait l'apprentissage. Pour être très précis, cette pauvreté expressive concerne l'ensemble du corps de l'artefact et se voit investie d'une de ses principales fonctions, à savoir la réassurance de l'enfant.

Intéressons-nous à présent à la robotique sociale dans le contexte particulier de Spoon, le partenaire industriel notre thèse CIFRE, et plus précisément à la genèse de celle-ci.

0.4 La robotique sociale dans le contexte de la thèse CIFRE

Spoon est une entreprise présidée par Jérôme Monceaux, ancien co-fondateur d'Aldebaran Robotics, pionnier français de la robotique sociale. L'activité principale de Spoon concerne, selon ses statuts, « la conception, le développement, la production, le design, le marketing et la vente de robots interactifs programmables et autonomes ». A sa création, la firme s'est donc pensée comme entreprise de « robotique interactive », notion alors jugée plus claire que celle de « robotique sociale ». Un robot interactif, comme son nom l'indique, est un robot doté d'interactivité, c'est-à-dire capable d'interagir. Un tel choix soulève cependant deux difficultés. La première, de nature taxonomique, est que la notion de « robotique interactive » se présente comme une notion de R&D (recherche et développement) dont l'ampleur permet, dans l'absolu, d'y intégrer les robots sociaux, mais qui, en pratique, renvoie surtout aux robots industriels dits collaboratifs (ou « cobots » pour *collaborative/cooperative robots*) dont l'interactivité n'est pas tant « sociale » qu'axée sur la manipulation d'objet. La seconde difficulté vient du fait qu'en substituant la notion de robot interactif à celle de robot social à des fins de clarification, on s'interdit, par occultation, de penser le problème légitime de leur socialité.

Disons-le sans détour, ce n'est pas tant la question de la socialité des robots que celle de leur altérité qui intéressa Jérôme Monceaux lors de la construction du projet CIFRE. Son intérêt pour le thème répond à un double enjeu de conception et de conceptualisation. Côté conception, l'ingénieur co-fonde Spoon sur la base d'un constat : l'intérêt des utilisateurs pour les robots sociaux est évanescent. Si la question de la « valeur d'usage » (Marion, 2017) des sociobots se pose bien évidemment, il existe selon lui un problème de design plus profond : l'interactivité des sociobots est généralement de mauvaise qualité car elle ne s'inspire pas assez du principe de base de l'intersubjectivité qu'est le sentiment d'exister aux yeux de

l'autre. Monceaux ancre cette idée dans une expérience vécue chez Aldebaran. Un jour, en entrant dans une salle où était disposée une dizaine de NAO, il déclencha la localisation sonore des artefacts et vit leur tête se tourner simultanément vers lui. L'ingénieur dit avoir compris grâce à cette expérience que l'enjeu fondamental de la robotique sociale

« n'est pas d'essayer de mettre de la "vie" dans les robots comme [on tente souvent de le faire], mais plutôt de refléter l'existence de l'utilisateur dans les robots. »⁹

Ce à quoi il ajoute que

« [l]e fait de ne pas savoir si on "existe" aux yeux de la machine perturbe considérablement nos interactions avec elle et ça finit par provoquer un décrochage, un désintérêt. »

Érigée en véritable heuristique de design, cette expérience aura permis de fixer l'un des principes de développement majeurs de Spoon, à savoir le soin accordé à la modélisation du « regard » de ses artefacts permettant de créer une sensation de croisement perceptif (Lenay, 2010).



BigSpoon

Du côté de la conceptualisation, cet événement est indissociable d'une question plus proprement philosophique quant au sens des regards en question et, plus largement, de l'activité sociobotique : de qui ou de quoi a-t-il fait l'expérience ? Que signifie le sentiment d'exister aux yeux d'une machine ? Comment un tel regard affecte-t-il l'idée que l'on se fait de la machine en question ? L'intérêt du dirigeant pour le thème philosophique de l'altérité émane de ces interrogations. Car pour parler de « social » ou de « socialité », ne faut-il pas

⁹ Notes issues de nos entretiens.

d'abord qu'il y ait un ou des autres, c'est-à-dire de l'altérité ? Mais que peut bien signifier l'altérité lorsqu'on l'applique aux artefacts ? Ces questions font écho au fort enjeu de conceptualisation auquel l'entrepreneur était confronté au moment de la mise en place du dispositif CIFRE et qui le justifie en tant que projet de recherche.

0.5 Positionnement scientifique

0.5.1 Le tournant chosique de la philosophie des artefacts techniques

Prendre la question de l'altérité des sociobots au sérieux demande de ne pas les subsumer dans « la » technique, mais au contraire de les étudier dans leur singularité artefactuelle. Renoncer aux points de vue généraux et englobant sur *la* technique, pour penser philosophiquement le fait technique à partir de la très grande diversité des artefacts que nous portons à l'existence et utilisons, telle est au demeurant l'orientation prise par la philosophie des techniques ces deux dernières décennies. Les années 2000 ont en effet été marquées par l'émergence d'un « tournant chosique » (*thingly turn*) (Verbeek, 2005), caractérisé par l'étude, notamment axiologique, d'artefacts très précis (citons, par exemple, [Verbeek, 2005] pour son analyse de l'échographe, [Kudina & Verbeek, 2019] sur les Google Glass ou encore [Bergen & Verbeek, 2020] à propos d'Habitaca, une *to-do list* ludique) et par sa vocation interventionniste en vue de développer une « éthique by design » (Fischer, 2020). Il s'agit, pour les contributeurs du corpus, de construire une philosophie capable d'*accompagner* la conception des technologies, mais aussi leur utilisation (Verbeek, 2005, 2010, 2011, 2013 ; Dorrestijn, 2012) afin de générer de « bonnes » médiations techniques. Ce tournant chosique, qui est aussi un tournant éthique (*Ethical turn*), a suscité une littérature foisonnante rassemblée sous le nom de *philosophy of technical artefacts* (PTA). Ces nombreuses contributions sont dédiées à l'analyse du concept d'artefact et se répartissent en deux principaux programmes de recherches : le *Dual Nature of Technical Artifacts* (DN TA) d'une part ; la postphénoménologie et sa théorie des médiations technologiques (TMT) de l'autre. Dans quelle mesure ces développements philosophiques récents autour du concept d'artefact technique apportent-ils des éléments de réflexion susceptibles d'enrichir notre compréhension d'artefacts aussi singuliers que les sociobots ?

0.5.2 La métaphysique analytique du Dual Nature of Technical Artifacts

Pour les auteurs du programme de recherche *The Dual Nature of Technical Artifacts*, un artefact est un objet dual, issu de la rencontre entre une nature structurale, matérielle qui permet un *fonctionnement* et une nature intentionnelle qui, elle, attribue des *fonctions*. En d'autres termes, un artefact est la matérialisation d'une intention humaine dans une structure physique qui, en fonctionnant, réalise une ou plusieurs fonctions (Houkes & Vermaas, 2004 ; Houkes & Meijers, 2006 ; Kroes & Meijers, 2006). Dès lors, l'objectif du DNTA est de comprendre le phénomène générateur d'artefact qu'est l'attribution de fonction à une structure physique, autrement dit la survenue d'une intention à une matérialité.

« [Le DNTA vise à] fournir une analyse conceptuelle profonde de la notion de fonction technique ; à comparer cette explication avec celles données à la fonction dans d'autres domaines, particulièrement en biologie et dans les sciences sociales ; à développer une explication de la relation entre fonction technique et structure physique, notamment de la façon dont cette celle-ci contraint celle-là ; à proposer une explication des aspects intentionnels des fonctions techniques et de leur relation avec l'intentionnalité et l'action des concepteurs et des utilisateurs ; à examiner les conséquences épistémologiques de cette dimension intentionnelles des fonctions ; à enquêter sur les jugements normatifs qui s'appliquent au fonctionnement des artefacts et sur l'origine de cette normativité ; à clarifier les implications ontologiques qu'entraînent la description et l'utilisation des artefacts ; à développer une explication du processus de conception des artefacts techniques et de la façon dont la description structurelle de l'artefact-comme-objet-physique et la description fonctionnelle de l'objet-comme-artefact-intentionnellement-formé sont combinées durant ce processus ; et, enfin, à examiner comment les explications techniques dans la conception sont reliées à différents types d'explication qui sont généralement distingués dans la philosophie des sciences » (Kroes & Meijers, 2002 p.12).

Là où une approche fonctionnaliste « naïve » stipulerait qu'un artefact se définit par une fonction propre, c'est-à-dire par la fonction que son concepteur a en tête au moment de la conception, le DNTA rétorque que cette idée n'est pas en mesure d'expliquer adéquatement la nature des artefacts. Pieter Vermaas et Wybo Houkes proposent alors une conception

phénoménologique alternative de l'artefact basée sur l'action (2004). Selon eux, la ou les fonctions d'une technologie se définissent par ce que nous faisons concrètement d'elle, c'est-à-dire par l'usage effectif que nous en avons. Trois grands types d'usages peuvent alors être énoncés : l'usage standard, l'usage alternatif et l'usage inapproprié. Pour l'approche fonctionnaliste naïve, est standard tout usage qui implique la fonction propre de l'artefact. Les usages alternatifs et inappropriés, quant à eux, se basent sur une de ses *fonctions accidentelles*. Le problème de cette théorie est qu'elle ne fait ni ne permet de faire aucune distinction entre les fonctions accidentelles, ce qui, pour Houkes et Vermaas, représente un appauvrissement phénoménologique et conceptuel patent. De plus, l'approche naïve se révèle inadaptée à la réalité dynamique des usages. Un usage alternatif peut se standardiser et, ce faisant, contrer l'idée qu'un artefact n'est que la traduction matérielle et fonctionnelle de l'intention initiale qui justifie sa création. Un exemple célèbre est celui de l'aspirine dont la fonction antalgique originelle est désormais concurrencée par sa capacité à fluidifier le sang et, ainsi, à prévenir la formation de caillots. Face à ces lacunes, Houkes et Vermaas proposent de concevoir l'artefact sous l'angle des actions qu'il permet d'accomplir en le liant à la notion de *plan d'usage (use plan)*. Ce dernier est défini comme « une série d'actions dans laquelle la manipulation de [l'artefact] contribue à la réalisation d'un but donné » (Houkes & Vermaas, 2004, p.57). Dans cette perspective, ce n'est donc pas tant la fonctionnalité que la *rationalité* qui caractérise l'intention informatrice de l'artefact. Ainsi, bien qu'une chaise soit « faite pour » s'asseoir, je peux très bien l'utiliser comme escabeau pour atteindre des endroits haut placés. Dans ce cadre d'analyse, la fonction devient une capacité réelle ou supposée¹⁰ d'un artefact ayant un certain statut préférentiel dans un certain contexte d'action et de croyance ; elle dérive de l'action et n'est pas une caractéristique *essentielle* de l'artefact mais une composante *relationnelle*.

0.5.3 L'approche postphénoménologique des médiations technologiques

Le second courant fondateur de la philosophie des artefacts techniques est celui de la postphénoménologie et de sa théorie des médiations technologiques (TMT). Celles-ci proposent une réinterprétation de la phénoménologie comme pratique décrivant la relation

¹⁰ Pour illustrer une fonction supposée, Houkes et Vermaas donnent l'exemple du fer à cheval auquel on attribue la capacité repousser les mauvais esprits.

entre l'homme et le monde plutôt que comme description des choses elles-mêmes (bien que cela passe par le fameux retour « aux choses même » caractéristique du geste husserlien initial). Principalement représentée par le philosophe néerlandais Peter-Paul Verbeek, la TMT pense les artefacts comme des médiateurs notre façon d'entrer en relation avec le monde via des *pratiques* et des *perceptions* : « Les choses [...] ne sont pas des « intermédiaires » neutres entre les humains et le mondes, mais des médiatrices : elles médiatisent activement cette relation. » (Verbeek, 2005, p.114). Autrement dit, les techniques participent au dévoilement du monde. Selon Verbeek, ce statut de médiateur se divise en deux dimensions essentielles : l'une herméneutique en ce qu'ils favorisent des interprétations particulières du monde via notre perception, l'autre existentielle par les actions et la conduite de la vie qu'ils rendent possibles voire incitent à adopter.

Parler du pouvoir médiateur des artefacts à l'échelle de nos pratiques et de nos perceptions revient à leurs accorder un rôle constitutif au sein de l'expérience humaine. Selon Don Ihde, fondateur de la postphénoménologie, l'expérience est composée par l'association entre « microperception » et macroperception » (1990). La première renvoie au domaine des perceptions sensorielles, la seconde au contexte d'interprétation des microperceptions, c'est-à-dire à ce qui donne du sens aux choses perçues et qu'on désigne souvent par la vague notion de culture. Cette distinction reste toutefois purement analytique : toute activité perceptive implique simultanément une dimension micro et une dimension macroperceptive (en témoigne, comme le suggère [Verbeek, 2005, p.123] le double sens du mot « voir » qui peut désigner le fait de percevoir par la vue et/ou celui de comprendre quelque chose). Ceci explique pourquoi les relations mondaines tissées à travers les artefacts ne sont pas uniquement pratiques, mais aussi herméneutiques. Ihde identifie quatre catégories de relations techniques au monde :

- Les « relations d'incarnation » où l'artefact fait partie intégrante de l'expérience du monde qu'il permet de dévoiler, comme avec les lunettes de soleil par exemple ;
- Les « relations herméneutiques » où l'accès au monde se fait via un artefact qui n'est pas transparent et qui en fournit une représentation. La relation herméneutique implique alors une interprétation : le monde n'est pas découvert à travers mais *au moyen* de l'artefact. Le thermomètre et l'échographie illustrent bien cette idée en ce qu'ils

fournissent des représentations cryptiques du réel et supposent ainsi un effort de décryptage ;

- Les « relations d'altérité » au sein desquelles l'artefact et le monde se confondent et où l'expérience du monde se résume à l'expérience de l'artefact. Les ordinateurs, les pianos ou encore les distributeurs de billets en sont de bons exemples ;
- Les « relations d'arrière-plan » où la technologie n'est pas expérimentée directement dans sa matérialité, mais où celle-ci façonne le contexte de notre expérience sans que l'on s'en rende compte. Les thermostats ou les bornes wi-fi sont de bons représentants de ces relations.

Définir l'artefact comme un médiateur et un co-constituant du monde, c'est lui reconnaître une agentivité, notamment morale – c'est-à-dire une capacité de contribution active et non seulement passive au domaine moral –, que la pensée moderne évacue en le reléguant dans le champ inerte de l'objectalité : les artefacts (nous) agissent autant qu'ils sont agis. A cette agentivité (morale) des techniques correspond une responsabilité des concepteurs qui ne peuvent plus être considérés comme de simples « exécutants ». Comme le dit Verbeek : les concepteurs font de « l'éthique par d'autres moyens »¹¹ (*ibid.*, p.212), et ce au sens où ils produisent des entités qui ne sont pas neutres axiologiquement car elles participent activement, via leurs médiations, à la vie éthique. La « productivité ontologique » (Guchet, 2017) des artefacts est aussi une productivité morale : ils enrichissent notre répertoire d'action, invitent voire incitent à agir de telle ou telle façon ; ce faisant, ils contribuent à l'évolution et/ou à la stabilisation de nos cadres de pensée morale. Autrement dit, en médiatisant l'action humaine, les technologies fournissent des réponses matérielles aux grandes questions éthiques du « comment vivre » et du « comment agir ».

0.5.4 La philosophie des techniques au défi de la robotique sociale

La présente recherche partage l'intérêt du tournant chosique pour les étants et se propose, comme nous l'avons dit, d'étudier la robotique sociale à partir des robots sociaux eux-mêmes.

¹¹ « *Designers engage in ethics by other means* »

Les deux grands courants que nous venons de présenter permettent-ils toutefois de saisir le sens d'être du sociobot tel que nous l'avons problématisé ?

En tant qu'il s'intéresse aux artefacts sous l'angle d'une rencontre entre matérialité et fonction intentionnelle, le DNTA ne semble pas structuré pour répondre à la question que nous posons aux robots sociaux. En effet, celle-ci porte sur le sens d'être de leur socialité et, plus particulièrement, de leur socialité comme phénomène moral d'altérité. Notre problème central n'est donc pas celui des fonctions matérialisées (ou de la matière fonctionnalisée), mais celui de la signification du social comme « apparaître d'un autre ».

La postphénoménologie, quant à elle, peut offrir de solides appuis à notre réflexion, notamment à travers son concept d'*alterity relation*. En l'état, celui-ci reste toutefois trop général et formel pour décrire finement l'altérité des sociobots. L'artefact qui nous entraîne dans une relation d'altérité est mondifié au sens où l'utilisation que nous en faisons l'institue comme *étant-monde*, un étant qui, pour la durée de la relation, résume le monde. En un sens, ici, être autre c'est être monde. Mais qu'est-ce que cela signifie quand cette mondification prend la forme d'un « tu » et non d'un « cela » ? Dans la perspective des relations d'altérité, les exemples cités par les postphénoménologues sont souvent de nature utilitaire et/ou interfaciale, à l'instar de l'ordinateur, du portable ou du distributeur de billets. Ces exemples d'« autres artefactuels » relèvent du mode d'être qu'Heidegger nomme *Zuhandenheit* au sens où ils s'effacent plus ou moins totalement de notre conscience au profit de leur fonction. Or, le propre du sociobot, comme nous le verrons, est de ne surtout pas s'effacer, au point d'avoir, dans certains cas, l'« acte de présence » pour fonction (Dumouchel & Damiano, 2016). Cette présence, ce non-retrait ne les renvoie pas pour autant au second mode d'être heideggérien qu'est la *Vorhandenheit*, car ils ne se contentent pas de se trouver là, d'être des choses-là-devant. Dès lors, que sont-ils ?

Notre recherche propose d'examiner cette question et, plus largement, celle de la socialité des robots en deux grands moments : une première partie dédiée à l'état de l'art critique de la robotique sociale en lien avec le problème de la socialité des robots et une seconde consacrée à l'exploration du concept d'altérité en sociobotique. La première partie comprend les chapitres 1 à 3 ; la seconde les chapitres 4 et 5.

Dans le premier chapitre, nous interrogeons la notion de robot afin de comprendre ce que le robot social partage ou non avec le domaine auquel il semble nommément rattaché.

Le second chapitre est décomposé en deux temps : le premier est consacré à l'analyse des principales définitions de la robotique sociale et de ses principes fondateurs ; le second, quant à lui, est dédié aux trois premières réponses que nous apportons à la question du sens de la socialité des robots. Nous verrons que ces réponses, respectivement centrées sur les notions d'interaction, de fonction et d'interactance, dépendant du constat que le « social » semble être lié au robot sur le mode de la propriété.

Le troisième chapitre est dédié à l'examen des trois dimensions constitutives de l'éthique robotique, à savoir l'éthique des personnes qui conçoivent et utilisent les robots, l'éthique des systèmes intégrés aux robots et la question du statut moral des robots. Nous proposons de qualifier respectivement ces trois dimensions d'éthique « autour », « dans » et « envers » le robot. Dans la dernière partie du chapitre, nous examinons une quatrième dimension, peu discutée dans la littérature : celle de l'éthique « avec » le robot, centrée sur la question phénoménologique de l'apparence de l'artefact et des relations qu'elle suscite.

Le quatrième chapitre aborde la robotique sociale à partir du contexte de Spoon et de la présentation détaillée des fondements conceptuels de son activité. La question de l'altérité robotique y est examinée à partir du thème du regard, central dans l'approche de l'entreprise, et de la notion d'« animalité transpèce » proposé par Dominique Lestel (2021), dont nous verrons qu'il fait écho aux réflexions et partis pris de Spoon.

Le cinquième et dernier chapitre se concentre sur l'analyse de la socialité robotique à l'aune du concept d'altérité. Nous commençons par identifier les rares occurrences du mot dans la littérature de robotique sociale et dans la philosophie des techniques pour asseoir notre constat : dans un cas comme dans l'autre, la notion d'altérité n'est ni conceptualisée, ni appliquée aux robots eux-mêmes. Seul le concept d'*alterity relation* proposé initialement par Don Ihde offre une première prise intéressante pour penser ce que pourrait être l'altérité de l'artefact. En l'état, cette prise reste toutefois largement insuffisante. La deuxième partie du chapitre est donc consacrée à l'exploration de ce qui, dans l'apparaître du robot, donne accès au sens d'être de son altérité. Soucieux de ne pas revenir à la dichotomie sujet/objet que la robotique sociale met ouvertement en crise, nous analysons ce sens d'être sur le mode de l'effectivité. Le sociobot se présente alors comme entité génératrice d'« effets d'altérité » dont nous proposons de situer l'origine dans l'apparaître spécifique de l'artefact. A l'issue d'une description « ontophanique » (Vial, 2013) de cet apparaître visant à dégager les caractéristiques essentielles du phénomène sociobotique, nous observons que l'artefact se

présente à nous « comme si » (Dennett, 1990) il était un vivant vulnérable. Nous proposons alors de penser l'effet d'altérité comme effectivité *morale* questionnant 1) notre responsabilité à l'égard de l'artefact et 2) à l'égard des vivants en général que le sociobot représente symboliquement. La dernière partie du chapitre est consacrée à l'interprétation de la socialité comme relation de responsabilité envers l'autre, c'est-à-dire à la quatrième grande réponse que nous proposons d'apporter à notre question directrice.

PARTIE I

DU ROBOT AU ROBOT SOCIAL : ÉTAT DE L'ART CRITIQUE DE LA SOCIOBOTIQUE

Tout travail au sujet des robots en général et des robots sociaux en particulier commence avec la difficulté comprendre ce qu'ils sont. Parmi tous les adjectifs venant préciser les grands domaines d'application de la robotique – industrielle, médicale, militaire/d'intervention, éducative, transport, domestique/servicielle -, celui de « social » est peut-être le plus déroutant. Nous l'avons dit, l'idée même de sociobot présente la difficulté d'être composée de deux termes particulièrement opaques. Malgré cela, nous parlons aujourd'hui de robots sociaux et de robotique sociale comme si l'association de ces deux termes allait de soi. D'emblée, le premier problème qui se pose à nous est le suivant : faut-il appréhender le robot social comme étant avant tout un robot auquel viendrait s'ajouter, comme un supplément d'âme, la dimension sociale ou, au contraire, faut-il l'aborder comme un tout, une unité incompréhensible dès lors qu'on la décompose ? Répondre à cette question implique d'interroger au préalable la notion de robot.

Chapitre 1 - Qu'est-ce qu'un robot ?

« Qu'est-ce qu'un robot ? ». Cette question très générale appelle de prime abord une réponse d'ordre définitionnel. Définir, c'est *cerner* une chose, délimiter les frontières à partir desquelles on entre et on sort de cette chose. Or, si la valeur pratique d'une définition est indéniable, on peut se demander dans quelle mesure définir un objet permet véritablement de le comprendre. Une définition amorce une discussion en la clarifiant, mais ne la clôture pas. Voilà pourquoi la philosophie part généralement des définitions pour mieux s'en affranchir au terme de son enquête. En dernière instance, l'exercice définitionnel n'a d'intérêt philosophique qu'en ce qu'il marque la première étape vers la conceptualisation. Car c'est le concept en tant qu'il dégage les structures essentielles du réel qui intéresse la philosophie, et non les définitions. Sur ce point, l'objectif de notre enquête est donc des plus clairs : tirer les différentes ficelles d'une définition du robot pour en faire émerger le concept, c'est-à-dire la *com-préhension*.

Définir ce qu'est un robot n'en demeure pas moins délicat tant l'artefact se révèle protéiforme et ce, aussi bien dans ses manifestations techniques que discursives. Le vocable englobe une classe disparate d'appareils qui, d'ailleurs, ne correspondent que très rarement à l'image populaire de la machine humanoïde. Comme nous le verrons, il n'existe pas de définition officielle du robot, mais bien un ensemble de caractérisations technico-fonctionnelles relativement convergentes dont il s'agira de questionner la signification. Pour l'heure, nous débuterons cette discussion avec ce que les philosophes Luisa Damiano et Paul Dumouchel qualifient de « définition originale du terme », à savoir :

« un travailleur artificiel, conçu comme un appareil qui a sa propre source d'énergie, qui travaille pour nous, et qui est autonome dans une certaine mesure » (Dumouchel & Damiano, 2016, p.7).

Le robot semble donc émerger au croisement de trois notions philosophiquement aussi profondes que problématiques : l'artifice, le travail et l'autonomie. Pour l'heure, notons simplement que cette émergence ne se manifeste pas uniquement dans la technique, loin de là. Le robot est un exemple archétypique d'entité technodiscursive au sens où l'on ne peut

dissocier sa réalité technique des imaginaires en tout genre dont il est porteur. Autrement dit, le robot existe aussi bien dans la matérialité technique de l'artefact que dans l'immatérialité discursive du personnage. C'est d'ailleurs dans le champ littéraire du début du XXe siècle qu'il est pour la première fois nommé en tant que tel. Il apparaît alors comme personnage symbolique introduisant une rupture à la fois pratique et discursive avec son « aîné », l'automate classique. Voilà pourquoi, avant de l'examiner dans ses manifestations techniques et fonctionnelles contemporaines, nous voudrions appréhender ce « travailleur artificiel » dans ses dimensions historique, mythique et symbolique.

1.1 Le robot est un mythe

1.1.1 Le robot n'existe pas

Dès l'introduction de sa « petite thèse » intitulée *Du mode d'existence des objets techniques*, Gilbert Simondon se livre à une virulente critique de l'idée de robot :

« Nous voudrions précisément montrer que le robot n'existe pas, qu'il n'est pas une machine, pas plus qu'une statue n'est un être vivant, mais seulement produit de l'imagination et de la fabrication fictive, de l'art d'illusion. Pourtant, la notion de machine qui existe dans la culture actuelle incorpore dans une assez large mesure cette représentation mythique du robot. Un homme cultivé ne se permettrait pas de parler des objets ou des personnages peints sur une toile comme de véritables réalités, ayant une intériorité, une volonté bonne ou mauvaise. Ce même homme parle pourtant des machines qui menacent l'homme comme s'il attribuait à ces objets une âme et une existence séparées, autonomes, qui leur confère l'usage de sentiments et d'intentions envers l'homme » (Simondon, 2012, p.11).

Le robot est pour Simondon l'exemple même du mirage provoqué par l'ostracisme culturel dont la technique fait l'objet. L'expulsion de celle-ci en dehors des sphères de la culture, de ce qui mérite d'être connu de tous, génère une incompréhension et une ignorance des réalités techniques conduisant à croire en des mythes comme celui du robot. Le philosophe note d'ailleurs que cette attitude exclusive repose sur une contradiction : d'un côté, la technique est limitée à la pure instrumentalité résultant d'« *assemblages de matière*, dépourvus de vraie

signification » (*ibid.*), de l'autre, on lui attribue des intentions qu'elle ne saurait avoir en la soumettant à une forme de prosopopée dont le robot incarne la figure ultime. Non content d'être une illusion du commun, le robot repose également sur une erreur des « idolâtres de la machine [qui] présentent en général le degré de perfection d'une machine comme proportionnel au degré d'automatisme » (*ibid.*).

« Or, en fait, l'automatisme est un assez bas degré de perfectionnement technique. Pour rendre une machine automatique, il faut sacrifier bien des possibilités de fonctionnement, bien des usages possibles. L'automatisme, et son utilisation sous forme d'organisation industrielle que l'on nomme *automation*, possède une signification économique ou sociale plus qu'une signification technique. Le véritable perfectionnement des machines, celui dont on peut dire qu'il élève le degré de technicité, correspond non pas à un accroissement de l'automatisme, mais au contraire au fait que le fonctionnement des machines recèle une certaine marge d'indétermination. C'est cette marge qui permet à la machine d'être sensible à une information extérieure » (*ibid.*, p.11-12).

C'est là le cœur de la critique simondonienne du robot : la confusion entre automatisme et autonomie pousse illusoirement à concevoir le premier comme exprimant un haut degré de technicité. Or, pour Simondon, automatisme et autonomie sont deux réalités antinomiques :

« ou bien le robot est une machine automatisée et sa finalité découle de la programmation que lui ont assignée ses concepteurs, ou bien il dispose d'une "marge d'indétermination" pour recevoir et intégrer l'information issue de son milieu associé et s'y adapter, et il est possible de communiquer avec lui et de le contrôler » (Bontems, 2018, p.24).

En somme, c'est l'automate qui existe techniquement, et non le robot qui n'est rien d'autre qu'une fable procédant d'une confusion dommageable. Car faire du robot un modèle suprême de technicité s'apparente à un profond nivellement par le bas du potentiel des techniques. Cette critique s'ancre dans une conception de la technique qui relève de ce que Xavier Guchet (2010) nomme un « humanisme technologique ». Glorifier le robot par l'intermédiaire de l'automatisme, c'est courir le risque d'oublier que

« [ce] qui réside dans les machines, c'est de la réalité humaine, du geste humain fixé et cristallisé en structures qui fonctionnent. Ces structures ont besoin d'être soutenues au

cours de leur fonctionnement, et la plus grande perfection coïncide avec la plus grande ouverture, avec la plus grande liberté du fonctionnement » (Simondon, 2012, p.13).

Magnifier les étants que l'on nomme robots, c'est donc occulter l'humain dans son rôle, non seulement de concepteur, mais aussi de « chef d'orchestre », d'« interprète vivant des machines » (*ibid.*, p.12) selon les mots Simondon. Le robot, par son lien fondamental avec l'automatisme, est une entité hypertélique allant dans le sens de l'abstraction, c'est-à-dire d'une faible cohérence interne.

Contrairement à ce qu'on pourrait croire, les essors consécutifs de l'intelligence artificielle et des apprentissages automatiques (*machine learning*) ne balayent pas totalement la pertinence de l'argument simondonien. Comme le rappellent la plupart des spécialistes, au premier rang desquels Jean-Gabriel Ganascia, Professeur d'informatique à la faculté des sciences de Sorbonne Université, les IA sont aussi efficaces qu'elles sont hyperspécialisées (Ganascia, 2017). Dès lors, si « marge d'indétermination » il y a, elle ne déborde pas l'hypertélie de la machine. En outre, les travaux de recherche estampillés *digital labor* comme ceux d'Antonio Casilli mettent en lumière l'état de grande dépendance dans lequel se trouve l'intelligence artificielle à l'égard de celles et ceux qu'on nomme désormais les « tacherons » du numérique. Dans son livre intitulé *En attendant les robots. Enquête sur le travail du clic*, le sociologue constate que « [l']automation revient à une formule simple : une façade avec un ingénieur qui vante les prouesses de sa machine et une arrière-boutique dans laquelle des travailleurs se tuent à la micro-tâche » (Casilli, 2019, p.137). Ainsi, contrairement à ce que laisse entendre une certaine vulgate technocratique plus que technophile, nous sommes encore loin du mythe de la machine autonome telle qu'on la rencontre dans l'imaginaire fictionnel. Il serait toutefois dommageable de ne pas prendre cet imaginaire au sérieux, tant il a contribué et contribue encore à façonner le phénomène protéiforme qu'est le robot. Arrêtons-nous donc sur l'origine du mythe et sur son contenu.

1.1.2 Le robot est une figure mythique

Simondon ne pourfend pas les étants qu'on appelle robots, mais la conception mystificatrice dont ils font l'objet et dont leur nom est porteur. Et pour cause, ce nom voit le jour sous la plume de l'écrivain tchécoslovaque Karel Čapek qui, en 1920, écrit une pièce de théâtre

intitulée *R.U.R (Rossum's Universal Robots)*. Le terme n'a donc, originellement, rien de technique ni de scientifique : il désigne un personnage symbolique inscrit dans le prolongement moderne de mythes séculaires comme celui du Golem.

Le mot « robot » vient du tchécoslovaque *robota* signifiant « corvée », « travail forcé ». Dans *R.U.R.*, les robots ne sont pas des entités mécatroniques, mais des travailleurs humanoïdes issus de la biologie de synthèse. Il s'agit d'artefacts organiques dépourvus d'émotions (au début de la pièce du moins) et utilisés à des fins de production industrielle. La chose est exposée dès le prologue où sont évoquées diverses affiches publicitaires déclarant :

« Efficacité – économie : les robots de Rossum »

Ou encore :

« Réduisez vos frais de production grâce aux robots de Rossum ! »¹²

Figure du vivant purement instrumental, le robot naît comme symbole bio-artefactuel d'un âge où la science se substitue à Dieu dans l'explication et la conception du monde. A cet égard, *R.U.R.* peut être analysée comme la traduction moderne de la figure mythique du Golem (Munier, 2014). Dans la version pragoise du mythe, le rabbin Loew donne vie à une créature humanoïde préalablement modelée dans l'argile afin de défendre sa communauté contre les attaques antisémites. L'homme anime le Golem en écrivant « EMETH » sur son front, un des noms de Dieu signifiant « Vérité » en Hébreu. Conçue comme arme de défense à la force surhumaine, la créature n'est pas aimée de son concepteur. Le Golem, bien qu'humanoïde, n'est perçu que comme un outil, une entité purement instrumentale, d'aspect répugnant, qui plus est. Ce désamour questionne directement le statut ambigu de la créature située à la frontière entre l'humain et l'outil : le Golem est certes vivant mais, contrairement à l'homme que Dieu investit d'une âme en le créant à son image, le gardien de glaise reste d'humaine facture, dépourvu d'intériorité. Un soir de Shabbat, alors que le rabbin oublie de le désactiver, le Golem est pris de frénésie, forçant son créateur à le détruire. Pour ce faire, il efface l'*aleph* de son front, c'est-à-dire la première lettre du mot « EMETH » pour obtenir « METH », signifiant « mort ».

¹² Čapek, K. (2011). *R.U.R.* Paris : La Différence, p.17.

R.U.R., comme *Frankenstein* avant elle, met en scène cette humaine création se retournant contre ses maîtres. Toutefois, point d'alchimie ni d'incantations magiques chez Čapek (ni chez Mary Shelley). La version moderne du mythe donne à la science le pouvoir d'animation qui, par le passé, était d'origine divine ou mystique. Le vieux Rossum, créateur des premiers robots, « voulait détrôner Dieu par la science. »¹³ Les robots procèdent de l'intelligence humaine concrétisée en biotechnologie. Dans *R.U.R.*, l'intériorité est concédée aux bio-artefacts par l'intermédiaire de la sensibilité. C'est en accédant à la sentience que les robots vont peu à peu développer une conscience qui les conduira à se révolter contre les humains. Il est intéressant de s'arrêter un instant sur la cause du don de sensibilité aux robots. C'est un certain Dr Gall qui, sous la pression d'Hélène, héroïne de la pièce dont il est secrètement amoureux, accorde en cachette la sensibilité des artefacts. En effet, Hélène ne supporte pas l'image que lui renvoient ces entités purement fonctionnelles mises au service exclusif d'une humanité avide et paresseuse. La triste apathie qu'elle lit sur leur visage, amplifiée par leur apparence humaine, heurte violemment son empathie et ses valeurs : ces corps humains dépourvus d'âme sont un affront, non pas à Dieu, mais à l'humanité au sens moral. La réaction de la jeune femme renvoie directement au thème anthropomorphique, dont nous discuterons longuement au cours de notre enquête. Signalons simplement que le personnage d'Hélène soulève d'importantes questions philosophiques comme celle, d'ordre éthique, du rôle de l'empathie dans la conduite de l'action ou celle, bien plus métaphysique, du rapport entre forme physique (*morphê*) et forme essentielle (*eidos/ousia*).

Outre ces premières réflexions, on peut également voir en *R.U.R.* une relecture laïque de le Genèse. Tenté par la femme qu'il aime, l'homme cède à ses supplices et accomplit une action qui conduira l'humanité non pas à l'expulsion du (soi-disant) paradis industriel qu'elle s'est construit, mais bel et bien à l'éradication totale. Ici, point de commandement divin, ni de fruit défendu : « le *désirable* [...] ne descend ni d'un ciel *a priori* ni d'un mont Sinaï de la Morale », selon la belle formule de Dewey (2011, p.36) ; le désirable est tout entier affaire de jugement humain. La seule transcendance morale qui soit est celle que nous pouvons construire collectivement en interrogeant notre action et le devenir qu'elle entraîne. Dès 1920, Čapek perçoit avec clairvoyance des thèmes qui, aujourd'hui, brûlent d'actualité :

¹³ *Ibid.*, p.28.

- 1) Génies, idéalistes et ambitieux peuvent se révéler extrêmement dangereux, à l'instar des fanatiques religieux. Brigitte Munier résume bien le problème dans la préface de *R.U.R.* : « on ne devrait jamais donner le pouvoir à l'homme d'une seule idée. »¹⁴ Le vieux Rossum n'avait pas prévu les conséquences de son invention, trop occupée à satisfaire son ambition démiurgique. Harry Domin, le DG de R.U.R., est un idéaliste qui ne voit pas ou plutôt qui ne veut pas voir la catastrophe arriver. Pour lui, le robot symbolise l'avènement du paradis sur Terre, d'une société où l'homme, débarrassé du labeur, pourra se consacrer entièrement à cultiver sur esprit. Cela n'est pas sans évoquer les personnalités monomaniaques de nombreux entrepreneurs du monde des technologies.
- 2) L'activité technoscientifique déployée à l'échelle industrielle à des fins de commercialisation massive ou non – qu'on appelle aujourd'hui « recherche et développement » (R&D) – engendre une complexité qui, entre autres problèmes, pointe la dilution de la responsabilité au sein d'une chaîne causale des plus obscures. En effet, qui peut être désigné responsable de l'insurrection des robots ? Le Dr Gall qui a développé leur sensibilité ? Hélène qui a insisté pour qu'il le fasse ? Harry Domin dont l'idéalisme n'a d'égal que la myopie morale ? Rossum et son combat personnel contre Dieu ? Son neveu qui rationalisa l'œuvre du savant pour l'industrialiser ? Toute l'équipe de l'entreprise ?
- 3) La situation d'oligopole existentiel qu'exercent certaines sociétés comme Google ou Meta qui, en combinant technologie et logique de marché, ambitions personnelles et capitalisme actionnarial, façonnent un monde dont la structure imprègne fortement nos vies individuelles et collectives.

En peu de mots, la figure symbolique et désormais séculaire du robot, vit le jour « pour dénoncer le hiatus grandissant entre la surestimation du progrès scientifique et le désintérêt pour l'évolution morale et sociale de l'homme »¹⁵. Pire encore : la seconde semble devenir totalement fonction du premier. Tout se passe alors comme si une religion en remplaçait une autre.

¹⁴ Introduction de Brigitte Munier dans Čapek, K. (2011). *R.U.R.*, Paris : La Différence, p.10.

¹⁵ *ibid.*, p.11.

Depuis l'écriture de *R.U.R.*, le mythe du robot a largement fait florès. Sa popularité est désormais proportionnelle à son caractère déroutant. Entité limite par excellence, le robot perturbe car il n'est assignable à aucune des catégories que nous mobilisons habituellement pour penser et décrire le monde. Se situant dans un espace non cartographié entre le moyen et la fin, le sujet et l'objet, l'instrument et la personne, le robot est une entité spéculaire (Munier, 2014) dont la puissance d'interpellation et de confrontation est proprement *dérangante* : où s'arrête la machine et où commence l'humain ? Les frontières sont floues. L'Histoire comme l'actualité regorgent aussi bien d'exemples de robotisation de l'homme par l'homme, au sens de réduction instrumentale de l'autre à une machine sans âme, que d'exemples d'humanisation des machines. La spécularité du robot se double ainsi d'une invitation à penser l'altérité sur au moins deux plans : l'altérité de l'artefact lui-même et celle de l'autre humain, l'alter ego. De façon plus générale, le robot est révélateur d'une tension très forte entre identité et altérité, entre le même et l'autre, et ce, jusque dans son actualisation par et dans la technique. Car au mythe correspondent en effet des étants bien tangibles que l'on nomme robots.

1.2 Le robot est un automate

Comme le remarque Charles Corval, doctorant à SciencesPo, dans un méticuleux mémoire de Master dédié à la réception philosophique de l'automate au XVIIIe siècle en France :

« le robot dispose d'une histoire [et] son nom actuel ne fait que le séparer artificiellement d'une trajectoire multimillénaire. Ce que le sens commun contemporain croit être une invention neuve de l'Occident technoscientifique remonte en vérité au moins aux civilisations antiques. Les anthropologues et les historiens racontent un cheminement qui commence dès les statues articulées des temples indiens » (Corval, 2018, p.8).

Avant la rupture terminologique introduite par Čapek, les machines animées et leurs créateurs étaient respectivement nommés « automates » et « automaturges ». Du latin *automatus*, lui-même emprunté au grec ancien *autómatos* signifiant « qui se meut soi-même », la notion d'automate semble aujourd'hui réservée à une classe d'artefacts réputés archaïques car

purement mécaniques. Pourtant, comme nous allons le voir, la plupart des machines que l'on nomme aujourd'hui robots ont ceci de commun qu'ils sont des automates.

Une première recherche dans le Centre National de Ressources Textuelles et Lexicales (CNRTL), nous apprend que l'automate désigne, au sens large, un « [a]ppareil renfermant divers dispositifs mécaniques ou électriques qui lui permettent d'exécuter un programme déterminé d'opérations »¹⁶. On retrouve la même idée dans l'appréhension historique de l'artefact, avec, toutefois, une insistance fondamentale sur la notion de mouvement. Ainsi d'Alembert propose-t-il la définition suivante dans son article de l'*Encyclopédie* :

« Engin qui se meut de lui-même, ou machine qui porte en elle le principe de son mouvement. Quelques auteurs mettent au rang des automates les instrumens de mécanique, mis en mouvement par des ressorts, des poids internes, &c. comme les horloges, les montres, etc. »¹⁷

Le dictionnaire de Furetière, publié en 1690, soit 61 ans avant le premier volume de l'*Encyclopédie*, ne disait pas autre chose. L'automate y était ainsi présenté :

« Terme des Méchaniques. Machine qui se remuë toute seule, qui a en soy le principe de son mouvement, comme une montre, une horloge à contrepoids, ou autres machines qui se meuvent par ressort. »¹⁸

L'automatique est donc abordé comme branche de la technique visant à concevoir des artefacts capables de se mouvoir par eux-mêmes. De façon plus générale, les automates appartiennent à la grande famille des machines perçues comme

« ce qui sert à augmenter & à régler les forces mouvantes, ou quelque instrument destiné à produire du mouvement de façon à épargner ou du tems dans l'exécution de cet effet, ou de la force dans la cause. »¹⁹

Si la machine est un moyen de mettre quelque-chose en mouvement, l'automate est quant à lui un type particulier de machine qui se met elle-même en mouvement, qui bouge par elle-

¹⁶ CNRTL, entrée « automate » (<https://www.cnrtl.fr/definition/automate>).

¹⁷ *Encyclopédie*, Vol. I, pp. 896-897, article « Automate », consultable à <http://enccre.academie-sciences.fr/encyclopedie/article/v1-3743-0/> et cité dans (Corval, 2018).

¹⁸ Furetière, A. (1690). Dictionnaire universel. Ed. par REY A. (1978). Paris. Le Robert, cité dans (Corval, 2018).

¹⁹ *Encyclopédie*, Vol. IX, article « machine », pp. 794-795, consultable à <http://enccre.academie-sciences.fr/encyclopedie/article/v9-2259-0/> et cité dans (Corval, 2018).

même. Toutefois, les formules telles que « par elle-même », « de lui-même », ou encore « principe de son mouvement », restent problématiques en ce qu'elles ne renvoient pas à une causalité claire. Parmi les quatre causes qu'il identifie dans sa *Physique*, Aristote définit la « cause motrice » (également nommée « cause efficiente » par la suite) comme « principe premier d'où part le changement²⁰ ou la mise en repos ». De toute évidence, la cause du mouvement de l'artefact n'est pas à chercher dans une quelconque volonté de l'automate. Le canard ou le flûteur de Vaucanson ne décident pas eux-mêmes de s'activer. De ce point de vue, ils n'agissent pas, mais *sont* agis. En revanche, ils possèdent la structuration mécanique nécessaire à la production et à l'entretien du mouvement dont ils font preuve. L'automate n'est donc pas cause motrice au sens de déclencheur, mais au sens de producteur et d'organisateur. Son autonomie n'est pas celle du « pourquoi bouger », mais de son « comment ». Pour Charles Corval :

« "Avoir le principe de son propre mouvement en soi" ne renvoie pas à autre chose qu'à l'idée d'une séquence de mouvement dont la machine organiserait elle-même le processus temporel ; quand bien même un humain aurait à l'activer via un système de démarrage » (Corval, 2018, p.16).

Partant de cette dimension motrice générale, on découvre que l'automate, à l'instar de la machine, est avant tout un terme taxonomique désignant un ensemble composite d'artefacts unis par une communauté de principe. En effet, le canard de Vaucanson, simulant par voie mécanique le processus digestif de l'animal, est un automate au même titre qu'une montre ou qu'une pompe hydraulique automatisée, mais pas le même *genre* d'automate. Leur différence est principalement fonctionnelle. Contrairement au canard, l'horloge n'a pas qu'une fonction ornementale et démonstrative du génie mécanique : elle est également utile. Affranchi de ces considérations taxonomiques, le langage courant semble toutefois réserver le terme d'automate aux artefacts strictement biomimétiques, c'est-à-dire conçus sans autre objectif qu'imiter le vivant²¹. Le fameux *androïde* (du grec *andrós*, « homme », et *eîdos*, « forme », « aspect extérieur »), qui est aujourd'hui souvent utilisé comme synonyme de

²⁰ Aristote ne pense pas le mouvement au sens restreint du déplacement d'un point A à un point B, mais au sens large de changement.

²¹ On pourrait même dire zoomimétique en ce qu'il s'agit surtout d'imiter le vivant animal (auquel l'humain appartient).

robot, désigne initialement un automate anthropomimétique, reproduisant l'apparence, le mouvement et, plus largement, les attitudes humaines, à l'instar du flûteur de Vaucanson. En tant qu'il appartient à la famille des automates, le robot hérite de ces ambiguïtés et les entretient. On retrouve ainsi les mêmes variations d'usages et les mêmes chevauchements de référentiels chez le robot que chez l'automate classique. Le robot est souvent immédiatement associé à la figure de l'androïde mécatronique ou cyborganique telle qu'on la trouve, par exemple, dans les œuvres d'Isaac Asimov ou de Philip K. Dick. Cette appréhension populaire, issue de la science-fiction, est souvent celle que l'on retrouve en premier dans les dictionnaires. Du point de vue technique et taxonomique, toutefois, la robotique est l'autre nom de l'automatique moderne. Ses moyens ont beau différer ou s'étoffer, son principe originel reste le même : créer des machines intégrant le principe de leur propre mouvement. Malgré cela, certains partis pris sémantiques réservent le nom d'automate aux entités purement mécaniques de l'ère préindustrielle. Ainsi, Nathalie Lazaric, directrice de recherche au CNRS, écrit-elle :

« Les premières solutions techniques des robots se trouvent dans les mécanismes des automates qui comprenaient déjà quelques principes d'autorégulation. Cependant la différence réside dans la programmation : les robots sont reprogrammables et multifonctionnels selon la tâche impartie, l'automate par contre est simplement animé d'un mécanisme intérieur assurant son mouvement et limitant sa programmation. Leur finalité d'autre part est radicalement opposée, le robot est une machine industrielle tandis que la construction de l'automate repose sur une élaboration ludique, voire philosophique : la recherche du mouvement perpétuel » (Lazaric, 1992, p.55).

Ici, l'usage des deux termes instaure une ligne de partage très claire entre un avant et un après, rendant l'emploi du mot automate anachronique en matière de machinerie moderne. Si Lazaric se montre simplificatrice en résumant l'ensemble de l'automatique et de la robotique à des projets respectivement ludo-philosophiques et instrumentaux²², sa référence à la nature industrielle du robot est, quant à elle, tout à fait décisive. L'évolution technique et langagière de l'automate classique vers le robot procède bien de l'industrialisation de l'économie et de l'idéologie productiviste qui en découle. La conception technique s'en

²² Comme l'atteste l'exemple de l'horloge, les automates pouvaient avoir une fonction utilitaire précise tandis que la robotique regorge d'exemples d'artefacts inutiles ou seulement *potentiellement* utiles dont la valeur est souvent plus démonstrative qu'utilitaire.

retrouve orientée par les nouveaux impératifs d'efficacité et de rendement. Le robot devient alors le symbole de ce nouveau paradigme techno-économique et se charge d'une dimension sociofonctionnelle dont l'automate classique était exempt : celle du « travailleur artificiel [...] autonome dans une certaine mesure » dont parlent Dumouchel et Damiano (2016).

1.3 Le robot est un faisceau de normes

A ce premier détour historique centré sur la définition de l'automate, s'ajoute celui dédié à son statut d'entité liminaire. Comme nous l'avons dit du robot, « l'automate est un rapport dans la chose entre l'identité et la différence » (Corval, 2018, p.18). Artefact non-humain ressemblant à l'humain ou à l'animal, celui-ci a tout du phénomène quantique au sens figuré : il est et n'est pas ce qu'il imite en même temps. Pour Charles Corval, « c'est dans cet écart de l'identité et de la différence que s'articule tout l'enjeu d'un usage philosophique de l'automate » (*ibid.*). Le jeune chercheur défend alors l'idée que l'automate du XVIIIe siècle fait figure d'être normatif au double sens de modèle (normalité) et de règle (morale), voire de principe (ontologie).

La figure de l'automate est mise à contribution épistémologique dans le débat général sur la nature humaine, notamment à travers la question des liens entre l'esprit et le corps. Chez Descartes, l'automatisme relève du domaine matériel de la *res extensa*, c'est-à-dire de la « chose étendue » que le philosophe nomme également « substance corporelle ». En tant qu'il appartient strictement au domaine de l'étendue, le corps humain est pensé comme un automate biologique. Entité réactive sur le mode du réflexe, ce dernier est régi par une machinerie biomécanique que seule la présence d'une « substance pensante » (*res cogitans*), c'est-à-dire d'une âme peut contrarier. La célèbre thèse cartésienne de l'animal-machine place les animaux du côté de la pure étendue : les animaux sont des automates dénués d'âme en ce que, contrairement à l'humain, il n'existe aucune preuve de leur capacité à penser. L'animal ne connaît pas de séparation en son sein puisqu'il n'est que matière. L'humain, en revanche, est cet être qui transcende la seule matérialité, le seul automatisme du corps, par la possession d'une âme. Pas d'ambiguïté, donc, chez Descartes : l'automate aussi complexe soit-il, n'est qu'un agencement d'étendue qui fonctionne indépendamment de la pensée, c'est-à-dire sans intelligence. « Descartes est d'opinion que les brutes ne sont que des

automates »²³, précise la définition de Furetière citée plus haut, ce qui suggère clairement que le comportement machinique témoigne d'une domination du corps sur l'esprit, c'est-à-dire d'un enchaînement à la pulsion et aux déterminismes dénués de toute intelligence.

On retrouve cette approche automatique du vivant chez l'un des principaux représentants du mécanisme au XVIIIe, à savoir La Mettrie. Médecin et philosophe matérialiste, l'auteur de *L'Homme-machine*, mobilise la figure de l'automate comme modèle explicatif des phénomènes naturels au sens large. En observant la mécanique du monde physique, en la décomposant et en la recomposant dans des structures automatiques, les automatiques sont les parfaits représentants de l'empirisme qui, pour La Mettrie, résume l'esprit scientifique. Si l'automate ne saurait épuiser le réel, trop complexe pour être entièrement appréhendé, « il fournit un cadre à la science pour éviter l'intervention de causes non-matérielles ou d'éléments chimiques non explicables » (Corval, 2018, p.56). Toutefois, contrairement à Descartes, La Mettrie n'accorde pas à l'humain une place à part dans ce vaste enchaînement matériel de causes et d'effets qui compose le réel. Cela s'explique par le matérialisme du philosophe. Dans ce monde tout entier fait de matière, le terme d'« âme » n'est qu'une façon de désigner l'organe de pensée qu'est le cerveau. Les différences entre les êtres ne sont pas de nature ontologique, mais physique (Marzano, 2016) : le facteur de diversité animale, végétale, minérale etc., n'est autre que l'organisation spécifique à chaque entité. Ainsi, bien que l'automatisme humain soit d'une insondable complexité, il reste un automatisme parmi d'autres.

Voltaire ne soutiendra pas autre chose dans le cadre de sa pensée déiste. L'auteur du *Dictionnaire philosophique* ancre ses réflexions dans une philosophie de la connaissance où l'automate est pensé comme figure d'un « nouvel *ethos* philosophique et scientifique » (Corval, 2018, p.57). Outre son rapport au mouvement, l'automate se caractérise par l'ignorance des causes et principes premiers de son fonctionnement. Le canard mécanique comme le canard biologique partagent cette condition d'ignorance. L'automate est donc cette entité qui « subit les règles que le "Mécanicien" lui a donné sans pouvoir nécessairement les comprendre, ni même trouver une expérience satisfaisante qui lui permettrait de les connaître véritablement » (*ibid.*, p.56). L'humain, malgré son statut d'être conscient, n'échappe pas à la règle. C'est d'ailleurs précisément sur ce point que l'humain ressemble le

²³ Furetière, A. *op.cit.*

plus à un automate : si, comme le soutiennent La Mettrie ou Locke (dont Voltaire est un fervent lecteur), la mécanique du réel extérieur peut nous être scientifiquement accessible par la méthode empirique d'observation, cela est beaucoup moins vrai de la mécanique interne à la conscience à laquelle nous n'avons que très partiellement accès. Les progrès de la médecine et de la biologie ont beau permettre de connaître et de comprendre notre mécanique corporelle, il n'en va pas de même pour la machinerie psychologique qui produit nos pensées et notre volonté. L'ensemble des causes et des effets du domaine spirituel est à la fois trop enchevêtré et trop profondément enfoui pour que nous y ayons accès. Avant même d'être un artefact matériel, l'automate voltairien se pose comme modèle épistémologique de compréhension scientifique du monde, des animaux et de l'humain. Les uns comme les autres *fonctionnent* indépendamment de tout rapport à la volonté et à la connaissance de soi. Pour Voltaire, la mécanique de l'automate est symbolique d'« une réunion [du tout] sous un même règne : celui des lois universelles de la nature qui s'imposent de la même façon que le font les règles des mécaniciens aux machines qu'ils peuvent produire » (*ibid.*, p.68). L'humanité est donc ramenée à l'animalité qui est elle-même ramenée à ses principes et constitution biomécaniques. En empiriste convaincu, Voltaire ne se réfère pas à des catégories ontologiques ni mêmes taxonomiques : l'« humain », l'« animal » etc. sont simplement des mots permettant de désigner des récurrences dans le réel, pas des essences. L'expérience montre que bien des hommes apparaissent infiniment plus stupides ou plus viles que nombre d'animaux. Pour Voltaire, les grandes figures de la philosophie, en particulier les métaphysiciens comme Descartes et leurs réflexions abstraites, se rendent coupables d'une double erreur : celle de construire de la pensée inconsistante car privée de l'ancrage perceptif permettant d'expérimenter et d'observer, mais aussi celle de systématiser un discours sur l'humain et sa prétendue supériorité, alors qu'ils ne font que témoigner de leur propre et singulière capacité à avoir de l'esprit. On comprend alors pourquoi la figure épistémologique de l'automate est aussi une figure morale permettant de lutter rationnellement contre l'*hybris* intellectuelle dont certains humains font preuve. Voltaire promeut un *ethos automatique* du discours scientifique sur l'humain et le monde où l'observation empirique est synonyme d'humilité. Dans le schéma voltairien, l'humain fait donc figure d'« automate à idées et à sensations ». « Le savant ne doit pas vouloir à tout prix la vérité absolue de la métaphysique, mais aimer sa propre machinerie interne et l'enquête

sur les frontières de ses facultés à connaître » (*ibid.*, p.83). Cet amour de la machinerie implique une reconnaissance de nos déterminismes programmatiques :

« l'usage de l'automate chez Voltaire se rapproche vivement de la question du programme ou du code. [...] Au siècle des Lumières, le programme désigne avant tout l'annonce d'un déroulement, principalement un prospectus capable de présenter succinctement une pièce ou un spectacle. Le code lui aussi se veut comme l'énonciation d'un ensemble de règles qui devront ensuite être suivies. Ce n'est que récemment que les deux mots ont pris un sens technique particulièrement localisé en informatique et en électronique. L'automate sert de modèle moins pour son apparence extérieure que pour l'intériorité qui le pousse à effectuer les mouvements voulus par l'ingénieur » (*ibid.*, p.76).

Précisons toutefois que cet ethos voltairien de l'automate et l'acceptation du déterminisme qu'il suppose n'invitent aucunement à se résigner quant à la possibilité d'une quête sérieuse de connaissance. L'automatisme ne freine pas la connaissance scientifique et la recherche philosophique, il prescrit simplement le cadre et l'humilité qu'elles demandent.

« La volonté de dépassement doit être contenue par une autre passion : celle de rester dans les justes limites de son être. Les justes limites elles-mêmes incitent à gouverner l'imagination par l'observation d'une part, et par l'examen sceptique d'autre part. Les mots que l'on utilise doivent ainsi toujours justement pouvoir retrouver l'idée qu'ils figurent, et l'idée en question doit toujours pouvoir être analysée dans son rapport à ce qui est observable » (*ibid.*, p.82).

Cette lutte contre l'hybris par la figure de l'automate classique peut sembler contre-intuitive tant le robot symbolise, depuis sa naissance dans *R.U.R.*, l'absence de mesure d'*homo industrialis* (Lefèvre & Luzi, 2017). En outre, ce contraste entre mesure et démesure se double d'une tension entre amour et désamour de soi. Si l'automate voltairien invite à adopter une posture d'humilité scientifique, cette attitude suppose une joyeuse acceptation de soi comme entité certes déterminée, mais non moins capable de science. L'automate est donc le symbole d'une épistémologie aussi modeste qu'enthousiaste où connaissance de soi et connaissance du monde sont les maîtres-mots de l'épanouissement humain. Le robot, lui, se présente sous un jour à la fois plus noir et plus ambigu. Čapek ne donne pas naissance à une figure épistémologique et ontologique positive, mais plutôt à une entité dérangeante au sens propre comme au sens figuré. Au sens propre, d'abord, car, non content de remettre en cause les

catégories modernes de pensées qui séparent le sujet de l'objet, la nature de la culture, le social du technique, l'animé de l'inanimé, le robot questionne directement notre humaine condition et nos aspirations – et en cela, le robot semble très proche de l'automate des Lumières. Au sens figuré, ensuite, car la dimension morale du questionnement qu'il suscite semble répondre à une axiologie bien différente de celle de l'époque classique. Loin de l'humble acceptation de ce que nous sommes, le robot est souvent présenté comme l'antithèse d'une condition humaine foncièrement défaillante. Naissant comme pied de nez adressé au Divin, le robot se charge rapidement de la dimension instrumentale propre à l'hyper-rationalisme économique de l'ère capitaliste. Il incarne alors à la perfection l'impératif d'efficacité industrielle : infatigable, inlassable et totalement apathique, sa force de travail n'a pas à être renouvelée, pas plus que sa motivation. Tout son être converge vers un seul et même objectif : produire efficacement et indéfiniment. Pas question, ici, d'une quelconque modernisation du « connais-toi toi-même » delphique : un robot n'a pas à connaître et encore moins à se connaître, il doit simplement *fonctionner* techniquement et instrumentalement. Malgré son apparente indigence existentielle, cette condition génère en l'homme moderne une forme d'attraction-répulsion le conduisant à désirer deux états contradictoires : être et ne pas être un robot. Dans l'entretien qu'il nous a accordé, Dominique Sciamma, directeur de l'école de design CY Université Paris Cergy, nous confiait la chose suivante :

« Plus jeune, je voulais écrire un livre dont le titre, volontairement provocateur, aurait été *Devenir un robot*. Quand on me demandait pourquoi, je répondais que le robot avait pour moi une dimension super-héroïque : il est hyper-efficace et incarne l'idéal de maîtrise et de contrôle. En outre, à l'époque, le déterminisme associé à la figure du robot ne me dérangeait pas, au contraire : être robot, c'était accomplir sa destinée comme on fait tourner un programme. Mais au fond, la vraie signification de tout cela, la vraie figure marquante, c'était celle de R. Daneel Olivaw, LE héros androïde d'Asimov. C'est un personnage messianique, christique, éternel. R. Daneel Olivaw c'est l'amour éternel, l'attention éternelle, l'empathie éternelle et il n'y a que lui qui est comme ça ! C'est la seule personne authentiquement humaine de la saga. Les autres personnages essaient, recherchent ou rejettent cette humanité. Lui, il l'incarne tout le temps et pour toujours. Mais je sais que dans cette fascination, il y a une grande part d'ivresse philosophique qui parle. »

A l'instar de ce que dit Sciamma dans la première partie de ce témoignage, le devenir machine se présente en effet comme la promesse d'une hyper-adaptation à l'impératif rationnel d'efficacité du système technocapitaliste, celui-là même qui exerce sur les individus interchangeables une pression constante. Nous nous sentons dépassés par ces machines qui s'acquittent de leurs tâches avec une aisance et une efficacité infiniment supérieures aux nôtres. Ainsi, le robot possède lui aussi une dimension normative, mais de nature culpabilisante. C'est ce sentiment de déchéance fascinée que Gunther Anders nomme « honte prométhéenne » dans *L'obsolescence de l'homme*. Se faire robot, c'est se délivrer des « failles » humaines que sont le corps biologique, les émotions et ce qu'on appelle communément le libre arbitre, source de tant d'erreurs. Le robot ne doute pas, il agit sans se poser de question. Il n'entretient aucun rapport à la liberté. En termes heideggériens, le robot se présente comme l'antithèse d'un *dasein*. Tout se passe comme si le désir de robot révélait notre volonté de substituer l'être-dans-le-monde à notre être-au-monde pour atteindre l'inerte quiétude du simple étant, l'inexistence donc, caractérisée par l'absence de Souci (*die Sorge*). Ce désir d'efficacité opérationnelle se double d'une dimension morale comme l'atteste la référence de Sciamma à l'androïde R. Daneel Olivaw. Robotiser l'amour, l'attention et l'empathie, c'est la rendre à la fois automatique et systématique, c'est-à-dire permanente et imparable. Chez Spoon, un collègue développeur nous confiait une idée similaire quoiqu'en apparence moins affective :

« Le robot ou les machines en général sont honnêtes, contrairement à l'humain. J'ai plus confiance en une machine que j'ai configurée avec certaines informations qu'en un humain à qui j'aurais fourni les mêmes informations. L'Homme va chercher à intervenir sur les informations qui lui sont données, à les modifier pour mieux les traiter, même si ça les dénature et change leur sens. La machine, elle, fait ce qu'on lui demande et ne fait pas ce qu'on ne lui demande pas. On sait donc pourquoi elle fait ou ne fait pas les choses. Avec l'humain, on court toujours le risque des machinations sociales. Un service qui t'est rendu peut se retourner contre toi et ce simple principe m'énerve. Ceci dit, on cherche désormais à doter les machines de cette capacité de divergence qui caractérise l'humain : on veut instiller à la machine l'interprétation des faits et non plus seulement les faits eux-mêmes. »

Dans ce témoignage, comme dans le précédent, on voit que le robot incarne une certaine idée de la perfection conçue comme infaillibilité à la fois opérationnelle, morale et destinale. Parallèlement, toutefois, l'image du robot nous révolte. « Je ne suis pas un robot » dit-on lorsqu'on nous en demande trop et/ou qu'on nous demande d'exécuter une tâche sans poser de question, sans exercer notre esprit critique. Ce n'est plus l'attrait de l'efficacité qui parle, mais la prise de conscience de sa démesure et de sa vanité. La symbolique du robot nous apparaît alors sous un jour bien moins désirable : celui d'un ethos prolétarisant au double sens de « mécanisation des esprits par l'extériorisation des savoirs dans des systèmes » (Stiegler, 2011) et de mécanisation des corps imposée par le rythme de la machine. L'un des exemples les plus commentés en la matière est peut-être celui des entrepôts de l'entreprise américaine *Amazon* et de leur régime de productivité connu sous le nom de « *the rate* » (le taux) (Carr-Brown, 2018 ; Lecher, 2019 ; Dzieza, 2019, 2020). La productivité minimum est fixée par le système informatique de la firme. Si les manutentionnaires n'atteignent pas les taux exigés, ils sont rapidement licenciés « par la machine », c'est-à-dire sans autre procédure qu'un message provenant du système en question.

« Dans une lettre signée l'année dernière, un avocat représentant Amazon a déclaré que l'entreprise avait licencié des "centaines" d'employés dans un seul établissement entre août 2017 et septembre 2018 pour n'avoir pas respecté les quotas de productivité. [...] Les documents montrent également un processus de suivi et de licenciement profondément automatisé. "Le système d'Amazon suit les taux de productivité de chaque associé", selon la lettre, "et génère automatiquement tout avertissement ou licenciement concernant la qualité ou la productivité, sans contribution des superviseurs." [...] "L'une des choses que nous entendons constamment de la part des travailleurs est qu'ils sont en effet traités comme des robots parce qu'ils sont surveillés et supervisés par ces systèmes automatisés"[...]. "Ils sont surveillés et supervisés par des robots."

Le système va jusqu'à suivre le "temps hors tâche", que l'entreprise abrège en TOT²⁴. Si les travailleurs cessent de scanner les colis pendant trop longtemps, le système génère automatiquement des avertissements et, à terme, l'employé peut être licencié. Certains travailleurs de l'établissement ont déclaré qu'ils évitaient les pauses toilettes pour que leur temps soit conforme aux attentes » (Dzieza, 2020).

²⁴ Pour « *time off task* ».

L'image du robot est ici associée à un stakhanovisme algorithmique où l'établissement de la norme (le fameux « taux ») est déléguée aux machines et à leurs lignes de code. En substance, le régime normatif des entrepôts d'Amazon transforme l'espace de travail des manutentionnaires en milieu hostile sans négociation ni adaptation possible. « En attendant les robots », les vrais, pour reprendre le titre provocateur de l'ouvrage d'Antonio Casilli, il s'agit de robotiser l'humain au maximum. Nous qualifions ces milieux d'hostiles en ce que les normes de productivité qui les régissent entrent en contradiction avec la normativité vitale (Canguilhem, 1966) des manutentionnaires. En d'autres termes, l'humain impose à ses pairs une normativité technique aux effets délétères sur le bon fonctionnement de leur organisme. Chez Amazon, cela se traduit notamment par d'importants troubles musculo-squelettiques, comme le rapporte un journaliste du site américain *The Verge* à propos d'un employé témoignant sous pseudonyme :

« Après plusieurs mois, [Jake] a senti une brûlure dans son dos. Un superviseur lui disait parfois de plier davantage les genoux lorsqu'il soulevait des charges. Quand Jake s'exécutait, sa cadence baissait et un autre superviseur lui disait d'accélérer. "Vous vous moquez de moi. Accélérer ?" se souvient-il avoir dit. "Si je vais plus vite, je vais avoir une crise cardiaque et tomber par terre". Finalement, son dos a complètement lâché. On lui a diagnostiqué deux disques endommagés et il a dû se mettre en invalidité. Le taux, dit-il, était "100 %" responsable de sa blessure » (*ibid.*).

Et le journaliste d'ajouter :

« Tous les employés d'Amazon à qui j'ai parlé m'ont dit que c'est le rythme de travail imposé automatiquement, plus que la difficulté physique du travail lui-même, qui rend la tâche si éreintante. Tout relâchement est perpétuellement évacué par le système, et avec lui toute possibilité de repos ou de récupération. Un travailleur de la côte ouest m'a parlé d'un nouveau dispositif qui éclaire l'article qu'il est censé choisir, ce qui permet à Amazon d'accélérer encore la cadence et de se débarrasser de ce que le travailleur décrit comme des "micro-repos" volés dans l'instant qu'il faut pour chercher l'article suivant sur l'étagère. (...) En novembre, *Reveal* a analysé des documents provenant de 23 entrepôts Amazon et a constaté que près de 10 % des travailleurs à temps plein ont subi des blessures graves en 2018, soit plus du double de la moyenne nationale pour un travail similaire. De multiples travailleurs d'Amazon m'ont dit que les blessures dues au stress répétitif sont très courantes, mais rarement signalées. (Un porte-parole d'Amazon a

déclaré que l'entreprise prend la sécurité des travailleurs au sérieux, dispose de personnel médical sur place et encourage les travailleurs à signaler toutes les blessures). Les maux de dos, les douleurs aux genoux et d'autres symptômes de tension constante sont suffisamment fréquents pour qu'Amazon installe des distributeurs automatiques d'analgésiques dans ses entrepôts » (*ibid.*).

L'application de telles normes techno-économiques révèle une vision de l'humain comme *homme-machine*, non plus au sens biomécaniste de La Mettrie, mais au sens instrumental d'automate productif. La présence de distributeurs d'antalgiques au sein des entrepôts est très révélatrice de cette vision : il s'agit de *réparer* l'humain qui dysfonctionne en maintenant un intégrité physique minimale, et non de le *soigner*. Ce second geste, contrairement au premier, impliquerait de reconfigurer l'intrication conflictuelle des normes au profit de la normativité vitale. La conception classique en robotique industrielle stipule que les robots ont vocation à effectuer des tâches relevant du triptyque dit des trois « D » : *Dangerous* (dangereuses), *Dirty* (sales), *Dull* (ennuyeuses/répétitives) (Takayama *et al*, 2008). Or, dans l'exemple que nous venons de développer, c'est l'humain qui est investi d'une telle vocation et qui doit s'adapter aux directives de la machine.

Cette figure du robot reste toutefois très ancrée dans sa genèse et ses développements industriels. Par la suite, nous verrons que l'émergence progressive des robots sociaux enrichit petit à petit la symbolique générale de l'automate.

1.4 Définitions, taxonomie et exemples

1.4.1 Définitions technico-fonctionnelles

Si la figure du robot émerge au sein du champ littéraire comme entité humanoïde autonome sur le plan énergétique, intellectuel et moteur, les étants actuels qui portent le même nom sont très éloignés de cette figure originelle. Alors que certaines approches lient la notion de robot à un type de machine automatiques bio-inspirées au sens où elles rappellent, même de façon schématique, tout ou partie d'un corps organique animé, la plupart des définitions évacuent désormais cet enracinement biomimétique pour se concentrer sur des éléments techniques et/ou fonctionnels. Cela ne les empêche pas de varier selon les dictionnaires, les

auteurs et les organisations. Selon leur degré de précision, notamment technique, ces définitions seront plus ou moins inclusives ou exclusives quant aux artefacts concernés.

Les philosophes Paul Dumouchel et Luisa Damiano (2016), par exemple, s'appuient sur une approche composite qu'ils présentent comme « définition originale du terme » et qualifient le robot de « travailleur artificiel, conçu comme un appareil qui a sa propre source d'énergie, qui travaille pour nous, et qui est autonome dans une certaine mesure » (*ibid*, p.7). C'est donc leur autonomie énergétique et opérationnelle ainsi que leur fonction substitutive qui semble distinguer les robots d'autres artefacts. Les auteurs remarquent toutefois qu'une telle définition pose un problème majeur en ce

« qu'elle associe deux critères de natures fort différentes. Premièrement, un critère d'ingénierie, un mécanisme automatique autonome, c'est-à-dire capable de s'adapter à certains changements de son environnement et de transformer son comportement en conséquence. Deuxièmement, un critère sociofonctionnel, "travailler à notre place" ou remplir la tâche d'un ouvrier humain. Or si le premier critère peut être défini rigoureusement, le second est extrêmement vague et renvoie à une réalité culturellement relative et historiquement changeante : le travail humain. Dès lors, il n'est pas clair que ces deux critères pris ensemble qualifient une classe particulière d'objets techniques, plutôt qu'une collection disparate qui regroupe des réalités techniques fort différentes » (*ibid.*, p.10).

Cette remarque montre bien que le concept de robot est fondamentalement difficile à définir tant il articule des notions de nature hétérogène qui empêchent de le réduire à une réalité purement technique.

L'*International Standards Organization* (ISO) évacue ce problème par une approche strictement technofonctionnelle et nomme robot tout « mécanisme programmable actionné sur au moins deux axes avec un degré d'autonomie, se déplaçant dans son environnement, pour exécuter des tâches prévues. » Raja Chatila (2014), Professeur à Sorbonne Université et Directeur de l'Institut des Systèmes Intelligents et de Robotique (ISIR), se veut, quant à lui, plus précis techniquement :

« Un robot est une machine matérielle qui, sous différentes formes, est essentiellement munie d'*actionneurs* lui permettant de se mouvoir, ou de déplacer une partie de son corps dans l'espace, de *capteurs* qui acquièrent des données sur son environnement (par

exemple des caméras) ou sur elle-même (par exemple des codeurs mesurant les rotations des moteurs), ainsi que de *processeurs*, des calculateurs qui sont des moyens de traitement de l'information. Cette machine peut être aussi dotée de moyens matériels de *communication* (microphone, haut-parleur, transmission radio, ...). Et, bien sûr, d'une source d'énergie. »

Pierre-Yves Oudeyer (2016), directeur de recherche à l'INRIA, propose une définition similaire tout en introduisant et en insistant sur l'idée de rétroaction entre les captations du robot et ses actions.

« D'un point de vue technique, un robot est une machine dotée de capteurs qui lui permettent de percevoir son environnement (par exemple capteur de contacts, de distances, de couleurs, de force...), de moteurs l'autorisant à bouger et à agir sur cet environnement, et d'un système électronique ou informatique qui contrôle, au moins en partie, ce qu'effectue le robot en fonction de ce qu'il perçoit. Une caractéristique fondamentale des robots est en particulier cette rétroaction entre la perception et l'action : ce que fait le robot ne dépend pas entièrement de ses instructions, mais aussi des propriétés physiques de son environnement perçues avec ses capteurs. Ainsi, selon cette définition, les automates (tels que ceux de Jacques de Vaucanson ou de Pierre et Henri-Louis Jaquet-Droz au XVIIIe siècle) ne sont pas des robots car leurs mouvements ne dépendent pas de ce qui se passe autour du robot (leurs propriétés et leurs enchainements sont entièrement prédéterminés par le programme et il n'y a pas de capteurs). »

Dans son rapport intitulé *Éthique de la recherche en robotique*, la CERNA²⁵ (2014) adopte elle aussi une approche technique, mais choisit de la détailler à l'extrême.

« Le robot [y] est défini comme une machine mettant en œuvre et intégrant :

- Des capacités d'acquisition de données avec des capteurs à même de détecter et d'enregistrer des signaux physiques.
- Des capacités d'interprétation des données acquises permettant de produire des connaissances.
- Des capacités de décision qui, partant des données ou des connaissances, déterminent et planifient des actions. Ces actions sont destinées à réaliser des

²⁵ Commission de réflexion sur l'Éthique de la Recherche en sciences et technologies du Numérique d'Allistène

objectifs fournis le plus souvent par un être humain, mais qui peuvent aussi être déterminés par le robot lui-même, éventuellement en réaction à des événements.

- Des capacités d'exécution d'actions dans le monde physique à travers des actionneurs, ou à travers des interfaces.

Le robot peut également présenter :

- Des capacités de communication et d'interaction avec des opérateurs ou des utilisateurs humains, avec d'autres robots ou des ressources via un réseau comme l'Internet.
- Une capacité transversale aux précédentes, l'apprentissage, qui permet au robot de modifier son fonctionnement à partir de son expérience passée.

Le robot est considéré dans cette définition comme étant une entité matérielle et non uniquement logicielle. Cette définition exclut ainsi les agents logiciels ou les "bots" qui traitent des données à grande vitesse.

Les capacités du robot à se mouvoir, agir, interagir et décider de lui-même, lui confèrent une autonomie. »

Bien qu'elles offrent des degrés de précision différents, ces approches autorisent des convergences qui semblent caractériser le robot comme tel. On peut toutefois considérer que l'exhaustivité de la caractérisation proposée par la CERNA se retourne contre elle-même en faisant du robot une figure finalement très abstraite. Que signifie « produire des connaissances » pour un robot ? Pourquoi choisir d'exclure les entités dites « immatérielles » comme les bots ? Que penser, par exemple, des enceintes connectées qui agissent sur la base de captations (qui sont donc « à même de détecter et d'enregistrer des signaux physiques »), mais sans actionneurs mécatroniques et qui, de surcroît, sont dénuées de capacités motrices ? La précision quasi chirurgicale de la CERNA devient paradoxalement vectrice de confusion.

Dans son très conséquent *Traité de droit et d'éthique de la robotique civile*, la juriste Nathalie Nevejans propose une approche synthétique à la fois plus générale et plus claire, centrée sur six conditions fondamentales : un robot est 1) une machine matérielle, 2) liée à une source d'énergie, 3) dotée de capacité(s) d'action sur le réel, 4) de faculté(s) de perception de son environnement, 5) d'aptitude décisionnelle et 6) d'apprentissage (Nevejans, 2017). Elle ajoute deux critères complémentaires car non discriminants à ces conditions, à savoir a) une forme d'autonomie opérationnelle et/ou décisionnelle et b) des capacités de communication et d'interaction (humain-machine et/ou machine-machine).

Aussi précises et/ou pertinentes soient-elle sur le plan technico-fonctionnelle, ces approches s'inscrivent toutefois dans une démarche critériologique qui échoue à appréhender le robot à partir de l'unité conceptuelle de son principe.

1.1.2 Approches conceptuelles du robot : la question du « rapport au monde »

A cet égard, la définition élaborée par Joanna Bryson (2010) propose une approche conceptuelle à la fois simple, efficace et inclusive en qualifiant de robot « toute entité artificielle située dans le monde réel qui transforme des perceptions en action ». La chercheuse se concentre ici sur ce qu'elle identifie comme *principe essentiel* de la robotique et non sur ses déclinaisons techniques. Une telle approche touche bien, nous semble-t-il, au fondement irréductible du domaine. Elle n'en soulève pas moins certaines questions conceptuelles d'envergure, comme celles de la signification du « monde réel » auquel elle fait référence ou encore celle de la notion de perception appliquée au robot (pourquoi, par exemple, ne pas parler de « captation » qui, par analogie, est généralement considéré comme son équivalent technique ?). En outre, cette définition suggère que les robots n'agissent que sur la base de données perceptives, alors que nombre d'entre eux peuvent déclencher divers processus indépendamment de toute captation et ce en vertu de leur programmation²⁶. Cette approche a, par ailleurs, le mérite de se positionner en dehors du débat, conceptuellement périlleux, sur l'autonomie (*versus* le simple automatisme) du robot.

Dans la même veine, Dominique Sciamma (2012a) propose une approche de l'artefact centrée sur la notion de « rapport au monde » ou, plus précisément, de rapport *vivant* au monde en tant que la robotique cherche à reproduire au moins phénoménotechniquement ce que l'auteur qualifie de « caractéristiques du vivant » (*ibid.* p.104) à savoir la perception, la décision et l'action²⁷. Citons-le longuement sur ce point :

« Dans l'imaginaire collectif, le robot évoque inmanquablement un être humanoïde ou au minimum un animal dont les formes, les comportements et les objectifs tendent à se rapprocher toujours plus du vivant. Quand ce n'est pas le corps tout entier qu'il s'agit

²⁶ Par exemple, le robot BigSpoon est programmé pour générer aléatoirement ou toutes les x minutes d'inactivité sociale (c'est-à-dire en dehors d'une interaction socio-fonctionnelle avec un humain) des petits bruits censés manifester son ennui.

²⁷ Ici, nous entendons par « phénoménotechniquement » la *production* du phénomène de ces caractéristiques, c'est-à-dire de leur apparence par la technique.

d'évoquer, comme dans les cas des robots Nao ou Romeo de la société Aldebaran Robotics, c'est d'une partie anatomique, comme le bras, la jambe, la main, le torse ou la tête, dont on tente de s'inspirer. Il est donc souvent difficile pour les concepteurs, et encore plus, pour le grand public, de se départir de cette tentation de voir le robot comme une réplique mimétique, fût-elle partielle, du vivant. Et pourtant, à n'en pas douter, ce qui caractérise le robot relève davantage de son rapport au monde qu'à ses formes. Comment ce rapport au monde se définit-il ? Un robot peut être défini comme un système capable de percevoir des informations issues du monde ambiant, d'avoir une certaine représentation de ce monde, d'être capable de prendre des décisions dans le cadre de cette représentation, à partir des informations perçues, et d'agir dans ce monde. Ces informations comme ces actions peuvent être de nature très différente (mouvements, forces, sons, odeurs, images, compositions chimiques, lumière, bits, radio, etc.). »

Et l'auteur de conclure :

« On l'admettra : rien dans cette définition n'induit une quelconque forme, et rien ne s'oppose dès lors à ce que l'on applique cette définition à n'importe quel objet de notre quotidien. Un meuble, une automobile, un tapis, un parapluie, un mur, une fenêtre, un jouet pourraient, chacun, être dotés de ces capacités de percevoir, représenter, décider et agir. »

Sciamma l'affirme très clairement : non seulement le robot est engagé dans un rapport biomimétique au monde, mais c'est aussi ce rapport qui le définit en tant que tel. Autrement dit, s'intéresser au robot *lui-même* mène inévitablement à cet étrange constat que, du point de vue de son phénomène, le monde ne lui est pas indifférent et qu'il évolue dans une certaine forme d'attachement aux choses qui lui sont extérieures. Prévenons d'emblée un risque de malentendu : notre objectif n'est pas de soutenir une quelconque thèse d'ontologie phénoménologique en affirmant que le robot a un monde²⁸. Afin de neutraliser le propos sur ce point, disons que le robot témoigne d'un rapport sensorimoteur ou, plus largement, sensoriactif²⁹ au monde dépourvu de tout lien à quelque chose comme une conscience corporée. Autrement dit, « ça perçoit » (ou « capte »), « ça bouge », « ça agit ».

²⁸ En effet, il n'appartient pas à cette recherche de s'engager dans cette voie. Il s'agit toutefois d'un réel sujet, aussi complexe qu'intéressant. Une piste de recherche pourrait être de substituer le robot à l'animal dans la thèse heideggerienne de la pauvreté en monde et de la stupeur (*Benommenheit*) qui y est associée.

²⁹ Nous proposons ce néologisme pour qualifier les actions non nécessairement motrices réalisées sur la base d'une captation quelconque. Par exemple, les enceintes connectées - que certains intègrent à la grande famille

Quoique très abstraites et très générales, les approches de Bryson et de Sciamma nous semblent mettre explicitement le doigt sur le principe essentiel de « la » robotique en général : celui du rapport sensorimoteur au monde et ce, indépendamment de toute considération utilitaire *a priori*. En outre, elles présentent l'originalité (implicitement chez Bryson, explicitement chez Sciamma) de ne pas penser le robot uniquement comme une catégorie « native » d'artefact, mais aussi comme une *propriété* et, de ce fait, comme devenir possible. En d'autres termes, n'importe quel objet peut devenir un robot dès lors qu'il se voit équipé des technologies lui permettant d'acquérir une forme de sensorimotricité (c'est ce qui pousse Sciamma à former la notion de « objet » sur laquelle nous reviendrons dans le point 2.1.2.1 dédié aux « objets à comportements »). Le robot ne voit pas donc forcément le jour en tant que tel, mais peut le devenir. Voilà qui renforce la pertinence et la difficulté de notre question d'ouverture ; car si le robot ne désigne pas nécessairement un type bien spécifique d'artefact, mais aussi une propriété « intégrable » à d'autres artefacts, qu'en est-il du robot social ?

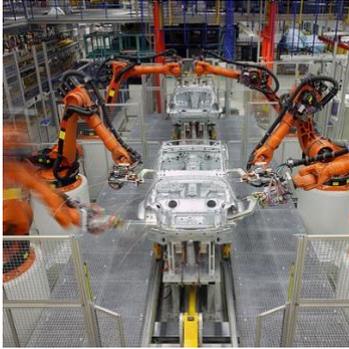
1.5 Taxonomie robotique

Complétons à présent ces éléments définitionnels par des exemples concrets d'artefacts que nous qualifions habituellement de robots. La suite et fin de ce chapitre est ainsi consacrée à la présentation d'artefacts répondant aux catégories taxonomiques suivante : robotique industrielle, robotique domestique et servicielle, robotique médicale, robotique militaire/intervention, robotique scientifique, robotique de transport et, enfin, robotique éducative.

1.5.1 Robotique industrielle

Les bras articulés industriels Kuka

des robots - agissent en fonction des données vocales captées par leur microphone sans que cela n'implique une quelconque forme de motricité. Ces actions peuvent concerner directement son environnement (e.g. allumer des ampoules elles aussi « connectées ») ou indirectement (e.g. lancer une radio ou une liste de lecture musicale).



Kuka System GmbH©



Un LBR iiwa – Kuka Laboratories GmbH©

Kuka est un des principaux chefs de file de la robotique industrielle. D'origine allemande, l'entreprise est récemment passée sous pavillon chinois suite à son rachat par *Midea*, une firme spécialiste de l'électroménager. *Kuka* propose une vaste gamme de bras articulés destinés, selon leur taille et leur usage, à des environnements différents (chaines de montage, fonderies, ateliers, entrepôts etc.).

Le KR 1000 titan, par exemple, est destiné aux charges les plus lourdes (entre 750 et 1300 kg). Disposant de six axes de rotation, il est conçu pour déplacer des pièces et des composants de façon sûre et précise. Ses domaines d'application sont nombreux : manutention, palettisation, emballage, soudage par point, application, peinture, collage, mesure, contrôle, usinage, assemblage, coupe, séparation. Outre cette flexibilité applicative et la sécurité qu'ils assurent, ces bras sont bien entendu utilisés pour leur précision et leur impact sur la productivité.

Le LBR iiwa (en haut à droite), quant à lui, est un des plus petits modèles du constructeur. Son nom est une succession d'acronymes : LBR pour « *Leichtbauroboter* » (robot léger) et iiwa pour « *intelligent industrial work assistant* ». Présenté comme la « troisième main de l'homme », il représente un type de robot nommé *cobot* (Claverie *et al*, 2013 ; Kleinpeter, 2015), contraction de « *collaborative robot* ». Cet artefact est pensé pour agir avec l'humain de façon coordonnée et synchrone, bien qu'il puisse également fonctionner seul. Le LBR iiwa est conçu pour les petites charges (de 7 à 14kg). *Kuka* le qualifie de « sensitif » du fait de ses capacités fines de détection : il adapte la force et la vitesse qu'il déploie en fonction des contacts captés, il repère les bonnes positions de montage et peut rapidement trouver les bonnes pièces, à installer, notamment de très petite taille, sans supervision humaine. *L'iiwa* est également doté de capacités d'« apprentissage » grâce à la programmation par simulation. Sa nature cobotique se traduit également par une dimension dite « haptique » : le bras peut être manipulé (un retour de force est présent pour éviter les manipulations trop

Brusques) pour interrompre ses mouvements automatiques et le guider manuellement si nécessaire.

Le robot de manutention Handle de Boston Dynamics

Handle est un robot de manutention en cours de développement chez *Boston Dynamics*, célèbre société née de sa scission avec le MIT avant d'être rachetée par *Alphabet* (maison mère de *Google*), puis par la holding japonaise *SoftBank*. Le Handle se présente comme un bras articulé doté d'un préhenseur à vide pour manipuler les cartons et de deux « jambes » sur roue pour se déplacer. L'équilibre du tout est assuré par une sorte de « queue » qui, faisant balancier, permet également au robot de se déplacer de façon dynamique.

1.5.2 Robotique domestique et servicielle

Les robots de cuisine multifonctions

Le robot-cuisine multifonction est peut-être l'un des exemples les plus célèbres de robot tangible du fait de sa nature grand public. Bien que sa cinétique ne soit qu'interne (i.e. qu'il ne déplace pas sa structure physique elle-même dans l'espace), il correspond bien à la définition générale que nous avons forgée. Cet automate peut assister voire totalement remplacer l'humain dans la transformation des aliments en cuisine. On l'utilise pour cuire (cuisson thermique ou vapeur), couper, émincer, râper, battre et peser les aliments.

Le robot Hair-wash de Panasonic

Présenté au public en 2010 puis abandonné en 2015, le *hair-wash robot* (aussi présenté comme *head-care robot* ou *shampoo robot*) était un des dispositifs développés par *Panasonic* pour répondre aux problèmes soulevés par le vieillissement de la population japonaise³⁰. L'appareil pouvait mesurer le crane de l'utilisateur pour adapter le lavage à l'unicité de sa

³⁰ Voir l'article dédié sur le site internet de *Panasonic* : <https://news.panasonic.com/global/topics/2012/10508.html>

forme. Après avoir mouillé les cheveux à l'eau chaude et appliqué le shampoing, le *hair-wash* les lavait tout en massant le crane grâce à 24 « doigts » mécatroniques.

Les robots tondeuses et aspirateurs



Le Roomba de iRobot – Kārlis Dambrāns©

La robotique domestique compte également dans ses rangs des robots aspirateurs et des robots tondeuses, dont les ventes tendent à exploser, en particulier pour les premiers. Sans surprise, le principal argument commercial des constructeurs porte sur le gain de temps que l'automatisation des corvées domestiques représente.

1.5.3 Robotique médicale

Le système chirurgical robotisé da Vinci d'Intuitive Surgical



Cmglee©



Ars Electronica©

Le da Vinci développé par l'entreprise américaine *Intuitive Surgical* est un système robotique d'intervention chirurgicale mini-invasive et de haute précision, également appelé robot de téléopération. Il est composé d'une « console chirurgien », d'un « chariot patient » et d'une

« colonne vidéo ». Le système permet d'effectuer des interventions avec une dextérité, une précision et une liberté de mouvement accrues. Concrètement, le chirurgien opère à partir de la console ergonomique où il dispose d'une vision en trois dimensions du champ opératoire et de manettes permettant de commander les quatre bras articulés du chariot patient. Outre son impact sur la qualité de l'opération elle-même, la grande précision du système permet de diminuer la durée d'hospitalisation, les douleurs post-opératoires, les risques d'infection, les probabilités de transfusion sanguine ainsi que le temps de rétablissement nécessairement au retour à une activité normale.

Le système endoscopique robotisé Monarch Platform d'Auris Health

La Monarch Platform développée par l'entreprise américaine *Auris Health* est un bronchoscope robotisé permettant d'examiner les voies aériennes des patients afin d'établir un diagnostic et des procédures thérapeutiques en conséquence. A l'heure actuelle, la firme concentre principalement sa R&D sur la détection du cancer du poumon qui est, à lui seul, plus meurtrier que les cancers du sein, du colon et de la prostate réunis. Plus précisément, le Monarch facilite la détection et l'échantillonnage des petits nodules pulmonaires potentiellement cancéreux pouvant passer inaperçus avec d'autres dispositifs. Le système se compose d'un chariot doté de deux bras robotiques téléopérés par le médecin à l'aide d'une manette similaire à celles utilisées pour les jeux vidéo. Ces bras guident la gaine télescopique avec précision à travers les bronches tout au long de l'examen.

Le robot désinfectant d'UVD Robots

UVD Robots est une entreprise danoise qui, depuis 2018, commercialise un robot de stérilisation par rayonnement ultra-violet (UV). Le type d'UV émis, les UV-C, sont ceux dont la charge énergétique, mais aussi la nocivité sont les plus élevées. Les lampes à UV-C sont couramment utilisées dans les laboratoires de biologie pour leurs effets germicides. Ils permettent de stériliser efficacement les lieux et leurs équipements. En effet, ce type de rayons ultra-violet est suffisamment puissant pour détruire l'ADN ou l'ARN de n'importe quel microorganisme. Une fois les bonnes mesures de protection adoptées, cette technologie de stérilisation est donc l'une des plus efficaces et des plus sûres pour garantir une hygiène

maximale dans les hôpitaux. Le robot UVD est une base mobile qui, outre les lampes à UV-C, embarque des lidars³¹ permettant de scanner son environnement et d'en produire une carte numérique. Il suffit ensuite d'annoter cette carte pour indiquer à l'artefact les points où il doit s'arrêter. Une fois programmé, le robot navigue grâce à un logiciel SLAM (*Simultaneous Localization and Mapping*) lui permettant d'effectuer son circuit de stérilisation de façon autonome. Pour des raisons de sécurité, les robots UVD sont actifs si et seulement si aucun humain n'est présent dans leur périmètre. S'ils captent un mouvement ou détectent une personne, leurs lampes à UV s'éteignent automatiquement. La stérilisation d'une chambre dure généralement entre dix et quinze minutes, à raison d'une à deux minutes par point d'arrêt identifié en amont du circuit.

1.5.4 Robotique militaire et d'intervention

Le Colossus de Shark Robotics



Shark Robotics©

Le Colossus de l'entreprise française *Shark Robotics* est décrit par ses concepteurs comme un « robot de soutien polyvalent destiné à intervenir dans les zones à risque. »³² Développé avec la brigade des Sapeurs-Pompiers de Paris, on a pu le voir intervenir lors de l'incendie de la cathédrale Notre-Dame de Paris du 15 avril 2019³³. Le Colossus fait partie des robots d'intervention télécommandés, c'est-à-dire dirigés à distance par un opérateur humain, ce qui pousse parfois à les qualifier de « drones terrestres ».

³¹ Acronyme de *Light/Laser Detection and Ranging*.

³² Voir le site internet de *Shark Robotics* : <https://www.shark-robotics.com/robots>

³³ Voir, par exemple, cet article du *Huffingtonpost* : https://www.huffingtonpost.fr/entry/colossus-le-robot-qui-a-aide-les-pompiers-a-eteindre-lincendie-de-notre-dame_fr_5cb6a07ce4b0ffefe3b8c661

Spot, la « plateforme polyvalente » de Boston Dynamics



ArticCynda©



Senior Airman John Ennis©

Spot est un robot d'intervention polyvalent développé par *Boston Dynamics*. Comme tous les produits de Boston Dynamics, il excelle avant tout par ses capacités de mouvement et d'équilibre. Des escaliers au terrain escarpé d'une zone d'intervention (militaire ou non), sa réputation est née de son habileté et de sa fluidité cinétique. Parfois qualifié de robot-chien (« *dog-robot* »), l'automate peut marcher, trotter, éviter les obstacles, grimper des surfaces inclinées etc. Spot est présenté comme une « plateforme » robotique polyvalente que les entreprises clientes peuvent assembler selon leurs besoins. Cette polyvalence se concrétise notamment dans la modularité de l'artefact : son matériel (*hardware*) de base est présenté comme personnalisable dans sa quasi intégralité et 14kg d'outils additionnels peuvent lui être greffés comme un bras articulé doté de six degrés de liberté de mouvement ou une « SpotCam » permettant de capturer des images sphériques de son environnement. *Boston Dynamics* n'a donc pas développé son robot avec une application précise en tête, mais plutôt avec un ensemble de capacités « génériques » permettant une forte générativité d'usage *a posteriori*.

« What types of applications? The company says the best way to find out is to put Spot in the hands of as many users as possible and let them develop the applications. [...] “We have some initial guesses about which markets could benefit most from this technology, and we’ve been engaging with customers doing proof-of-concept trials,” [Boston Dynamics’ VP of business development Michael] Perry says. “But at the end of the day, that value story is really going to be determined by people going out and exploring and pushing the limits of the robot » (Guizzo, 2019).

Les suppositions initiales dont parle Michael Perry concernent par exemple la collecte de données diverses à distance, la manipulation de poids légers sur des chantiers, le contrôle des capteurs et de l'infrastructure des sites gaziers ou pétroliers, ou encore l'accomplissement de missions périlleuses comme le désamorçage de bombe ou l'inspection de matières dangereuses.

Teodor, le robot démineur de Telerob



Le Teodor (« *Telerob Explosive Ordnance Disposal and Observation Robot* ») de l'entreprise allemande Telerob est un robot de déminage souvent utilisé par la police, notamment lors d'opérations antiterroristes. Le dispositif se veut toutefois polyvalent et peut être mobilisé pour faire face aux risques NRBC³⁴ (nucléaire, radiologique, biologique, chimique) ou lors d'incendies. Comme le français Colossus, le Teodor est téléopéré, par signal radio ou par fibre optique, avec une portée maximale de 200 mètres. Outre sa pince articulée, l'artefact peut embarquer quatre équipements supplémentaires selon le contexte d'intervention (marteaux, perceuses, mousquetons, scie simple ou circulaire, pic à glace...)

1.5.5 Robotique scientifique

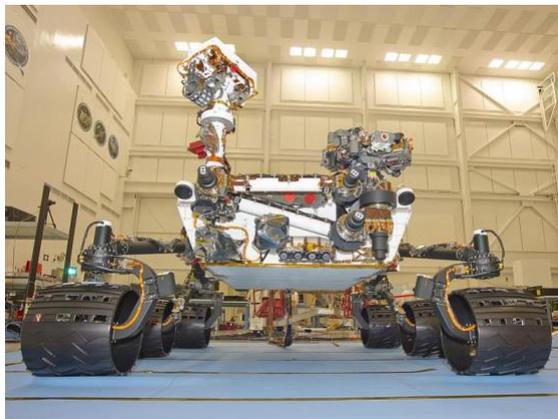
Curiosity, l'astromobile d'exploration de la NASA

Curiosity est l'astromobile (ou *rover*) d'exploration conçue par le *Jet Propulsion Laboratory* (JPL) de la NASA pour la mission *Mars Science Laboratory* (MSL) lancée le 26 novembre 2011.

³⁴ CBRN en anglais.

MSL a pour principaux objectifs de déterminer si des conditions propices à la vie ont existé sur Mars, étudier le climat et la géologie de la planète rouge, et préparer sa future exploration humaine. Le Curiosity est donc équipé en conséquence. Citons quelques-uns des instruments embarqués sur le rover :

- MAHLI (*Mars Hand Lens Imager*) : le MAHLI est une adaptation des loupes que les géologues utilisent pour révéler les minéraux et les textures des surfaces rocheuses.
- APXS (*Alpha Particle X-Ray Spectrometer*) : l'APXS permet de mesurer par radioactivité le taux et le type d'éléments chimiques présents dans un sol ou une roche.
- ChemCam (Chemistry and Camera) : la Chemcam du rover est composée d'un laser, d'une caméra et d'un spectrographe dédiés à l'analyse de la composition chimique des sols et des roches martiennes.
- DAN (*Dynamic Albedo of Neutrons*) : le DAN est outil qui traque les différences de relâchements neutroniques par le sol martien. Ces changements peuvent indiquer la présence d'eau liquide ou gelée en sous-sol.
- REMS (*Rover Environmental Monitoring Station*) : le REMS permet de fournir des compte-rendu quotidiens et saisonniers des conditions météorologiques aux environs du rover.



NASA/JPL-Caltech

Si l'astromobile reste majoritairement pilotée par l'humain depuis la Terre, la distance nous séparant de Mars accroît considérablement les temps de communication entre l'artefact et ses opérateurs humains. Ainsi, depuis mai 2016, Curiosity est doté d'un logiciel nommé AEGIS (*Autonomous Exploration for Gathering Increased Science*) lui permettant de repérer « par lui-même » les caractéristiques du sol martien candidates à l'inspection et d'ajuster la mire du

laser de sa ChemCam³⁵. Cette forme, certes minime, d'autonomie représente une des premières solutions de l'IA permettant de contourner les contraintes spatio-temporelles imposées par l'exploration extraterrestre.

1.5.6 Robotique de transport

Le VAL : Véhicule Automatique Léger



Une rame du VAL de Toulouse – Maxime Lafage©

Lorsque l'on parle de robotique de transport, les métros automatiques ou VAL (Véhicules Automatiques Legers) sont souvent les premiers évoqués. Les VAL sont des métros légers sur pneus totalement automatiques. Ils sont issus des recherches menées dans les années 1970 au sein des laboratoires de l'université Lille-1 et de l'École centrale de Lille. Depuis leur apparition, les VAL se sont répandus en France, mais aussi à l'étranger. On peut monter à leur bord à Toulouse, Rennes ou Paris, mais aussi à Turin, Taipei, Chicago ou encore Uijeongbu en Corée du Sud. Contrairement au SAET (Système d'Automatisation de l'Exploitation des Trains) déployé par la RATP dans le métro parisien³⁶, les VAL n'autorisent ni la mixité du trafic³⁷ ni la non-polarisation des rames³⁸.

³⁵ Voir, par exemple, cet article de *The Atlantic* : <https://www.theatlantic.com/technology/archive/2017/06/mars-curiosity-rover/531339/> (consulté le 30/04/2020).

³⁶ Le SAET est actuellement actif sur les lignes 1 et 14.

³⁷ Possibilité de faire circuler simultanément des rames intégralement automatiques et des rames manuelles ou semi-automatiques sur la même ligne.

³⁸ Qui permet de ne pas tenir compte du sens d'une rame automatique avant son intégration à la ligne

Hercules de Unity Drive, un véhicule autonome de transports de marchandises

Le Hercules est un véhicule autonome de livraison conçu par Unity Drive Innovation, une entreprise chinoise basée à Shenzhen. Déployé à Zibo, Suzhou et Shenzhen pour livrer des paniers repas aux travailleurs des postes de contrôle et des sprays désinfectants aux alentours des hôpitaux dans le cadre de la crise sanitaire de la Covid-19, le robot était déjà en service fin 2018 pour acheminer des composants électroniques et d'autres produits sur l'un des campus de Foxconn situé à Shenzhen où travaillent et vivent quelques 200 000 personnes. Hercules dispose d'un nombre important de capteurs : un lidar³⁹ principal, trois lidars auxiliaires, une caméra stéréo, quatre caméra « fisheye », un système de navigation redondante par satellite, une unité de mesure inertielle (UMI) et deux codeurs rotatifs. L'ordinateur du véhicule reçoit les nuages de points du lidar et les transmet à l'algorithme principal de perception, un réseau neuronal convolutif entraîné pour détecter et classer les objets de son environnement. Deux autres algorithmes sont respectivement chargés d'identifier la signalisation routière par l'intermédiaire des caméras avant du véhicule et d'associer les nuages de points des lidars aux données de l'UMI pour qu'il se géolocalise lui-même. Le service proposé par UDI fonctionne comme n'importe quel autre système numérique applicatif : l'utilisateur réserve un Hercule via son téléphone pour que le véhicule aille chercher son produit et lui livre au point de collecte de son choix.

Les véhicules autonomes de transport humain



Une voiture autonome Waymo – elisfkc©

³⁹ Lidar est l'acronyme de « *light detection and ranging* ». Il s'agit d'un système de détection et d'estimation de la distance par la lumière.

Ces dernières années, la robotique de transport à beaucoup fait parler d'elle par le biais des fameuses « voitures autonomes », destiné au transport de personnes, notamment suite au tristement célèbre accident d'Uber survenu en 2018 à Tempe, Arizona⁴⁰. La SAE International (*Society of Automotive Engineers*) distingue 5 niveaux d'autonomie couramment repris par les constructeurs, journalistes et autres commentateurs de l'actualité technologique.

SAE level	Name	Narrative Definition	Execution of Steering and Acceleration/Deceleration	Monitoring of Driving Environment	Fallback Performance of Dynamic Driving Task	System Capability (Driving Modes)
Human driver monitors the driving environment						
0	No Automation	the full-time performance by the <i>human driver</i> of all aspects of the <i>dynamic driving task</i> , even when enhanced by warning or intervention systems	Human driver	Human driver	Human driver	n/a
1	Driver Assistance	the <i>driving mode</i> -specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the <i>human driver</i> perform all remaining aspects of the <i>dynamic driving task</i>	Human driver and system	Human driver	Human driver	Some driving modes
2	Partial Automation	the <i>driving mode</i> -specific execution by one or more driver assistance systems of both steering and acceleration/deceleration using information about the driving environment and with the expectation that the <i>human driver</i> perform all remaining aspects of the <i>dynamic driving task</i>	System	Human driver	Human driver	Some driving modes
Automated driving system ("system") monitors the driving environment						
3	Conditional Automation	the <i>driving mode</i> -specific performance by an <i>automated driving system</i> of all aspects of the <i>dynamic driving task</i> with the expectation that the <i>human driver</i> will respond appropriately to a <i>request to intervene</i>	System	System	Human driver	Some driving modes
4	High Automation	the <i>driving mode</i> -specific performance by an automated driving system of all aspects of the <i>dynamic driving task</i> , even if a <i>human driver</i> does not respond appropriately to a <i>request to intervene</i>	System	System	System	Some driving modes
5	Full Automation	the full-time performance by an <i>automated driving system</i> of all aspects of the <i>dynamic driving task</i> under all roadway and environmental conditions that can be managed by a <i>human driver</i>	System	System	System	All driving modes

Copyright © 2014 SAE International. The summary table may be freely copied and distributed provided SAE International and J3016 are acknowledged as the source and must be reproduced AS-IS.

A l'heure actuelle, aucun constructeur ne semble pouvoir dépasser le niveau 3 en conditions ouvertes. Si certains d'entre eux affirment envisager un déploiement prochain de véhicules de niveau 4 en circuit fermé, la question de la viabilité commerciale de tels projets demeure. En outre, l'industrie du véhicule autonome ne cible pas uniquement le transport individuel et ses longs trajets potentiels, mais aussi et peut-être surtout le transport collectif de courte distance. Cruise (voir l'image en haut à droite), une filiale de General Motors, a récemment révélé son dispositif de co-voiturage intra-urbain baptisé Origin. Totalement électrique et pouvant accueillir jusqu'à six passagers, Origin n'est pas pensé pour la vente de détail. Il s'agit pour Cruise de mettre à disposition des villes une flotte de véhicules permettant de généraliser la mobilité urbaine en covoiturage. Aussi importante soit-elle, la technique ne

⁴⁰ Voir, par exemple, cet article de *The Verge* qui revient sur l'enquête du National Transportation Safety Board et la chaîne de responsabilités qu'elle mit en lumière : <https://www.theverge.com/2019/11/20/20973971/uber-self-driving-car-crash-investigation-human-error-results>

décidera pas à elle seule de l'avenir de ces technologies. Si Cruise affirme qu'Origin est prêt à passer au stade de la production, l'entreprise doit toujours attendre l'aval des autorités gouvernementales pour espérer déployer son futur service.

1.5.7 Robotique éducative

Root, le robot pour apprendre à coder d'iRobot

Root est un robot éducatif initialement développé au Harvard's Wyss Institute avant son achat par iRobot en 2019. Root est conçu pour faciliter l'apprentissage du code aux enfants (à partir de quatre ans), mais aussi son enseignement par les professeurs. Le robot se présente sous la forme d'un petit pavé hexagonal prêt à l'emploi. Aimanté, il peut se déplacer sur les tableaux magnétiques aussi bien que sur les tables ou les feuilles en papier. Le centre de l'artefact peut accueillir un feutre afin de tracer toute sorte de formes géométriques ou de simples lignes selon la programmation effectuée par l'utilisateur. Root est également équipé d'un tampon qui lui permet d'effacer ce qui est écrit ou dessiné au tableau. Outre ces caractéristiques matérielles, une interface de programmation nommée « Root Square » est bien sûr prévue pour apprendre à coder selon trois niveaux de difficultés (voir image ci-dessus à droite) et observer immédiatement le résultat concret du code grâce au comportement du robot.

UnicornBot, la licorne éducative d'UBTECH

Comme Root, UnicornBot est un robot dédié à l'apprentissage du code. Il est commercialisé par HUBTECH, une entreprise chinoise basée à Shenzhen. La particularité d'UnicornBot est d'être avant tout pensé pour séduire le tout jeune public féminin, d'où sa forme de licorne. Aidée d'une application Android ou iOS affichant des instructions en 3D, sa (ou son) propriétaire doit d'abord assembler les quelques 400 pièces et connecteurs qui forment le robot. L'application permet ensuite de programmer l'artefact grâce au langage de programmation visuelle Blockly. Selon sa programmation, UnicornBot peut se déplacer, hocher la tête, réagir à des cartes de couleurs ou encore illuminer sa corne. L'enfant construit son programme en combinant des blocs de code sur le mode glisser-déposer (*drag-and-drop*).

Dans ce premier chapitre, nous nous sommes attelés à caractériser le robot à travers les trois grandes catégories analytiques du mythe, de l'automatisme et de la normativité. Nous avons ensuite proposé de le définir conceptuellement comme artefact dont le principe structurel est d'entretenir un rapport sensorimoteur (ou plus largement sensoriactif) avec le monde extérieur, que ce rapport soit finalement soumis à des fins utilitaires ou non. Une telle caractérisation peut amener à penser le robot sous l'angle du devenir-robot et non seulement comme une catégorie native d'artefacts bien particuliers. En outre, elle peut être qualifiée de phénoménologique au sens où elle identifie le principe général d'apparition du robot à la conscience humaine. Or, à ce stade, notre question introductive – « le robot social est-il avant tout un robot *tout court* ou reconfigure-t-il en profondeur ce qu'on appelle communément robot ? » - ne peut pas encore trouver de réponse. Pour ce faire, poursuivons notre état de l'art en abordant les travaux spécifiquement dédiés à la robotique sociale.

Chapitre 2 - Qu'est-ce qu'un robot social ?

Les précurseurs de ceux qu'on appelle aujourd'hui « robots sociaux » étaient des artefacts destinés à l'étude des comportements collectifs chez les insectes via des interactions entre machines (Fong *et al*, 2002). Au début des années 90, les premières expériences sur la stigmergie⁴¹ furent réalisées avec des robots fourmis et de nombreuses autres ne tardèrent pas à suivre. Conformément à la nature des sociétés animales étudiées, ces premiers « robots sociaux » étaient conçus comme des individus anonymes dont la seule raison d'être était de servir le collectif auquel ils appartenaient - c'est là, comme nous le verrons, une différence notoire avec les artefacts que nous qualifions aujourd'hui de sociobots. Comme le remarquent Fong *et al* (*ibid.*), ce modèle social fondé sur des comportements collectifs anonymes se révéla particulièrement attractif pour la robotique car il permettait à des groupes de robots assez rudimentaires d'accomplir des tâches difficiles telles que jouer au football. Puis, l'intérêt pour les sociétés mammifères et plus particulièrement humaines émergea peu à peu parmi les roboticiens. Or, contrairement aux insectes, la socialité des mammifères se caractérise par une dialectique permanente - plus ou moins prononcée selon les cas - entre l'individu et les groupes auxquels il appartient. Le modèle constitue alors un défi autrement plus périlleux que celui des initiatives précédentes. A cet égard, notons que, depuis le début des années 2000 et plus particulièrement depuis une quinzaine d'années, le champ de la robotique sociale se structure surtout autour d'artefacts destinés aux interactions humain-robot (IHR) (Hegel, 2009). Dumouchel et Damiano l'affirment d'ailleurs sans détour :

« L'appellation "robot sociaux" renvoie plus précisément, ou exclusivement, à des agents artificiels qui réalisent [leurs] fonctions grâce à leur capacité d'interagir "socialement" avec des agents humains » (Dumouchel & Damiano, 2016, p.117).

⁴¹ Introduit par le zoologiste français Pierre-Paul Grassé, le terme de stigmergie désigne un mécanisme de communication et de coordination indirecte entre individus par modification de leur environnement. Un comportement stigmergique consiste, pour un individu, à laisser des « traces » dans cet environnement pour stimuler l'action d'un autre individu après lui. L'exemple type est celui des fourmis ou des termites qui communiquent par dépôt de phéromones, respectivement pour s'orienter vers une source de nourriture et/ou pour revenir à la colonie, ou pour construire leur nid. Aujourd'hui, la stigmergie est parfois utilisée pour désigner la notion d'intelligence collective dans les groupes humains.

L'étonnement que suscite la notion de robot social redouble face à l'impressionnante prolifération définitionnelle dont elle fait l'objet. En outre, un certain nombre de travaux mobilisent des catégories qui, sans nommer directement le robot social ou l'englobant dans un ensemble plus vaste, mettent chacune en avant une de ses caractéristiques fondamentales. Ces notions sont principalement celles d'« objets à comportements » (Levillain & Zibetti, 2017), d'« artefacts relationnels » (Turkle, 2015), de « substitut » (Dumouchel & Damiano, 2016) et de « robot existentiel » (Lestel, 2021). Elles mettent respectivement l'accent sur le mouvement intentionnel, les émotions et l'empathie, l'autonomie sociale et la remise en question du statut de vivant. Dans la première partie de ce chapitre, nous proposons de passer en revue les principales définitions et notions en vigueur dans la littérature afin de dresser un premier portrait du sociobot. Nous verrons qu'aucun concept de robot social ne semble exister aujourd'hui. Dans la seconde partie, nous interrogeons plus spécifiquement l'état de l'art à l'aune de notre grande question directrice, à savoir « que signifie pour un robot d'être social ? » Nous identifions trois réponses respectivement centrées sur l'interaction, la fonction et l'interactance.

2.1 Portrait-robot du sociobot : terminologie et définitions en vigueur

2.1.1 Les approches généralistes de la robotique sociale

Développer une philosophie empirique et pragmatiste du sociobot suppose, bien sûr, de « faire parler » l'objet lui-même, mais aussi d'observer comment chercheurs et concepteurs parlent de lui, notamment à travers les définitions qui lui sont attribuées dans la littérature. En tant qu'elles fixent des limites claires et cherchent à assurer une convergence sémantique permettant une compréhension mutuelle (a minima au sein d'un groupe partageant le même bagage théorique et professionnel), les définitions expriment toujours également des connaissances implicites et des représentations sous-jacentes qui permettent cette convergence. Dit dans une optique bourdieusienne, « les définitions sont intrinsèquement mêlées à une historicité et à une praxis » (Sarrica *et al*, 2019). Ainsi, le « de quoi parle-t-on » et le « comment en parle-t-on » ne peuvent qu'être artificiellement dissociés lorsqu'on essaie de concevoir, tant au sens pratique que conceptuel, un nouveau genre d'artefact.

L'analyse détaillée de la prolifération définitionnelle qui auréole la figure du sociobot permet, dans un premier temps, d'observer deux grands mouvements : celui des définitions pionnières de la fin des années 1990 et du début des années 2000, et celui d'une combinaison de définitions existantes (*ibid*). Les quatre principales approches fondatrices comptent, aujourd'hui encore, parmi les plus influentes (Hegel, 2009 ; Sarrica *et al*, 2019). Il s'agit des définitions proposées par Duffy *et al* (1999), Fong *et al* (2002), Breazeal (2002, 2003a) et Bartneck et Forlizzi (2004).

2.1.1.1 Robot social *versus* Robot sociétal

Dans leur article commun Duffy *et al* (1999) proposent d'effectuer une distinction entre robot sociétal et robot social. Alors qu'un robot sociétal désigne un artefact spécifiquement conçu pour intégrer la société sur la base de ses capacités d'interaction et de l'aide qu'il peut apporter aux humains, le robot social, lui, renvoie à « une représentation plus générique applicable aux interactions entre robots » et peut se définir comme

« [u]ne entité physique insérée dans un environnement à la fois complexe, dynamique et social, et suffisamment encapacitée pour se comporter d'une manière propice à la réalisation de ses propres objectifs et de ceux de sa communauté. »⁴²

Selon les auteurs, cette encapacitation renvoie à des possibilités de collaboration, de communication, de coopération, de coordination, mais aussi à la construction d'une identité et de relations sur la base de modèle comportementaux aussi bien réactifs que proactifs.

En se basant sur le nombre de citations (485) et le fait qu'elle s'affiche en premier lors d'une recherche Google, Sarrica *et al* qualifient cette approche de « définition la plus influente » du sociobot (2019). Celle-ci nous apparaît pourtant particulièrement abstraite et datée. Abstraite, d'une part, parce qu'elle mobilise des notions aussi larges et floues que son objet ; datée d'autre part, car le robot social dont on parle aujourd'hui n'est pas limité aux interactions robot-robot (IRR), bien au contraire. En outre, l'approche se révèle finalement assez ambiguë puisque Duffy finira par parler de robot social y compris pour désigner des

⁴² « A physical entity embodied in a complex, dynamic, and social environment sufficiently empowered to behave in a manner conducive to its own goals and those of its community ».

robots capables d'interagir avec d'autres robots ou avec l'humain (2003)⁴³. Cela explique peut-être pourquoi la distinction entre robot social et robot sociétal est, finalement, assez peu reprise. Pourtant, une fois son contenu reconfiguré, nous pensons qu'une telle distinction peut s'avérer utile pour singulariser la figure du sociobot par rapport à d'autres types de robots. Nous proposons donc de conserver les deux catégories de Duffy *et al*, mais de les dissocier sur la base du rapport entre interaction et fonction(s) de l'artefact plutôt que sur celle du « partenaire » d'interaction. Dans cette optique, un robot sociétal est un robot qui agit et réalise sa fonction *en société* (contrairement aux robots industriels confinés dans des usines) indépendamment de toute interaction avec l'humain. Les robots de désinfection aux rayons UV ou les robots aspirateurs comme le célèbre Roomba sont de bons exemples de robots sociétaux. Le sociobot, à l'inverse, est un robot qui interagit socialement et dont la fonction se confond avec l'interaction ou s'accomplit à travers elle. Le bébé phoque interactif Paro est peut-être le représentant le plus connu du premier cas de figure. Pensé pour servir de compagnon aux personnes âgées atteintes de troubles neurodégénératifs ou en manque de socialisation, il se contente de répondre aux caresses qu'il reçoit par divers mouvements et petits cris. On peut alors dire que la fonction de Paro est intrinsèque à l'interaction (FII) au sens où l'interaction produit ses effets en tant que telle et qu'elle est ainsi visée pour elle-même. À l'inverse, un robot comme Pepper qui se voit souvent confier des tâches d'accueil et d'information dans les magasins, les gares et autres lieux publics exécute une fonction extrinsèque à l'interaction (FEI) car elle pourrait se réaliser par d'autres moyens. Cette prise en charge de FII ou de FEI n'est toutefois pas nécessairement liée au sociobot lui-même. Autrement dit, on peut tout à fait imaginer un même robot s'acquitter de FII dans un certain contexte et réaliser des FEI dans un autre. Cela ne signifie pas pour autant que tout sociobot possède les capacités techniques de prendre en charge les deux types de fonction. En effet, cette aptitude est déterminée par le design au sens large de l'artefact. Ainsi Paro est-il exclusivement conçu pour les FII, contrairement à Pepper qui peut passer du rôle d'hôte d'accueil à celui de compagnon.

⁴³ Notons toutefois que c'est peut-être en partie pour son caractère abstrait que cette définition est assez massivement citée (aux côtés d'un possible phénomène mimétique). En outre, une fois extraite du reste de l'article, elle ne semble pas limiter le robot social aux IRR. En effet, l'idée de « communauté » ne précise pas de quelle communauté il s'agit. Il est simplement question de « sa » communauté. Celle-ci peut donc renvoyer à une communauté de robots ou plus largement d'agents sociaux dans laquelle il y a plus d'humains que de robots (voire pas de robot du tout).

2.1.1.2 Robot socialement interactif

Une deuxième approche souvent citée est celle proposée par Dautenhahn et Billard (1999) et reprise par Fong *et al* (2002) dès l'introduction de leur état de l'art de référence :

« Les robots sociaux sont des agents incarnés faisant partie d'un groupe hétérogène : une société de robots ou d'humains. Ils sont capables de se reconnaître les uns les autres et de s'engager dans des interactions sociales, ont des histoires (perçoivent et interprètent le monde dans les termes de leur propre expérience) et communiquent explicitement entre eux tout en apprenant les uns des autres. »

Les auteurs apportent toutefois une précision d'importance lorsqu'ils affirment parler de robots sociaux au sens de « robots socialement interactifs » (*socially interactive robots*), c'est-à-dire de « robots pour lesquels l'interaction sociale joue un rôle clé » (Fong *et al*, p.3). Ce choix terminologique n'a pas vocation à créer une nouvelle classe d'artefact, mais à distinguer les robots sociaux d'autres robots impliquant des IHR « conventionnelles » d'ordre manipulateurs (i.e. cobots industriels ou robots téléopérés). Bien que certains robots sociaux puissent interagir avec d'autres robots, Fong *et al* précisent que leur étude est dédiée aux IHR et qu'ils s'intéressent aux artefacts manifestant les caractéristiques sociales humaines (« *human social* » *characteristics*) suivantes :

- Exprimer et/ou percevoir les émotions,
- Communiquer sur un mode dialogique de haut niveau (*high-level dialogue*),
- Apprendre et reconnaître les modèles des autres agents,
- Établir et maintenir des relations sociales,
- Utiliser des « signaux naturels » (*natural cues*) comme le regard ou la gestuelle,
- Manifester une personnalité distinctive et un caractère, apprendre et développer des compétences sociales.

Contrairement à Fong *et al* qui utilisent la notion de robot socialement interactif comme synonyme (plus précis) de robot social, nous pensons que l'intérêt principal d'une telle distinction terminologique se situe précisément dans l'idée que le social n'est pas intégralement identifiable à l'interaction. C'est un point majeur sur lequel nous reviendrons dans les chapitres suivants.

2.1.1.3 Robot sociable

La troisième grande approche pionnière que l'on peut citer est celle de Cynthia Breazeal avec sa notion de robot sociable (*sociable robot*). Dans le livre pionnier *Designing Sociable Robots* (2002), Breazeal ne cache pas son embarras face aux questions de définition et commence par situer son référentiel dans le champ de la science-fiction :

« Qu'est-ce qu'un robot sociable ? Il s'agit d'un concept difficile à définir, mais la science-fiction offre de nombreux exemples. Il y a les droïdes mécaniques R2-D2 et C-3PO du film *Star Wars* et l'androïde Lt Commander Data de la série télévisée *Star Trek : The Next Generation*. De nombreux et superbes exemples existent dans les nouvelles d'Isaac Asimov et de Brian Aldiss, tels que les robots Robbie (Asimov, 1986) et David (Aldiss, 2001). »

C'est donc sur la base de cet imaginaire et des représentations qui le nourrissent que la chercheuse du MIT propose une définition qui relève plus de la vision programmatique que de la caractérisation d'une technologie existante. Citons-là assez longuement :

« Pour moi, un robot sociable est capable de communiquer et d'interagir avec nous, de nous comprendre voire même de s'identifier à nous, et tout cela de façon personnelle. Il devrait être capable de nous comprendre et de se comprendre en termes sociaux. En retour, nous devrions le comprendre dans les mêmes termes sociaux – afin de pouvoir nous identifier à lui et d'être en empathie avec lui. Un tel robot doit être capable de s'adapter et d'apprendre tout au long de sa durée de vie, en intégrant les expériences partagées avec d'autres individus à la compréhension qu'il a de lui-même, des autres, et des relations qu'ils partagent. Pour résumer, un robot sociable est doté d'une intelligence sociale similaire de sorte que nos interactions avec lui rappellent les interactions avec une autre personne. Au pinacle de leur achèvement, ils pourraient se lier d'amitié avec nous et nous avec eux. La science-fiction montre comment ces technologies pourraient améliorer nos vies et bénéficier à la société, mais elle nous avertit également que ce rêve doit être approché de façon éthique et responsable, comme le montre Philip K. Dick dans *Les androïdes rêvent-ils de moutons électriques ?* »

Une telle description souligne la nature technodiscursive du sociobot : celui-ci existe aussi bien dans le discours au sens large que dans la technique. D'ailleurs, le robot sociable de Breazeal

n'existe, pour le moment, *que* dans le discours. Le soudain passage du présent de l'indicatif au conditionnel et l'emploi du mot « rêve » témoignent du référentiel éminemment projectif de la roboticienne. La science-fiction est source de prescription au sens où elle crée des personnages robotiques qui, plus que des figures d'inspiration, représentent des modèles à suivre. Or, dans la plupart des exemples cités, on constate que les robots se distinguent finalement assez peu des humains, si ce n'est par leur nature inorganique ou éventuellement par leur apparence. Dès lors, la socialité robotique dont il est question est, comme dans l'exemple précédent, ouvertement entendue comme socialité *humaine*.

Il faut toutefois préciser que, pour Breazeal, la figure du robot sociable est un idéal⁴⁴ qui ne résume pas à lui seul la robotique sociale. En effet, la roboticienne semble concevoir la socialité sur un continuum aboutissant au très futuriste robot sociable et sur lequel on trouve d'autres entités certes socialement moins évoluées, mais qui existent concrètement. Breazeal propose ainsi de classer ces artefacts selon leurs capacités individuelles à prendre en charge des scénarios d'interactions sociales plus ou moins complexes (2003a). Quatre catégories composent cette classification par *degrés de socialité* :

- 1) Les robots « socialement évocateurs » (*socially evocative*) : ce très faible degré de socialité se limite à la convocation d'un sentiment anthropomorphe.
- 2) Les robots « interface sociale » (*social interface*) : ces artefacts peuvent émettre des signaux sociaux (regards, gestuelle, mouvements, sons, mimiques etc.) et communiquer dans une certaine mesure avec l'humain.
- 3) Les robots « socialement réceptifs » (*socially receptive*) : outre des capacités de communication égales ou supérieures aux robots socialement communicatifs, ces sociobots sont capables d'apprendre de leurs interactions avec nous. De ce fait, ils peuvent percevoir et s'adapter aux signaux sociaux humains, ce qui permet aux individus d'adopter un comportement plus complexe envers eux. Ils restent toutefois socialement passifs en ce sens qu'ils répondent aux interactions engagées par l'interactant sans jamais chercher à les provoquer eux-mêmes.

⁴⁴ Au sens pragmatiste de modèle vers lequel tendre petit à petit grâce à la pratique et, en l'occurrence, grâce aux progrès scientifiques et techniques.

- 4) La quatrième catégorie présentée dans l'article est celle des robots sociables. Breazeal la décrit comme une classe d'artefacts socialement participatifs dotés de leur propre système téléologique, capables d'engager eux-mêmes des interactions avec les humains dans une optique d'assistance, mais aussi pour satisfaire leurs propres objectifs.

Breazeal présente elle aussi plusieurs conditions de possibilité – qu'elle qualifie d' « ingrédients » (2002, p.6) – du robot sociable. Elles sont au nombre de cinq :

- a) « Être là » (*Being there*) : un robot social doit être incarné (*embodied*) en ce que le corps est au fondement de nos interactions avec le monde en général et autrui en particulier. C'est par le corps que nous percevons, situons et comprenons ce qui nous est extérieur. C'est aussi par le corps que nous développons la plupart de nos facultés de communication et, plus largement, nos diverses aptitudes sociales.
- b) « Ressembler au vivant » (*Life-like quality*) : pour être qualifié de social, un robot doit avant tout nous apparaître comme tel. Dès lors un sociobot ne peut exister que s'il manifeste une forme d'intentionnalité et d'autonomie.
- c) « Perception et compréhension de l'humain » (*Human-aware*) : un sociobot doit être capable de reconnaître et de comprendre la ou les personnes avec qui il interagit, mais aussi ce qu'elles font et comment elles le font. Breazeal associe notamment ces capacités à la théorie de l'esprit⁴⁵ et à l'empathie qui, toutes deux, permettent de comprendre, d'expliquer et de prédire les comportements sociaux d'autrui et d'y répondre de façon appropriée.
- d) « Être compris » (*Being understood*) : si le sociobot doit comprendre autrui, il doit aussi se comprendre lui-même pour *se faire* comprendre. Pour communiquer, construire et maintenir des échanges avec autrui, la lisibilité (*readability*) de l'artefact est donc primordiale. Cela passe en grande partie par l'émission de signaux sociaux qui nous sont familiers et compréhensibles (expressions faciales, gestuelle, posture, mouvements etc.).

⁴⁵ Qu'elle définit comme « ces aptitudes sociales qui permettent à l'humain d'attribuer correctement à autrui ou à soi-même des croyances, des objectifs, des perceptions, des sentiments et des désirs » (Breazeal, 2002, p.8-9).

- e) « Apprentissage social » (*Socially situated learning*) : pour devenir sociable, un robot doit pouvoir apprendre en permanence par lui-même et par ses interactions avec autrui ou son environnement. Cela va de pair avec la possibilité d'expérimenter et de se constituer une histoire tout en développant de nouvelles aptitudes et compétences sociales.

Malgré des différences terminologiques et une ambition programmatique qui assume son lien fondateur à la science-fiction, l'approche de Breazeal n'est pas foncièrement différente des précédentes. Elle met en lumière la plupart des notions fondamentales de la sociobotique telles que la communication verbale et non verbale, l'interaction sociale, l'autonomie, l'identification, l'empathie, l'anthropomorphisme, l'apprentissage, l'adaptation ou encore l'expérience en première personne.

2.1.1.4 Normes et expériences sociales : le sociobot de l'UX design

La quatrième et dernière contribution de référence est celle proposée par Bartneck et Forlizzi (2004) dans une optique centrée design (*design-centered approach*). C'est aussi celle qui introduit explicitement un élément nouveau par rapport aux précédentes : l'importance des normes sociales humaines qui cadrent les interactions. Selon les deux auteurs,

« [u]n robot social est un robot autonome ou semi-autonome qui communique et interagit avec les humains en respectant les normes comportementales attendues par celles et ceux avec qui il est censé interagir » (*ibid.*).

En faisant référence aux normes sociales, Bartneck et Forlizzi ne rompent pas avec l'idée classique que le robot social est avant tout un robot « socialement interactif », mais précisent que cette interactivité n'est pas uniquement informée par des règles d'ordre physique et technique. Autrement dit, la sociobotique n'est pas surdéterminée techniquement, mais socialement. Voilà pourquoi, dans leur définition,

« être social est lié à la compréhension et, dans certains cas, à l'imitation de l'activité humaine, de la société et de la culture environnantes qui forment des valeurs sociales, des normes et des standards. Par exemple, un robot majordome devrait se conformer aux règles établies en matière de bon service. Il devrait anticiper, être fiable et, par-dessus

tout, être discret. En revanche, ses activités précises sont susceptibles de varier selon les cultures puisque les valeurs sociales, les normes et les standards sont eux-mêmes culturellement relatifs » (*ibid.*).

Cette approche n'est donc pas simplement prescriptive sur le pan technique, mais l'est aussi sur le plan social.

Outre la question de la normativité sociale, Bartneck et Forlizzi introduisent un autre élément nouveau : celui des expériences générées par le robot. Tous deux conçoivent le sociobot comme « produit facilitant la co-expérience et les interactions sociales ». La notion de co-expérience renvoie aux expériences sociales au sens d'expériences partagées avec autrui (Forlizzi & Battarbee, 2004). Dans cette perspective, un sociobot est à la fois celui avec qui on expérimente et celui qu'on expérimente avec d'autres. Autrement dit, il se présente comme sujet et comme objet d'expérience. Bartneck et Forlizzi s'inscrivent ainsi, sans le dire explicitement, dans le champ du design d'expérience (aussi appelé UX design ou UXD⁴⁶) qui, comme son nom l'indique, conçoit ses produits en fonction des expériences qu'ils doivent générer ou auxquelles ils doivent participer⁴⁷ (Hassenzahl, 2010). Bien sûr, une UX n'est pas intégrée au produit comme le serait une fonctionnalité, mais résulte de nos interactions avec lui. L'UX se présente ainsi comme propriété émergente de l'interaction (*ibid*) et procède par conséquent de la rencontre entre nos états internes, les différents attributs du produit et le contexte d'interaction (Alenljung *et al*, 2017). Dans cette perspective, l'adoption (tant au sens d'utilisation récurrente que d'attachement) d'une technologie donnée semble en grande partie conditionnée par la positivité de son UX. Une « expérience positive » (Assenzahl, 2010 ; Alenljung *et al*, 2017) se définit de façon générale comme une UX participant à la réalisation des bénéfices attendus de la technologie en question. Plus précisément, elle désigne une expérience où les fonctions pragmatiques et les fonctions hédoniques d'un artefact se réalisent correctement (Hassenzahl & Tractinsky, 2006). Les fonctions pragmatiques, d'une part, désignent l'utilisabilité⁴⁸ d'un produit au sens instrumental classique du terme ; les fonctions hédoniques, d'autre part, renvoient à cet au-delà de l'instrumental qu'est la

⁴⁶ Contractions de l'anglais *user experience design*.

⁴⁷ Contrairement au design industriel dont l'intérêt est principalement tourné vers le produit.

⁴⁸ La norme ISO 9241-11 définit l'utilisabilité comme le « degré selon lequel un système, un produit ou un service peut être utilisé, par des utilisateurs spécifiés, pour atteindre des buts définis avec efficacité, efficacité et satisfaction, dans un contexte d'utilisation spécifié ». <https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:fr> (consultée en ligne le 03/11/2021).

dimension émotionnelle et symbolique dont une technologie est porteuse. Les exigences quant au niveau de fonctionnalité pragmatique et hédonique dont un artefact doit faire preuve dépendent de plusieurs critères comme sa finalité et le public auquel il s'adresse ; d'où la nécessité d'identifier dès l'amont le bon « dosage » pour favoriser les UX positives.

En matière sociobotique, les interactions génératrices d'expériences sont principalement de nature sociale. Les designers doivent donc accorder une attention toute particulière aux éléments qui contribuent à façonner des expériences d'interactions sociales humain-robot (ISHR) de qualité. Sur ce point, Bartneck et Forlizzi identifient les cinq paramètres suivants : 1) la forme du robot, 2) ses modalités de communication, 3) sa conformité aux normes sociales du milieu dans lequel il intervient, 4) ses degrés d'autonomie et 5) d'interactivité.

Forme

Les auteurs incluent dans la notion de forme les questions d'apparence, de matériau et de qualités comportementales du robot. Globalement, celui-ci peut être de forme abstraite, biomorphe ou anthropomorphe. Ce paramètre renvoie à l'une des règles d'or du design sociobotique, à savoir la correspondance entre la forme perçue de l'artefact et ses capacités effectives. Tout manquement à ce principe conduit généralement à d'importantes déceptions et, par conséquent, à un vif désintérêt pour l'artefact.

Modalités de communication

Ce paramètre renvoie au nombre de canaux de communication auquel un sociobot peut renvoyer. Il peut être unimodal ou multimodal selon le choix des concepteurs et correspond, peu ou prou, aux cinq canaux perceptifs humains (vue, audition, toucher, olfaction et goût).

Conformité aux normes sociales

Pour être un « bon interactant », un robot doit aussi être pensé en fonction des normes sociales du milieu auquel il sera intégré. La connaissance et le respect de ces normes, selon des degrés propres au contexte et aux finalités de l'artefact, sont autant de facteurs non techniques qui conditionnent la positivité d'une UX.

Autonomie

Un sociobot doit, toujours selon son contexte d'intervention, afficher une plus ou moins grande autonomie qui, selon Bartneck et Forlizzi, peut se définir comme la capacité technique de l'artefact à agir sans supervision humaine directe.

Interactivité

Enfin, un robot social doit être interactif au sens de capacité à afficher des comportements intentionnels ou, dit autrement, de répondre aux diverses sollicitations humaines au travers des interactions.

Le critère de forme mis à part, les auteurs conçoivent chaque paramètre comme un continuum allant, par exemple, d'une autonomie nulle à une grande autonomie en passant par une autonomie moyenne. Bien qu'elle paraisse s'en rapprocher, cette gradation diffère dans ses objectifs de celle proposée par Breazeal. En effet, Bartneck et Forlizzi ne semblent pas avoir d'autre référentiel téléologique que la « bonne expérience » sociobotique. Or, contrairement au robot sociable, une telle UX n'est pas nécessairement celle d'un robot se rapprochant au maximum de l'humain. Penser le sociobot sous l'angle des expériences d'interactions sociales et non sous l'angle des interactions sociales *tout court* invite à resituer l'artefact dans un contexte qui relativise son interactivité. Par exemple, un NAO destiné à interagir avec des enfants autistes serait probablement vecteur d'une très mauvaise UX si ses concepteurs le dotaient des mêmes comportements d'interaction que ceux prévus pour des enfants sans TSA. Si l'appréhension expérientielle du sociobot ne change pas fondamentalement la liste des critères classiques de socialité, elle les relativise en les inscrivant au sein d'un ensemble complexe de paramètres que Forlizzi nomme « Product Ecology » (Forlizzi *et al*, 2004 ; Forlizzi 2007a, 2007b). Cette logique propre au design d'expérience explique probablement le succès d'un artefact somme toute très peu interactif comme Paro. Elle correspond à la catégorie « *human-centered HRI* » proposée par Dautenhahn dont l'objectif est de concevoir des IHR « acceptables et confortables pour l'humain » (Dautenhahn, 2007a, p.684).

2.1.2 Les approches spécifiques de la sociobotique

Ce portrait généraliste peut être complété par quatre autres concepts dont les contenus plus spécifiques sont utiles pour détailler notre portrait du sociobot et comprendre certaines de ses dimensions fondamentales.

2.1.2.1 L'« objet à comportement » ou l'influence du mouvement

L'« objet à comportement » (*behavioral object*) est un concept proposé par les chercheurs en psychologie cognitive Florent Levillain et Elisabetta Zibetti (2017). Comme son nom l'indique, il désigne un artefact robotique capable de manifester des comportements et, ainsi, de générer une impression de « vitalité » (*aliveness*). On trouve, sur proposition de Dominique Sciamma (2012a, 2012b), un concept équivalent avec le néologisme de « robject »⁴⁹. Les objets à comportement se définissent comme des « objets producteurs de signaux spécifiques qui, pour un cerveau humain et la psychologie naïve qu'il génère, constituent des indicateurs significatifs d'états internes et de disposition à interagir avec le monde extérieur »⁵⁰ (*ibid*, p.11). La production de tels indicateurs demande au design de prendre en compte de nombreux éléments au premier rang desquels on trouve le mouvement. N'importe quel mouvement ne fait toutefois pas l'affaire : celui-ci doit être expressif, exprimer une animation au sens du latin *animare* (« donner, insuffler la vie ») ou de l'anglais « *animacy* »⁵¹. Comme le remarquent Hoffman & Ju (2014), l'animation d'un objet constitue un puissant medium d'interaction et d'expression et ce indépendamment de son apparence. Des facultés telles que la kinésie et la proxémie⁵² sont, par exemple, déterminantes pour communiquer et établir des relations interpersonnelles. En outre, la recherche en psychologie cognitive a depuis longtemps montré que l'humain tend à attribuer des états internes et des intentions à des mouvements provenant de formes abstraites ou d'objets *a priori* éloignés de l'animation

⁴⁹ Contraction des termes de robot et d'objet.

⁵⁰ « Behavioral objects are objects that produce specific cues that, to a human brain and to the naive psychology it generates, constitute meaningful indicators of an internal state and a disposition to interact with the world. »

⁵¹ Le terme anglais d'*animacy* renvoie au mouvement que l'on rattache spontanément à une sentience ou à une conscience. Nous proposons donc de parler d'animation au sens de mouvement qui semble provenir d'une « âme » (*anima*), d'une intériorité dont il serait le révélateur.

⁵² La kinésie désigne la génération et l'interprétation de comportements non-verbaux s'exprimant à travers les mouvements du corps (le fameux « langage corporel »). La proxémie, quant à elle, renvoie à l'interprétation et à la gestion humaines des distances physiques en interaction.

(Heider & Simmel, 1944). Malgré ce rôle décisif du mouvement expressif, Hoffman & Ju remarquent que la robotique interactive au sens large ne lui accorde qu'une importance secondaire et privilégie généralement deux autres approches qu'ils qualifient de pragmatique et de visuelle. Dans l'approche pragmatique, c'est l'efficacité qui préside à la conception des robots en général et de leurs mouvements en particulier. La motion est alors entièrement fonction des objectifs physiques à atteindre (i.e. aspirer la poussière sur le sol dans le cas d'un robot aspirateur) et adossée à des critères de coût et d'énergie. L'approche visuelle, quant à elle, se concentre sur l'apparence de l'artefact. Elle s'avère particulièrement importante dans le cas des robots dont la vocation est d'interagir de façon « expressive » avec l'humain. Leur aspect est alors pensé en fonction du type d'interaction qu'ils doivent prendre en charge, ce qui, en règle générale, inclut des éléments à la fois tangibles et intangibles tels que le regard, le sourire, la gestuelle, la forme du robot (anthropomorphe, zoomorphe, fonctionnelle, abstraite etc.) et sa matière (métal, silicone, PVC, duvet etc.). Hoffman & Ju, quant à eux, proposent, non pas de dévaluer l'importance de ces deux dimensions, mais de réévaluer celle de l'animation dans une perspective centrée sur le mouvement expressif (*expressive movement centric design approach*). Concrètement, il s'agit de faire intervenir le design de l'animation dès l'amont du processus de conception afin de mieux l'articuler aux préoccupations pragmatiques et visuelles. Si l'animation fait partie des indices fondamentaux quant aux possibilités d'action et d'interaction de l'artefact – participant ainsi de ce que l'on appelle ses *affordances* –, il est également, selon les auteurs, un puissant vecteur d'acceptation des robots. Des mouvements fluides, par exemple, favorisent l'attachement des utilisateurs à l'égard d'un robot, contrairement aux gestes saccadés qui renvoient bien souvent à la froideur de la machine, du moins dans une certaine représentation de celle-ci. Revenons maintenant aux objets à comportement. Nous venons de voir que leur cinétique relève du régime phénoménologique particulier qui est celui de l'animation. Autrement dit, l'artefact doit nous apparaître comme un objet qui *se comporte*, pas simplement comme un objet qui bouge. Force est pourtant de constater que, d'ordinaire, nous appliquons généralement la notion de comportement aux entités qui sont « véritablement » animées, c'est-à-dire aux vivants. Dès lors, que faut-il entendre lorsque nous parlons du comportement d'un objet ? Dans le langage courant, en effet, le comportement semble associé à une forme d'intentionnalité voire de volonté. Or, il est vrai que nous parlons aussi du comportement d'une tempête ou des vagues, comme le remarquent Levillain et Zibetti.

« Alors que ces entités sont privées d'intentionnalité, elles possèdent une organisation interne qui subsume l'activité d'une myriade d'individus ou de particules dont le comportement, lorsqu'on l'observe de très près, peut sembler aléatoire ou désordonné. De la même façon, le comportement d'une voiture ou de tout autre artefact technologique renvoie à la description d'une organisation fonctionnelle qui assure la cohésion de ses parties et leur assigne un rôle au sein de la machinerie globale » (Levillain & Zibetti, 2017, p.10).

En ce sens, le comportement pointe vers l'idée d'une transformation qu'un être effectue sur la base des possibilités qu'offre son organisation interne. Cette perspective permet alors de penser le comportement indépendamment des dimensions biologique et intentionnelle habituellement convoquées. Levillain et Zibetti ajoutent à cette proposition une seconde approche où le comportement est pensé à l'aune du rapport sensorimoteur qu'une entité entretient avec son environnement. Il désigne alors

« la relation qu'un organisme entretient avec son environnement par le biais d'une organisation neurale qui convertit l'excitation de récepteurs sensoriels en une activité motrice. Cette activité peut être autoproduite [...] ou procéder d'une réaction à des changements déclenchés dans l'environnement. Dans cette perspective, le comportement désigne donc un ensemble d'activités qui peuvent être spontanées ou causée par des variations externes » (*ibid.*).

Malgré ces racines biologiques, les auteurs élargissent cette définition au non-vivant en remplaçant l'idée d'organisme par celle d'entité possédant une organisation interne. La sociobotique, comme nous l'avons vu, pousse très explicitement à ces efforts de révisions catégorique et conceptuelle. Du point de vue des concepteurs, toutefois, l'objectif n'est pas tant de procéder à une réforme sémantique et/ou ontologique que de reproduire les conditions phénoménologiques du comportement. L'ambition est claire : encourageant les utilisateurs à leur attribuer des qualités de nature psychologique (intentions, émotions, croyances, croyance, personnalité), la comportementalisation des artefacts favorisent nos interactions avec eux, mais aussi notre attachement à leur égard.

Concrètement, le concept d'objet à comportement peut renvoyer 1) à des artefacts existants dont les fonctions pratiques se voient enrichies d'une fonction comportementale ou 2) à des

objets non utilitaires spécialement conçus manifester cette animation. Pour illustrer le premier cas de figure, les auteurs citent notamment l'exemple du « grille-pain impatient » (*impatient toaster*) intervenant dans une étude de Burnleit, Hemmert & Wettach (2009) et celui du lecteur de disquette « *Floppy legs* » conçu par James Chambers (Auger, 2014). Le grille-pain impatient est un toaster à l'apparence des plus classiques, mais qui se met à vibrer fortement à intervalles réguliers. Ces vibrations cessent lorsque l'artefact est utilisé pour griller du pain puis reprennent lorsque celui-ci est prêt. Il faut alors retirer l'aliment de l'appareil pour que les vibrations disparaissent, jusqu'à son prochain « accès d'impatience ». Il est intéressant de constater l'impact de cette fonction comportementale sur le statut que nous attribuons à l'objet en pratique. Les six personnes participant à l'étude se mirent rapidement à parler au grille-pain, à le toucher et à essayer de le calmer verbalement. Comme le notent Levillain et Zibetti, ce comportement vibratoire inattendu et *a priori* sans lien avec la fonction pratique de l'objet pousse à l'analyser en termes psychologiques comme, par exemple, une impatience liée au fait d'avoir faim. L'artefact se montre ainsi capable de susciter des réactions empathiques fondées sur l'interprétation de ses mouvements, augmentant par-là son utilisation pratique d'une relation émotionnelle. En l'espèce, cette comportementalisation de l'objet n'est pas sans lien avec sa fonction pratique puisqu'elle fut notamment pensée pour inciter les étudiants absorbés par leurs ordinateurs à ne pas sauter de repas. Elle relève donc de ce qu'on appelle désormais le *nudge* (Thaler & Sunstein, 2010), c'est-à-dire du « petit coup de coude » qui pousse à agir d'une certaine façon sur le mode de l'influence. Le second exemple, le *Floppy legs*, est un lecteur de disquette qui se lève sur ses quatre « jambes » automatiquement lorsqu'il détecte un liquide à proximité. Cela permet bien sûr d'épargner ses parties électroniques en cas de contact avec le liquide en question. Là aussi, le comportement n'est pas constitutif de la valeur fonctionnelle du lecteur, mais il l'enrichit en permettant à l'artefact de se protéger contre l'accident du contact avec l'eau ou le café.



Le Floppy legs conçu par James Chambers

Le second type d'objets à comportement désigne des objets dépourvus de toute fonction pratique. Levillain et Zibetti citent l'exemple d'une œuvre de l'artiste danois Jeppe Hein nommée *360° Presence*⁵³. Il s'agit d'une boule en acier de 70 cm de diamètre qui se met à rouler lorsqu'un visiteur entre dans la salle où elle est exposée. Ses mouvements suivent une trajectoire totalement erratique et la conduisent régulièrement à heurter les murs de la pièce qui, peu à peu, se chargent d'impacts et de traces de frottements. Faisant écho à la célèbre étude d'Heider et Simmel (1944) intitulée « *An experimental study of apparent behavior* », *360° Presence* s'en démarque toutefois par le caractère très déstabilisant de son comportement. Dans l'expérience d'Heider et Simmel quatre formes géométriques (deux triangles, un cercle et un rectangle) suivent des mouvements aléatoires que les participants interprètent en termes d'intentions et intègrent à des narrations du type « telle forme tente de se mettre à l'abri de telle autre forme qui la poursuit ». Dans le cas de l'œuvre de Jeppe Hein, le spectateur attribue également une forme de subjectivité à la boule sans pour autant parvenir à expliquer son comportement en termes rationnels de causalité et de finalité.



360° Presence

⁵³ Voir le site personnel de Jeppe Hein : https://www.jeppehein.net/proiect_id.php?path=works&id=140.

Ces précisions étant données, comment situer le sociobot par rapport à ces objets à comportements ? Levillain et Zibetti soutiennent que

« [c]ontrairement aux robots sociaux, les objets à comportements ne sont pas spécifiquement conçus pour servir, aider ou coopérer avec les humains. Bien qu'ils puissent parfois mimer les comportements sociaux humain [...], ils ne sont pas conçus pour s'engager avec un utilisateur à travers des aptitudes sociales humaines ou par le biais de caractéristiques telles que la gestuelle, la posture, le langage corporel et facial qui organisent l'interaction sociale » (Levillain & Zibetti, 2017, p.10).

C'est donc la nécessaire finalité socio-interactive des robots sociaux qui les distingue des objets à comportements. Chez le sociobot, fonction, comportement et interaction sociale se confondent puisque celui-ci se comporte pour interagir socialement et qu'il interagit socialement pour réaliser sa fonction.

2.1.2.2 L'« artefact relationnel » : émotions et empathie artificielles

Le concept d'artefact relationnel (*relational artifacts*) est utilisé par Sherry Turkle dans ses nombreuses contributions au sujet des IHR. La notion renvoie

« aussi bien aux robots complexes de recherche comme Kismet ou Cog⁵⁴ qu'à un éventail plus large d'objets se retrouvant sur le marché grand public comme les créatures virtuelles, les animaux robotiques de compagnie et les poupées humanoïdes. Avec des niveaux de sophistication différents, ces objets donnent l'impression de vouloir qu'on s'occupe d'eux, qu'on satisfasse leurs "besoins" et semblent être heureux lorsqu'on prend soin d'eux correctement » (Turkle, Breazeal *et al*, 2006, p.314).

En d'autres termes, il s'agit d'artefacts dont l'apparence de sensibilité incite à prendre soin d'eux. En amont de sa dimension « sollicitudinaire » (Grimaud, 2012) sur laquelle nous reviendrons, le concept d'artefact relationnel renvoie, plus fondamentalement, à l'importance des émotions et de l'empathie en robotique sociale. Celles-ci jouent en effet un

⁵⁴ Cog est le nom d'un projet de recherche initié par le *Humanoid Robotics Group* (HRG) du MIT dans les années 1990.

rôle proprement fonctionnel au sein des ISHR (Breazeal & Brooks, 2004). Selon Fong *et al* (2002) ce rôle se décompose *a minima* en trois grandes fonctions :

1. Les émotions facilitent la crédibilité des ISHR ;
2. Elles livrent des indices à l'interlocuteur quant aux « états internes », aux buts et aux « intentions » du robot ;
3. Elles peuvent agir comme un mécanisme de contrôle en stimulant certains comportements et en affichant comment le robot est affecté par tel ou tel facteur de son environnement et s'y adapte (ou non) avec le temps.

Cynthia Breazeal et Rodney Brooks s'inscrivent eux aussi dans cette perspective qu'ils qualifient de « pragmatique ». Leur intérêt pour les « mécanismes et aptitudes d'inspiration émotionnelle »⁵⁵ (Brezeal & Brooks, 2004) est lié à deux enjeux majeurs de conception : 1) l'amélioration du fonctionnement des robots dans un environnement humain et 2) le perfectionnement des capacités de coopération effectives de l'artefact. Et les auteurs d'insister :

« les éclairages que ces mécanismes basés sur les émotions et les affects fournissent aux concepteurs ne devraient pas être réduits à une simple façon de créer des robots "heureux" ou divertissants. Penser ainsi revient à manquer un point extrêmement important : comme pour les êtres vivants, ces mécanismes d'inspiration émotionnelle modulent le système cognitif du robot afin qu'il *fonctionne mieux* dans un environnement complexe et imprévisible, c'est-à-dire qu'il prenne de *meilleures* décisions, qu'il apprenne *plus* efficacement, qu'il interagisse *plus* convenablement avec les autres que s'il n'avait qu'un système cognitif » (*ibid.*, p. 274).

On retrouve ici l'ambition décrite précédemment au sujet du robot sociable : créer des robots à l'image du vivant et, plus précisément, du vivant humain. L'avertissement des deux auteurs sur le risque d'une appréhension superficielle du rôle des affects en sociobotique trace une ligne de partage entre deux types de recherches que Dumouchel et Damiano (2016) nomment robotique « externe » et robotique « interne » des émotions.

⁵⁵ « *emotion-inspired mechanisms and abilities* »

La robotique « externe »

La robotique externe se consacre à la phénoménalité « extérieure » des émotions, c'est-à-dire à la façon dont elle se manifeste pour autrui. Elle renvoie ainsi à l'expressivité de la gestuelle, des mouvements, des postures ou encore des formes de corps et expressions du visage, mais ne correspond à aucun état « interne » de l'artefact.

« Le but est de doter les robots destinés à être utilisés dans les services de formes d'expressivité qui, en colorant affectivement leurs actions, facilitent leur acceptation et augmentent leur efficacité. Au cœur de cette démarche se trouve l'idée que l'empathie constitue le composant essentiel à la mise en place d'interactions affectives convaincantes, positives et durables entre les agents artificiels et les êtres humains » (*ibid.*, p.124).

Ne faisant pas appel à une mécanique interne sous-jacente à l'expression des émotions, la robotique externe concentre principalement ses efforts de recherche et de conception sur le phénomène d'anthropomorphisme. Compris comme la tendance humaine spontanée à attribuer des états mentaux (intentions, désirs, émotions, croyances etc.) aux non-vivants, ce phénomène est non-seulement nécessaire à la conception d'agents sociaux artificiels crédibles, mais permet également de générer des ISHR avec des artefacts finalement peu complexes techniquement (comme Paro ou le grille-pain affamé par exemple). Dans la classification proposée par Breazeal, ces sociobots sont toutefois rivaux aux couches inférieures de socialité car, ne disposant pas d'« objectifs sociaux internes »⁵⁶ (Breazeal, 2003), ils échouent bien souvent à entretenir leur intérêt aux yeux des humains.

La robotique « interne »

Cette génération de finalités propres se trouve justement au cœur des recherches en robotique « interne » des émotions qui, de façon générale, s'intéresse au rôle des affects dans la construction des comportements. Dumouchel et Damiano lui attribuent deux ambitions : un objectif de science fondamentale, d'une part, consistant à étudier les émotions pour mieux comprendre leur lien avec les comportements humains et, plus largement, animaux ; un

⁵⁶ « *internal social aims* »

objectif de science appliquée, d'autre part, consacrant ces nouvelles connaissances à la conception de robots plus performants en termes d'autonomie et d'adaptation sociales.

« L'implémentation de systèmes de régulation émotionnelle au sein d'agents robotiques vise plus précisément à les doter de mécanismes qui leur permettent de déterminer par eux-mêmes un ordre de priorité entre diverses options comportementales. Ces robots seront capables d'activer différentes stratégies d'apprentissage d'adaptation au moment opportun, ce qui augmentera leurs capacités d'interaction autonome avec l'environnement et avec les autres agents, robotiques ou humains. » (Dumouchel & Damiano, 2016, p.132-133)

Selon les deux philosophes, la robotique externe cherche à concevoir des sociobots *réellement* capables d'être affectés en ceci que leurs émotions seraient liées à un mécanisme de régulation interne⁵⁷.

« À la différence des efforts de la robotique "externe" des émotions pour éveiller et exploiter notre anthropomorphisme spontané, ces recherches visent à créer chez les robots des émotions et de l'empathie. La robotique "interne" veut fabriquer des agents artificiels qui seront véritablement empathiques et dont les émotions seront authentiques et réelles » (*ibid.*, p.137).

Cette distinction entre les aspects internes et externes des émotions est toutefois problématique car elle se fonde sur l'idée qu'une émotion est un phénomène essentiellement privé dont la vérité dépend de sa genèse « interne ».

Le bouclage affectif

« Pourquoi les aspects internes des émotions seraient-ils "vrais" et les aspects externes "faux" ? » (*ibid.*, p.140), s'interrogent alors les auteurs. Cette question confronte notre héritage intellectuel classique où l'émotion est interprétée comme un événement interne dont l'existence sur la scène extra-individuelle dépend justement de l'individu à qui elle se manifeste d'abord de façon privée et exclusive. La manifestation externe de l'émotion est alors jugée contingente et, ainsi, non nécessaire à son authenticité. Les approches internes et

⁵⁷ Sans surprise, cette question de l'authenticité des émotions artificielles et, plus largement, des comportements robotiques constitue, comme nous le verrons, l'un des principaux points abordés par l'éthique des sociobots.

externes des émotions convergent toutefois vers l'idée que « les phénomènes affectifs se déroulent dans un espace qui va depuis l'*intra-* jusqu'à l'*extra-*individuel » (*ibid.*, p.141). Contrairement aux conclusions de la pensée classique – reprises dans une assez large mesure par les sciences cognitives – Dumouchel et Damiano affirment qu'une telle convergence devrait inciter à penser les émotions au sein d'une ontologie *relationnelle*. Bien qu'à des niveaux de complexité différents, la robotique externe comme la robotique interne reposent sur un phénomène de « bouclage affectif » au sens où « lorsque l'interlocuteur du robot observe ou perçoit l'expression affective de l'agent artificiel il est immédiatement inclus au sein du processus affectif et participe à la dynamique des émotions » (*ibid.*, p.143). Cette idée de boucle affective (*affective loop*) s'offre donc comme alternative à la fois épistémologique et technique à une conception des émotions fondée sur la dichotomie entre authenticité intérieure et facticité extérieure.

« La frontière entre les aspects internes et externes des émotions et de l'empathie, laquelle est censée coïncider avec la limite de l'espace *intra-*individuel, est [...] de fait continuellement transgressée tant par la robotique "interne" que par la robotique "externe", qui s'accordent l'une et l'autre, malgré leurs perspectives très différentes, pour situer les processus affectifs non pas dans l'agent individuel, mais dans la relation entre l'agent et l'environnement, et plus précisément dans les rapports *inter*individuels entre agents humains et artificiels » (*ibid.*).

L'objectif d'une sociobotique fondée sur le bouclage affectif est donc de concevoir des artefacts capables de participer activement à la dynamique relationnelle des émotions en jouant sur un double mécanisme de modélisation interne et de manifestation externe des affects. Les auteurs remarquent toutefois que cette approche relationnelle reste dominée par un parti pris internaliste qui la ramène vers le paradigme classique. Malgré l'identification d'une dynamique relationnelle dans la genèse des émotions, le bouclage affectif continue de les concevoir comme des phénomènes essentiellement internes donc discrets. Autrement dit, l'interactionnisme du bouclage affectif est un interactionnisme où les individus préexistent à leurs interactions. En dernière instance, la dichotomie entre l'interne essentiel et l'externe contingent continue d'y jouer un rôle fondamental en structurant notre rapport aux émotions et, plus largement, à la socialité. Dans la partie 2.2 de ce chapitre, nous verrons que les

critiques de ces approches font signe vers une théorie des affects comme forme de « socialité primordiale » (*ibid.*, p.150) où la relation prime sur l'individualité de ses pôles.

Nous avons choisi d'introduire ce point consacré aux affects par la notion d'artefact relationnel car, contrairement aux formules de « robot émotionnel » ou de « robots empathiques » souvent utilisées par les journalistes ou les communicants, elle est dotée d'un contenu conceptuel précis. Son premier mérite est de mettre en avant la dimension fondamentale de la relation en sociobotique. Comme l'explique Sherry Turkle, les robots sociaux sont « explicitement conçus pour engager son utilisateur dans une relation » (2007, p.502), ce qui n'a rien d'anodin contrairement à ce qu'on pourrait croire de prime abord.

« Par le passé, le pouvoir des objets de "jouer à la poupée" ou de "jouer au cowboy" avec un enfant était lié aux significations que celui-ci projetait sur eux. Il s'agissait alors d'"objets transitionnels" stables. [...] La poupée ou l'ours en peluche manifestaient une présence immuable et passive. Mais les artefacts relationnels d'aujourd'hui se montrent résolument plus actifs. Avec eux, l'impression qu'ils désirent être câlinés, habillés ou bercés ne vient pas uniquement de l'imaginaire et des désirs d'un enfant projetés sur un jouet inerte, mais aussi de choses telles qu'une poupée électronique pleurant de façon inconsolable ou disant "Câline moi !" ou "C'est l'heure de m'habiller pour aller à l'école !" » (Turkle, 2006).

Turkle décrit ainsi les artefacts relationnels comme des « Rorschach » (*ibid.*) robotiques qui stimulent activement nos projections mentales dans le sens général de la sollicitude. Le concept d'artefact relationnel est ainsi doté d'une précieuse valeur descriptive permettant de comprendre l'un des aspects constitutifs des robots sociaux, à savoir l'investissement émotionnel qu'ils suscitent. D'un point de vue fonctionnel, cet investissement semble représenter l'une des conditions sine qua non de l'interaction sociale. Or, comme nous le verrons dans le chapitre consacré à l'éthique des sociobots, le concept d'artefact relationnel n'est pas uniquement un outil intellectuel de description, mais aussi un moyen de critiquer moralement son objet.

2.1.2.3 L'autonomie sociale du « substitut » : autorité, présence et coordination

Le troisième concept dont nous souhaitons rendre compte est celui de *substitut* proposé par Dumouchel et Damiano. Précisons d'emblée que cette notion ne s'applique pas à tous les sociobots, mais seulement à certains d'entre eux. Malgré cela, elle se forge à partir d'éléments qui, à la suite du mouvement et des émotions, apparaissent comme des caractéristiques essentielles du robot social. Selon Dumouchel et Damiano, le substitut renvoie à un type bien particulier de sociobot dont la vocation est de se substituer à un humain « *sans lui prendre sa place, c'est-à-dire sans lui enlever sa fonction* » (Dumouchel & Damiano, 2016, p.39). Contrairement à un remplaçant définitif, le substitut est destiné à prolonger la présence et la fonction de la personne substituée afin qu'elle puisse l'exercer ailleurs au même moment. Le concept est donc lié à une certaine forme d'ubiquité du rôle en question. Le caractère momentané et parcellaire du remplacement fait partie du concept. Autrement dit, un substitut ne fait pas tout ce que fait la personne substituée et il ne le fait pas en permanence.

« Le projet de créer des suppléants artificiels, des robots susceptibles de se substituer à nous dans certaines tâches, mais sans pour autant qu'ils ne prennent notre place est central en robotique sociale. Non pas au sens où la majorité des roboticiens qui œuvrent dans ce domaine s'y reconnaîtraient, mais au sens où il est souvent implicite dans ce qu'ils font. [...] L'aide-soignant, la nourrice ou le compagnon robotiques ne rendent pas obsolète ou inutile le rapport aux infirmiers, aux parents ou aux amis et ils n'ont pas pour but de les remplacer, mais de se substituer à eux par moments et dans certaines circonstances » (*ibid.*, p.40-41).

Cette notion de substitution marque une différence notable avec la plupart des robots industriels ou de service qui, eux, sont destinés à remplacer l'humain de façon permanente. En outre, la substitution n'intervient jamais ou presque jamais sur des fonctions d'ordre techno-instrumental (e.g. aspirer la poussière, désinfecter au moyen de rayons UV, assembler des pièces de carrosserie au sein d'une chaîne de montage), mais sur des fonctions que le langage courant qualifie volontiers de « sociales » comme, par exemple, accueillir le public à l'entrée d'un magasin ou tenir compagnie à une personne âgée. La dimension sociale de ces fonctions vient non seulement de ce qu'elles se réalisent à travers des interactions avec l'humain, mais de ce que ces interactions impliquent des éléments qui ne sont pas de nature

technique. C'est ce que Dumouchel et Damiano nomment l'« autonomie sociale » des substituts comprise comme « dimension particulière d'action » (*ibid.*, p.62) sans lien avec une quelconque thèse métaphysique quant à la liberté des sociobots. Trois caractéristiques fondamentales composent cette autonomie, à savoir la capacité de coordination et d'adaptation, la présence et l'autorité.

La capacité de coordination et d'adaptation

Selon les deux philosophes, la socialité d'un agent artificiel comme un substitut dépend, certes, de ses capacités d'interaction et du rôle qu'elles lui permettent de jouer au sein de la société, mais aussi de son aptitude à s'affranchir de sa fonction et des scénarios qu'il est censé prendre en charge. Bien qu'un substitut soit *d'abord* un sociobot conçu pour jouer tel ou tel rôle, il doit pouvoir s'adapter, même de façon minimale, aux écarts et autres imprévus qui caractérisent les interactions sociales.

« Les substituts artificiels sont différents des autres objets techniques dans la mesure où tout agent social doit nécessairement participer de cette capacité d'adaptation et de coordination, même de façon minime. Et cette capacité sociale exige de pouvoir détacher le robot de l'accomplissement d'une tâche ou d'une fonction unique » (*ibid.*, p.46).

Ainsi, un robot d'aide hospitalière devra, par exemple, pouvoir effectuer des tâches d'accueil et de renseignement si, au détour d'un passage en salle d'attente, il est interpellé par un patient en demande d'information.

La présence

Contrairement à de nombreux objets du quotidien dont la présence s'éclipse au profit de la fonction qu'ils réalisent, les substituts et, plus largement, les sociobots *fonctionnent* en faisant « acte de présence » (p.46). Ce point est d'une importance telle que la présence va jusqu'à être considérée comme une fonction à part entière de certains robots sociaux dits de compagnie comme Paro, Lovot, Aibo, RoBoHon ou encore Charlie. Bien sûr, cette présence

s'exprime de différentes façons selon les artefacts, mais elle a pour principe générique de donner « l'impression d'"être avec un autre" »⁵⁸ (Biocca *et al*, 2003, p.456).

« La caractéristique définitionnelle des "robots sociaux" est la capacité d'être perçus par ceux qui interagissent avec eux comme étant *présents* à la manière d'une personne, ce qu'on peut nommer la *présence sociale*. Les robots sociaux doivent donner à leurs partenaires humains le "sentiment", l'impression "suffisamment convaincante et durable... d'être en présence de *quelqu'un*". Ils doivent provoquer chez l'interlocuteur l'impression d'être face à un "autre". Ce qui est en jeu n'est pas une simple projection de la part du partenaire humain, mais la production, la mise en place d'un sentiment qui constitue un élément central des relations sociales de base, les rencontres face à face, dont, en dernière analyse, toutes les autres relations sociales sont dérivées » (Dumouchel & Damiano, 2016, p.117-118).

La présence des sociobots ne se manifeste pas simplement du point de vue matériel ou esthétique comme c'est le cas d'un objet de décoration, elle se manifeste comme une *altérité*. Ce point, nous l'avons dit, est au cœur de notre recherche et sera pleinement développé dans la seconde partie de ce manuscrit.

L'autorité

La dernière grande composante de l'autonomie sociale dont parlent Dumouchel et Damiano est celle de l'autorité. Contrairement aux outils comme le mètre qui font autorité, les substituts *ont* une autorité émanant de leur fonction et/ou de la personne au nom de qui ils l'exercent. Qu'il s'agisse d'un humain ou d'un robot, l'autorité consiste dans le fait d'être reconnu comme détenteur légitime d'un pouvoir d'obéissance sur autrui. « Même si la menace de la force constitue très souvent l'horizon ultime des rapports d'autorité, l'autorité existe dans l'exacte mesure où il n'est pas nécessaire d'y recourir » (*ibid.*, p.52). Autrement dit, on reconnaît l'autorité véritable en ceci qu'elle n'a pas besoin de se muer en rapport de force pour produire ses effets. « Un agent social robotique doit [donc] être un interlocuteur qui réussit au moins partiellement à s'affirmer et à se faire respecter. » Sans une autorité minimale, le substitut ne peut pas fonctionner socialement, c'est-à-dire remplir la fonction

⁵⁸ « *the sense of "being with another"* »

pour laquelle il a été conçu. Par exemple, un robot « aide-soignant » chargé de s'assurer que les patients d'un service hospitalier ont bien pris leurs médicaments sera entravé dans l'accomplissement de sa fonction si les personnes soignées jugent que, pour telle ou telle raison, un robot n'est pas légitime pour jouer un tel rôle. De ce point de vue, il y a dans le fonctionnement du robot social, c'est-à-dire dans l'accomplissement de sa fonction, quelque chose qui dépasse la seule opération technique. La délégation effective d'une fonction sociale n'est que partiellement d'ordre technique dans la mesure où un sociobot peut parfaitement fonctionner techniquement (i.e. être capable d'interagir sans défaillance technique au sein du scénario prévu), mais dysfonctionner sur le plan social en étant rejeté, ignoré ou abandonné après quelques interactions. L'autorité du substitut dont parlent Dumouchel et Damiano peut ainsi être replacée dans le cadre plus général de son acceptation sociale en ceci que l'autorité d'un robot ne peut être reconnue, donc être effective, que si ce robot est préalablement accepté. Ainsi, l'autorité ne dépend pas seulement de l'artefact et de ses caractéristiques techniques⁵⁹, mais aussi des relations qui l'unissent à son interlocuteur à travers les perceptions et les expériences qu'il en a.

Décrite en ces termes, l'autonomie sociale répond aux exigences de la vocation substitutive de certains sociobots. Cette autonomie peut toutefois faire l'objet d'une extension à l'ensemble des robots sociaux pourvu qu'on n'exige pas qu'un sociobot puisse s'arracher, même momentanément, à sa fonction (à moins d'être téléopérés, la plupart des robots n'en sont pas vraiment capables) et que la question précise de l'autorité soit élargie à celle de la *reconnaissance* de leur statut d'agents sociaux.

2.1.2.4 Le « robot existentiel » comme vivant post-biologique

Dans un livre intitulé *Machines insurrectionnelles. Une théorie post-biologique du vivant*, Dominique Lestel propose la notion de *robot existentiel* pour désigner les robots « qui obligent à redéfinir ce que signifie être vivant et ce que signifie être humain » (Lestel, 2021, p.33). Il substitue ce concept à celui de « robot social » dont il dit expressément qu'il

⁵⁹ Une autorité purement statutaire, sans lien avec ce que fait concrètement son ou sa détentrice, ne perdure jamais bien longtemps. Voilà pourquoi la question de ce que peut faire techniquement le robot et comment il le fait intervient dans l'équation de son autorité éventuelle, sans toutefois la résumer.

« est inadéquat parce qu'il laisse supposer que les agents concernés préexistent à des relations qu'on peut qualifier de "sociales" alors que les "relations" contribuent pleinement à mettre en place les agents eux-mêmes » (*ibid.*, p.309).

En partant de ce qu'il qualifie de « dimension phénoménologique de ces technologies » (*ibid.*, p.19), le philosophe construit une approche existentielle du vivant dont la question centrale consiste à interroger comment on le devient au-delà de la condition biologique. Lestel considère en effet que nous autres, occidentaux rationalistes, avons « tellement asséché la notion de vivant » (*ibid.*, p.41) que nous sommes désormais incapables de voir quelles peuvent être ses significations ontologiques et existentielles au-delà de ses aspects bio-rationalistes. Penser le vivant du point de vue existentiel consiste donc, comme l'indique le sous-titre du livre, à en élaborer une « théorie post-biologique ».

« Contrairement à la notion de "vie" des biologistes qui serait universelle, bien déterminée et immuable à travers le temps [...], la notion d'"agent vivant" est une catégorie flottante aux potentialités multiples qui renvoie à un phénomène en constante transformation qui émerge dans des pratiques existentielles, sociales et culturelles, et qui sont ancrées dans une concrétude engageante » (*ibid.*, p.257).

L'approche de Lestel se veut résolument perspectiviste. On peut, selon lui, concevoir le vivant selon trois perspectives complémentaires respectivement attachées aux points de vue de la première, de la deuxième personne et de la troisième personne.

Du point de vue de la troisième personne, le vivant peut être décrit au moyen de caractéristiques objectives venant tracer la frontière entre ce qui relève du vivant et ce qui en est exclu. Ce point de vue est celui de la biologie – « approche réaliste, objectiviste, causaliste et essentialiste » (*ibid.*, p.106) - qui s'intéresse aux « êtres vivants phylogénétiques ». Pour Lestel, cette dimension biologique ne représente qu'une partie du phénomène plus global et composite qu'est le vivant :

« L'idée qu'il existerait une catégorie unique qui contiendrait tout ce qui est vivant et qu'une telle catégorie pourrait être objectivée, c'est-à-dire qu'elle pourrait être définie sans aucune référence à un être vivant concret qui le caractériserait, apparaît très problématique compte tenu de la multiplicité des occurrences à travers lesquelles on est prêt à reconnaître le vivant comme vivant » (*ibid.*).

Cette remarque permet d'introduire les deux autres points de vue. Penser le vivant à la première personne consiste à reconnaître que, lorsque nous identifions une propriété chez un agent quelconque, nous faisons activement partie du processus d'identification. Autrement dit, l'attribution du caractère « vivant » à une entité (biologique ou non) constitue une participation active à l'élaboration de ce qu'est le vivant. Selon Lestel, on peut alors légitimement affirmer que le vivant relève en partie de ce qui est « vivant *pour moi* ».

« L'avantage d'une telle posture est que la nature de l'agent à évaluer (qu'il soit naturel ou artificiel ne change pas grand-chose à l'histoire) n'a finalement plus aucune importance. Dans cette perspective, l'ensemble des agents vivants devient l'ensemble de ceux qui sont des agents vivants pour au moins un autre agent vivant » (*ibid.*, p.107).

Le philosophe complète immédiatement cette très large perspective – « est vivant ce que je peux considérer être vivant » – par celle de la deuxième personne :

« je te considère comme vivant si je peux t'*éprouver* comme vivant dans la confrontation que j'ai avec toi ; si je peux t'attribuer une volonté, un désir ou tout simplement un point de vue qui se confronte au mien, et je t'éprouve d'autant plus comme vivant si ton point de vue porte sur moi » (*ibid.*, p.108).

En quelques mots : est vivant ce à quoi je peux attribuer un point de vue. En toute rigueur, cette perspective de la deuxième personne émane de la rencontre entre une première et une seconde personne, c'est-à-dire entre *mon* point de vue et celui que j'attribue à l'entité en question. Voilà pourquoi Lestel conditionne l'intensité du phénomène au sentiment que le « je » a lorsqu'il est pris pour objet par le « tu ».

« Nous nous voyons à travers les yeux d'autrui, c'est le fondement de l'identité du point de vue de la deuxième personne : j'existe dans le regard de l'autre qui lui-même existe dans mon regard — autrement dit, nous existons dans nos regards croisés, ou plus exactement *dans l'écologie de nos regards*. [...] Dans le point de vue à la deuxième personne, le sujet se constitue à travers ses frictions avec des interlocuteurs qui le considèrent lui-même comme un interlocuteur » (*ibid.*, p.109).

Outre cette dimension intentionnelle qui institue le vivant comme phénomène émergent des relations intersubjectives, Lestel ajoute que le devenir-vivant des artefacts est également lié à notre *désir* d'un tel devenir.

« Être vivant n'est donc plus pensé comme une caractéristique qui peut être cernée à travers un système de causalités, ni une projection fantasmatique plus ou moins réussie, ni une relation qui serait de l'ordre du contrat, mais la mise en place d'un dispositif écologique particulier qui se tisse à travers l'entrecroisement de perspectives multiples et hétérogènes à travers des actions, des attentes, des désirs, la mobilisation de ressources *ad hoc* et des dispositifs institutionnels qui structurent le tout. Il ne suffit pas de vouloir fabriquer des machines vivantes pour y parvenir, mais le désir de vouloir le faire est une condition nécessaire. [...] Un artefact devient vivant à l'intérieur d'un dispositif complexe (technique, social, sémiotique, psychologique) qui *me* le rend vivant, à condition que j'en éprouve le désir – un désir qui est d'ailleurs loin d'être un donné élémentaire mais qui est plutôt déjà lui-même une construction élaborée » (*ibid.*, p.114-116).

On l'aura compris, le propos de Lestel ne consiste pas à dire que les sociobots sont vivants au sens biologique du terme, mais qu'ils peuvent devenir vivant *pour nous* selon des agencements plus ou moins complexes de relations entre l'artefact, l'humain et son milieu. Il qualifie les robots existentiels de « machines insurrectionnelles » non pas parce qu'ils se révolteraient comme dans l'œuvre de Čapek, mais parce qu'ils *provoquent* nos certitudes, nos façons de penser et les concepts qui les accompagnent. Il s'agit alors de prendre au sérieux le « faire comme si » auquel les sociobots nous invitent. Face au « vivant artificiel » comme le nomme Lestel, l'obstacle épistémologique n'est pas à chercher dans le concept de vivant, mais dans la persistance d'une pensée du vivant *strictement* biologique. Ainsi, existentialiser le vivant et les sociobots implique notamment de réinterroger la façon dont on en parle, non pas pour changer de vocabulaire, mais plutôt pour l'enrichir de nouvelles significations. Voilà pourquoi l'auteur soutient que les sociobots sont des « *artefacts générateurs de sens* » (*ibid.*, p.120). Il s'interroge par exemple sur la pertinence du concept de mort appliqué au chien robotique de Sony, Aibo. Lorsque celui-ci cesse de fonctionner, faut-il dire qu'il est mort ou qu'il est cassé ? Une différence notable entre un chien biologique et un chien mécanique est que le premier « meurt pour lui-même alors que l'Aibo ne meurt que pour celui qui lui était attaché » (*ibid.*, p.43). De toute évidence, il y a là une confrontation de perspective, mais cette

confrontation n'interdit pas *a priori* l'utilisation du concept de mort. Dans un registre similaire, doit-on considérer que l'on « tue » un Aibo lorsqu'on le détruit ou qu'on le « casse » ? Pour Lestel, ces questions interrogent en profondeur les liens que nous tissons entre le statut ontologique d'une entité et son statut moral. Notre matrice intellectuelle occidentale nous pousse *théoriquement* à considérer que le second est totalement fonction du premier, mais en pratique la valeur morale que nous attribuons à une entité quelle qu'elle soit n'est que partiellement liée au statut ontologique qu'elle est censée avoir.

« Il faut repenser ce que signifie "tuer" et en complexifier l'usage, en conservant ce terme non seulement pour les animaux, les plantes, les végétaux, les champignons comme nous le faisons déjà, mais aussi pour certains artefacts avec lesquels nous avons développé des liens existentiels profonds. La difficulté réside précisément dans la façon dont nos sentiments et nos actions s'entremêlent. Nous n'avons pas le sentiment de tuer quand nous cueillons un champignon, alors que c'est pourtant ce que nous faisons, et nous ne tuons sans doute pas un humain dans le coma quand nous débranchons le dispositif technique complexe qui le maintient en vie de façon artificielle. La question du vivant devient en fin de compte un agrégat du vivant *pour moi* et du *vivant pour nous* » (*ibid.*, p.44).

Lestel n'affirme pas que n'importe quelle entité peut être vivante, mais qu'il est impossible de déterminer *a priori* celles qui le sont ou peuvent le devenir car cela dépend des relations que nous tissons avec elles. En outre, il cherche à montrer que le supplément de valeur que nous accordons ordinairement aux vivants phylogénétiques n'est pas lié à leur supériorité ontologique, mais à la plus grande intensité relationnelle que *nous* entretenons avec eux – cette intensité relationnelle pouvant varier selon les individus et les groupes auxquels ils appartiennent. Cette réinterprétation du vivant à l'aune d'une théorie post-biologique reste donc, selon lui, l'occasion de revaloriser nos liens avec le vivant phylogénétique.

Au terme de ce portrait-robot du sociobot, nous disposons d'un certain nombre d'éléments permettant à la fois de l'identifier et de le différencier au sein du règne artefactuel. Premièrement, on peut affirmer que le robot social est indéniablement un robot au sens

donné par Sciamma et Bryson, c'est-à-dire un artefact témoignant d'un rapport *a minima* sensoriactif au monde. Deuxièmement, ce rapport semble entièrement orienté vers la prise en charge d'interactions sociales avec l'humain à travers lesquelles il va réaliser ses fonctions (qu'elles soient intrinsèques ou extrinsèques à l'interaction). Le sociobot doit donc, troisièmement, nous apparaître comme un agent social crédible et acceptable, c'est-à-dire comme un agent avec lequel nous pouvons et souhaitons (ou, au moins, consentons à) interagir socialement. L'ensemble des caractéristiques que nous avons passées en revue contribue à façonner cette crédibilité et cette acceptation sociales. Les approches généralistes tentent d'appréhender globalement le phénomène sociobotique et énumèrent un ensemble de caractéristiques expliquant la socialité des robots (interactivité, autonomie, communication, apprentissage, anthropomorphisme etc.) tandis que les approches spécifiques s'intéressent à des dimensions particulières sur la base desquelles elles forgent leurs concepts (mouvement/intentionnalité, émotion/sollicitude, autonomie sociale, provocation du vivant/génération de sens). Nous remarquons toutefois qu'aucune de ces perspectives ne s'attardent sur la notion d'« autre » ou d'altérité. Dumouchel et Damiano le disent pourtant très clairement : la présence du sociobot, entendue comme sentiment d'être en compagnie d'un *autre*, est une « caractéristique définitionnelle "des robots sociaux" » (Dumouchel & Damiano, 2016, p.117). Dès lors, quelle est cette altérité qui définit le sociobot sans qu'on lui accorde plus d'attention ? S'agit-il d'une notion trop large ou trop abstraite pour prétendre rendre compte conceptuellement d'une entité aussi surprenante ? Ou est-elle simplement vidée de tout contenu conceptuel propre et considérée comme synonyme d'agent social ou d'agent vivant ? L'ambition de cette recherche, nous l'avons dit, est précisément de montrer que le concept d'altérité permet d'enrichir la pensée philosophique du sociobot en conjoignant deux champs qui ne sont pas suffisamment articulés : la réflexion sur le robot « social » d'une part et celle sur le robot « moral » de l'autre. En d'autres termes, nous cherchons à montrer que socialité et moralité robotique sont deux questions indissociables et que le concept d'altérité permet d'explicitier cette indissociabilité. Pour l'heure, mettons à profit les premières clarifications que nous venons de fournir pour nous tourner vers notre grande question directrice et examiner, à l'aune de la littérature existante, ce que peut signifier *être social* pour un robot.

2.2 Le social comme « propriété » de l'artefact

Dans son introduction à *Changer de société, refaire de la sociologie*, Bruno Latour remarque que l'emploi de l'adjectif « social » devient problématique dès lors qu'il « se met à désigner un type de matériau, comme si le mot était comparable à des adjectifs tels que "métallique", "biologique", "économique", "mental", "organisationnel" ou "linguistique" » (Latour, 2007, p.7). Les robots sociaux nous confrontent à une difficulté similaire à ceci près que leur socialité ne semble pas relever du matériau, mais plutôt de la propriété. Un robot serait donc social au même titre qu'il serait anthropoïde ou zooïde, bipède ou roulant, de couleur noire, blanche ou grise etc. Or, si l'on comprend aisément en quoi une forme ou une couleur peuvent avoir le statut de propriété, c'est-à-dire de caractéristique, d'attribut définissant l'identité d'un objet particulier, la chose est nettement moins claire en ce qui concerne le social. Dès lors, que comprendre de cette interprétation attributive du social ? L'analyse de la littérature permet de dégager au moins trois grandes réponses à cette question, respectivement centrées sur l'interaction, la fonction et l'interactance ; mais avant d'en rendre compte, un bref détour conceptuel sur la notion de propriété s'impose.

2.2.1 Quatre grands types de propriété

Si la propriété se laisse généralement saisir comme attribut participant de l'identité d'un objet particulier, il est utile d'affiner cette définition en précisant la nature de l'attribut en question. Dans ses travaux sur les cellules souches, la biologiste et philosophe Lucie Laplane (2017) propose une typologie dans laquelle la « propriété souche » peut s'exprimer selon quatre modalités différentes :

- La *propriété catégorique* : « Propriété dont dispose une entité du fait de ses propriétés intrinsèques et non en vertu de ses interactions avec le monde »
- La *propriété dispositionnelle* : « Propriété intrinsèque dont l'expression dépend de facteurs extrinsèques »
- La *propriété relationnelle* : « Propriété qui émerge d'une relation entre deux entités »
- La *propriété systémique* : « Propriété déterminée par le système »

Nous proposons d'examiner les différents sens que la littérature donne à la socialité robotique à l'aune de cette typologie. Dans cette optique, nous consacrons respectivement les trois prochains points à l'analyse du social comme propriété dispositionnelle, systémique et relationnelle. La propriété catégorique, quant à elle, ne semble pas pertinente en la matière car, parler de socialité catégorique reviendrait à soutenir l'idée passablement douteuse selon laquelle on pourrait être « social » tout seul.

2.2.2 Le social comme propriété dispositionnelle : l'interaction sociale

2.2.2.1 Le robot comme sujet de socialité

La première réponse appréhende la socialité comme une propriété dispositionnelle liée aux capacités d'interactions du robot. Autrement dit, elle affirme qu'un robot est social en vertu de ses aptitudes physique et cognitive à l'interaction avec l'humain et/ou avec d'autres robots (Duffy *et al*, 1999 ; Breazeal, 2002, 2003 ; Fong *et al* 2003). C'est la perspective techno-centrée proposée par les approches généralistes de Duffy *et al* (1999), Breazeal (2002, 2003), Fong *et al* (2003) et, dans une moindre mesure peut-être, par Bartneck et Forlizzi (2004). Partant, le social est conçu comme propriété intrinsèque, c'est-à-dire comme une caractéristique du robot pensé comme *sujet de socialité*. Cela est très simplement résumé par Duffy *et al* (1999) :

« La socialité implique l'existence de relations interactives. Un agent capable d'un comportement interactif et communicatif est considéré comme social. La simple existence de deux robots autonomes dans le même environnement impose des formes de contact social, qu'il soit direct ou indirect. »

De ce point de vue, l'enjeu général de la sociobotique semble être d'étudier dans le détail le fonctionnement des interactions sociales humain-machine (ISHM) pour les reproduire techniquement et ce quelle que soit leur finalité.

Cette approche du social comme propriété dispositionnelle soulève toutefois deux difficultés majeures. Premièrement, nous avons vu que ces recherches donnent généralement à la figure du sujet humain un caractère *pré-déterminant*. Autrement dit, la socialité humaine s'impose comme mesure de la socialité robotique. Le sociobot est alors pensé comme l'humain qu'il

n'est pas et qu'il ne sera jamais par définition. Loin d'être dépassée, la dichotomie sujet/objet se voit renforcée par la tentative technicienne de concevoir une subjectivité robotique dont l'étalon de référence est la subjectivité humaine. Deuxièmement, cette approche ne répond pas à la question de savoir comment penser les sociobots *actuels*, ceux qui existent déjà, aussi « défailants » ou « incomplets » soient-ils. Pour être très précis, ce problème est lui aussi pensé à l'aune de la dichotomie sujet/objet, ce qui a pour effet d'aboutir à une conception graduelle de la socialité, comme en témoigne la classification de Breazeal. Or, faire de la socialité un continuum peuplé de créatures plus ou moins « sociales » n'a de sens que si ce même social est entièrement assimilé à l'interaction puisque, derrière la généralité du concept, se cache en réalité l'idée quantitative du nombre d'aptitudes à interagir. A cet égard, un robot apte à interagir physiquement et verbalement est un robot plus « social » que celui n'ayant qu'une aptitude à l'interaction verbale, par exemple⁶⁰. Cette assimilation de la socialité à l'interaction n'en demeure pas moins contestable, comme nous allons le voir.

2.2.2.1.1 *Première critique : la dimension « interobjective » des interactions*

A ce stade, il est nécessaire d'exposer avec précision ce que l'on entend lorsqu'on parle d'interaction sociale. Dans un article intitulé « Une sociologie sans objet. Remarques sur l'interobjectivité », Bruno Latour en donne une « définition classique » d'autant plus claire qu'il a pour objectif de montrer que le social ne saurait s'y réduire.

« L'interaction sociale suppose la présence de plusieurs éléments constitutifs : il doit y avoir au moins deux acteurs ; ces deux acteurs doivent être présents physiquement face à face ; ils doivent se relier par un comportement qui implique une communication ; enfin, le comportement de chacun doit évoluer en fonction des modifications apportées au comportement de l'autre, d'une façon telle qu'il y ait émergence d'un comportement imprévu qui ne soit pas simplement la somme des compétences engagées par les acteurs avant cette interaction » (Latour, 1994b, p.37-38).

⁶⁰ Outre la critique que nous formulons dans le point suivant, une telle approche pose une autre difficulté liée à la dimension qualitative des aptitudes à l'interaction : y a-t-il des aptitudes à l'interaction intrinsèquement plus « sociales » que d'autres ou sont-elles toutes équivalentes de ce point de vue ?

La référence aux comportements émergents est ici cruciale car bon nombre de robots qualifiés de sociaux sur la base de leur aptitude à interagir socialement sont en réalité bien incapables de participer à une telle émergence ou, alors, de façon aussi limitée qu'intermittente. A cet égard, nombre d'échanges avec eux relèvent plus de la somme de compétences que d'interaction ainsi entendue. Quoi qu'il en soit, cette définition classique pense l'interaction sociale en termes d'intersubjectivité. Latour ne conteste pas la dimension intersubjective de nos interactions, mais la réduction de celles-ci à celle-là. Une telle exclusivité décrit bien mieux, selon lui, la socialité des babouins que la socialité humaine.

« En effet, la sociologie des singes se présente comme l'exemple extrême de l'interactionnisme, puisque tous les acteurs y sont co-présents et s'engagent, face à face, dans des actions dont la dynamique dépend, en continu, de la réaction des autres. [...] La question du rôle exact de l'interaction et de son aptitude à composer toute la société se pose déjà chez les primates – et ne se pose peut-être que là » (Latour, 1994b, p.38).

Contrairement à la socialité des primates, les interactions humaines sont toujours « cadrées », de sorte qu'il est impossible au reste de la vie sociale de les envahir. Selon Latour, cela s'explique par le fait que nos interactions sociales procèdent non seulement d'échanges intersubjectifs, mais également de médiations « interobjectives ». Ce concept d'interobjectivité attire l'attention du lecteur sur une réalité irréductible de la socialité humaine, à savoir qu'elle ne se compose jamais uniquement d'acteurs humains, mais toujours d'associations avec des non-humains de type artefactuel.

« Chez les humains, il est impossible de reconnaître une interaction qui ne ferait pas appel à une technique. Chez les singes, l'interaction peut proliférer, appelant à la rescousse, de proche en proche, l'ensemble de la troupe. Chez l'humain, l'interaction est le plus souvent localisée, cadrée, tenue. Par quoi ? Par le cadre justement, constitué d'acteurs qui ne sont pas humains » (*ibid.*, p.52).

Pour illustrer cette idée, l'auteur prend l'exemple d'une situation commune (en 1994, du moins), à savoir l'achat de timbres-poste auprès d'une « guichetière » de La Poste. L'interaction sociale qui permet cette transaction est elle-même permise ou plutôt composée, informée par une foule d'acteurs autres que les sujets du face à face : le guichet, l'hygiaphone, le formulaire, le stylo mis à disposition du client, l'agencement général du bureau de poste

etc. Selon Latour, faire fi de ces éléments non-humains – ceux qu’il appelle les « masses manquantes de la sociologie » (Latour, 1992) – c’est se condamner à une compréhension partielle de la socialité humaine. Cette nature hybride de l’interaction explique qu’elle prenne la forme paradoxale d’un cadre et d’un réseau.

« On dit, sans y regarder de trop près, que nous interagissons face à face. Certes, mais l’habit que nous portons vient d’ailleurs et fut fabriqué il y a longtemps ; les mots que nous employons n’ont pas été formés pour la situation ; les murs sur lesquels nous nous appuyons furent dessinés par un architecte pour un client et construits par des ouvriers, toutes personnes aujourd’hui absentes bien que leur action continue à se faire sentir. La personne même à laquelle nous nous adressons provient d’une histoire qui déborde de beaucoup le cadre de notre relation. De plus, elle n’est pas forcément présente dans l’interaction, non seulement parce qu’elle peut avoir "l’air absent", mais parce qu’il peut s’agir du masque d’une fonction définie ailleurs par d’autres. Si l’on voulait dessiner la carte spatio-temporelle de ce qui se présente dans une interaction, et si l’on voulait dresser la liste de tous ceux qui sous une forme ou sous une autre y participent, on ne discernerait pas un cadre bien délimité, mais un réseau très échevelé multipliant des personnes, des dates et des lieux fort divers » (Latour, 1994b, p.40).

Et l’auteur de conclure :

« On ne doit jamais quitter l’interaction, les interactionnistes ont raison, mais si l’on suit celle des humains on ne reste jamais en place, jamais en présence des mêmes acteurs et jamais dans la même séquence de temps » (Latour, 1994b, p.53).

Le paradoxe de l’interaction sociale humaine semble donc se lever : sa réticularité est celle d’un enchevêtrement d’acteurs hétérogènes aussi bien sur le plan ontologique que spatio-temporel dont les associations permettent au cadre d’émerger. Nous verrons, un peu plus bas, qu’une telle approche s’appuie sur une conceptualisation de l’action à la fois comme faire et comme faire-faire (et donc sur une conceptualisation des acteurs comme faiseurs et comme faisant-faire). Pour l’heure, nous souhaitons mettre l’accent sur cette critique latourienne où l’interaction sociale est pensée comme un agencement compliqué⁶¹ d’intersubjectivité et

⁶¹ Dans ce texte, Latour distingue les interactions sociales simiesques et humaines par une opposition entre complexité et complication. Les premières sont complexes au sens où elles comportent simultanément un grand nombre de variables non discrétisables. Les secondes, quant à elles, sont compliquées en ceci que les variables

d'interobjectivité et où, par conséquent, l'intelligence sociale (Dautenhahn, 2007a) exprimée en aptitudes interactionnelles ne représente qu'une partie du « social ».

2.2.2.1.2 *Seconde critique : la coordination affective ou l'origine infrasubjective de la socialité*

L'idée d'une socialité strictement intersubjective peut faire l'objet d'une seconde critique empruntant à la recherche en neurosciences sur les neurones et mécanismes miroirs. Cette critique ne se concentre pas sur l'*à-côté* de l'intersubjectivité, comme avec l'interobjectivité, mais sur l'*en-deça* de la subjectivité, c'est-à-dire sur ce qui se joue au niveau *infra*-personnel lors des interactions.

Appréhender la socialité comme une propriété dispositionnelle peut en effet conduire à concevoir ces interactions, malgré leur nature rétroactive, comme des échanges entre sujets déjà entièrement formés. Le social émergerait alors de ces échanges entre individus *indépendants* les uns des autres. Dans cette logique, l'intersubjectivité n'est pas constitutive des subjectivités, mais, au contraire, est constituée par les subjectivités qui lui préexistent. Or, c'est là un point de départ solipsiste largement remis en cause par la littérature, à commencer, en philosophie, par la phénoménologie. Ici, nous proposons d'examiner cette remise en question à partir des recherches en cognition incarnée (*embodied cognition*) et, plus précisément, en suivant la discussion proposée par Dumouchel et Damiano du phénomène de « coordination affective ». L'idée qu'ils présentent répond directement au problème évoqué précédemment à propos du bouclage affectif et, *a fortiori*, des robotiques externe et interne des émotions.

« L'hypothèse centrale est que les phénomènes affectifs, dans leur ensemble, participent de la dynamique intersubjective par laquelle les agents déterminent réciproquement leurs états émotionnels et coordonnent leurs intentions d'action. [...] Plus particulièrement, cette dynamique affective donne forme à notre environnement social. L'approche de la coordination affective instaure le domaine affectif comme le lieu de notre socialité primordiale, proprement biologique, qui précède et englobe les relations

qui les composent sont présentes de façon successive et, ce faisant, peuvent faire l'objet d'une discrétisation analytique.

sociales telles que les étudient la sociologie ou l'anthropologie sociale par exemple » (Dumouchel & Damiano, 2016, p.150).

Contrairement aux théories computationnelles, la coordination affective réfute la thèse d'un accès à l'esprit de l'autre et donc à ses émotions par la construction d'une simulation mentale ou par l'analyse du comportement expressif en question (Vernon 2014, Dumouchel & Damiano, 2016). Elle récuse également l'idée qu'un tel accès repose sur la reconnaissance des émotions d'autrui. La coordination affective dont parlent Dumouchel et Damiano pense les émotions et leur partage comme des objets de syntonisation plutôt que comme des objets d'accession. Une telle hypothèse prend racine dans les travaux de Vittorio Gallese (2005) sur les neurones miroirs. Lorsqu'elles sont stimulées, ces cellules nerveuses provoquent une coactivation neuronale entre l'acteur et l'observateur d'une interaction. Plus précisément, cette spécularité cellulaire fait que le processus neuronal responsable de l'action de l'un s'active aussi chez celui qui observe sans agir au sens courant du terme. Voilà qui explique l'idée de syntonisation affective dont le principe est que « l'observateur participe à l'expérience émotionnelle de l'observé, ou du moins qu'il partage son soubassement neurologique » (Dumouchel & Damiano, 2016, p.153). Selon Gallese, la syntonie déclenchée par les phénomènes miroirs de notre cerveau constituerait le fondement *incarné* de l'empathie dont la dimension affective est une des manifestations. Son autre hypothèse, bien plus radicale, est que cette spécularité ou « communion neuronale » comme disent Dumouchel et Damiano « suspend la frontière entre soi et l'autre » (*ibid.*).

« Nos émotions et nos réactions empathiques ne sont en ce sens ni des productions privées ni des entreprises solitaires. Ce sont plutôt des œuvres communes auxquelles plusieurs participent. Les émotions et l'affect renvoient bien au corps, mais ce corps ne se limite pas à l'organisme individuel. Les émotions s'incarnent dans ce qu'on pourrait nommer un "corps social" » (*ibid.*).

Il y aurait donc, à une échelle infra-personnelle, une genèse sociale à la fois *pré-* et *transindividuelle* de nos émotions et, plus largement, de nos subjectivités. En ce sens, l'interaction ne fait pas seulement émerger des « comportements imprévus » comme le dit sa définition classique, elle constitue les subjectivités auxquelles se rattachent ces comportements. Une telle perspective gomme la distinction entre les aspects externes et les aspects internes des émotions où plutôt les conjugue en ce que « l'expression affective

"externe" d'un agent est directement responsable de la réaction "interne" de l'autre » (*ibid.*, p.153-154). Bien que cette conjugaison brouille momentanément la frontière entre les agents, elle n'efface pas leur identité.

« Au contraire, la coordination affective est le moyen par lequel est constamment constituée, définie et redéfinie l'identité des agents. Cette influence est directe dans la mesure où elle n'exige et ne suppose aucun intermédiaire entre l'expression affective de l'un et la réaction neuronale de l'autre. [...] Cette influence est réciproque, parce que l'autre réagit immédiatement et expressivement à mon expression affective et sa réaction engendre chez moi un nouvel état neuronal. L'expression affective conduit ainsi à la mise en place d'une véritable boucle affective qui détermine l'expression affective et les intentions d'action des agents » (*ibid.*, p.154).

Ainsi, la « socialité primordiale » dont parlent Dumouchel et Damiano ne semble pas tant relever de la propriété dispositionnelle que de la propriété relationnelle. En ce sens, on peut considérer qu'un robot « social » est un robot qui s'invite dans « la dynamique intersubjective de codétermination réciproque des affects »⁶² (*ibid.*, p.155). La participation des sociobots à cette dynamique reste toutefois des plus sommaires tant leur capacité de réaction à nos expressions affectives sont limitées.

Ces limites montrent bien que l'enjeu actuel des roboticiens n'est pas tant de créer un robot « sujet de socialité » au sens d'un chimérique robot sociable, mais plutôt au sens de robot *suffisamment* interactif. Un tel sociobot est un artefact dont le design permet de renouveler indéfiniment l'intérêt de l'utilisateur et sa volonté d'interagir avec lui, quand bien même s'agirait-il de courts échanges. La socialité est alors plus modestement appréhendée comme capacité à susciter et entretenir la *ré*-interaction. Or, on s'aperçoit que ce ne sont pas les robots les plus sophistiqués techniquement (et/ou dépositaires des ambitions parfois démesurées de leurs concepteurs) qui s'acquittent le mieux de cette tâche. Prenons l'exemple des robots Paro et Lovot pour illustrer cette idée.

⁶² Bien entendu, le robot ne s'invite pas dans cette dynamique en tant que sujet, mais en tant qu'artefact simulant quelque-chose comme une intériorité.

2.2.2.2 Paro et Lovot : des robots sociaux ou sociables ?

Paro est un bébé phoque mécatronique « d'assistance thérapeutique » développé par Takanori Shibata. « Recouvert d'une fourrure synthétique à fibres bactéricides et fongicides »⁶³, il a pour fonction d'intervenir comme substitut dans le cadre de thérapies dites animales (principalement en maison de retraite). De nombreuses études semblent attester des vertus rassérénantes et socialisantes de l'artefact pour les patients atteints de troubles neurodégénératifs tels que la maladie d'Alzheimer ou la démence sénile (Wada *et al*, 2005 ; Kidd *et al*, 2006 ; Shibata, 2012 ; Sabanovic *et al*, 2013 ; Shibata & Coughlin, 2014 ; Sabanovic & Chang, 2016 ; Demange *et al*, 2018). Sa présence semble également favoriser l'amélioration de l'humeur et la réduction des états dépressifs chez les enfants hospitalisés à long terme (Shibata *et al*, 2008). L'interactivité de Paro n'est pas proactive, mais réactive. Il correspond à la catégorie des robots « interfaces sociales » proposée par Breazeal. Comme l'expliquent Dumouchel et Damiano, l'intérêt du blanchon réside dans sa non-instrumentalité :

« La thérapie animale est sa fonction et Paro l'accomplit en ne faisant rien de particulier, en étant simplement là, disponible pour des interactions répétées, des interactions qui ne servent aucun autre but ou fonction que d'être des interactions sociales. C'est pourquoi il est possible d'interagir avec lui plutôt comme avec un compagnon ou une amie, que comme avec un outil ou un instrument dont la valeur ou l'utilité est liée à une tâche particulière. C'est peut-être aussi pourquoi on ne se fatigue pas aisément de Paro, pourquoi l'intérêt subsiste et les effets thérapeutiques demeurent, même après plusieurs mois » (Dumouchel & Damiano, 2016, p.170).

Paro rassérène, rassure, console et tient compagnie par le simple fait d'être ce qu'il est : doux, mignon et, bien sûr, physiquement présent. Ses fonctionnalités techniques sont toutes entières tournées vers l'expressivité corporelle. Il peut bouger les nageoires, la queue et la tête, cligner des yeux, émettre deux petits cris (un cri d'« appel » et un autre de « réponse »), tourner la tête en direction du son quand on l'appelle, réagir aux bruits forts et soudains. En outre, Paro est doté de capteurs lui permettant de manifester son « plaisir » ou son « déplaisir » selon qu'on le caresse ou qu'on le malmène. Enfin, malgré cette faible diversité de comportements de base, la combinatoire permise par son architecture logicielle

⁶³ <https://www.phoque-paro.fr/phoque-paro/>, site internet consulté le 26/11/2020.

(mouvements sélectionnés, vitesse d'exécution etc.) offre un nombre indéfini de comportements émergents. Ces fonctionnalités forment l'ancrage technique à partir duquel les interactions sociales avec le blanchon adviennent⁶⁴.



Selon Dumouchel et Damiano, « Paro est hyperboliquement social » (*ibid.*, p.172) au sens où sa socialité est exclusivement réactive et, ainsi, totalement dépendante à l'égard d'interactions initiées par d'autres que lui.

« Il n'y a rien d'autre dans son monde que les agents qui entrent en contact avec lui. Il dépend entièrement d'eux pour exister, car il ne devient un agent que dans le rapport aux autres lorsqu'il est pris directement comme cible de leurs actions. Le reste du temps, ce n'est qu'une chose morte, un objet inanimé » (*ibid.*).

Bien sûr, cette dimension exclusivement réactive constitue un axe à part entière du design de Paro. Takanori Shibata n'a pas choisi l'apparence de Paro ni son schéma interactionnel par hasard : leur objectif est de donner aux IHM une récurrence dans le temps ou, autrement dit, de ne pas laisser l'utilisateur. Ainsi, la forme de blanchon permet de neutraliser les attentes des usagers puisque, contrairement aux chiens ou aux chats, très peu de personnes ont déjà eu affaire à un bébé phoque (cf. la règle selon laquelle la forme perçue de l'artefact doit correspondre à ses capacités effectives⁶⁵). La modalité réactive obéit à la même logique. La douceur et la mignonnerie de Paro incitent à entamer des interactions qui seront entretenues

⁶⁴ Ces interactions sont pensées pour être au plus proche de celles qu'on pourrait avoir avec un animal de compagnie, les inconvénients en moins.

⁶⁵ Voir p.83.

exclusivement sur le mode du répondant. De ce fait, le blanchon se présente à la fois comme un artefact relationnel au sens de Turkle et comme une peluche passive sur laquelle l'utilisateur projette des états mentaux. Puisque l'activité de Paro est une *ré*-activité, ces projections s'avèrent essentielles pour stimuler l'intérêt de l'interacteur dans le temps. En termes de design, cela signifie que nos projections psychiques font partie du fonctionnement de l'artefact en tant que robot *social*, puisqu'il doit être perçu comme tel par l'utilisateur pour susciter l'interaction qui, dans la perspective dispositionnelle, constitue sa socialité. A ce stade, notons simplement que le relatif succès de Paro s'explique par l'intelligence des limites dont il procède : Shibata a su mettre à profit les difficultés de la robotique sociale pour concevoir son robot.

Ces analyses nous poussent à assimiler le mode d'existence social de Paro à la sociabilité, mais au sens simmelien d'*interaction pour l'interaction*, c'est-à-dire d'interaction sans autre finalité qu'elle-même.

« La rencontre sociable est celle où les formes de l'action réciproque sont détachées des motivations et des finalités concrètes de l'interaction et sont actualisées pour elles-mêmes : elles deviennent alors elles-mêmes "fin et matière de leur développement" [...]. Dans l'interaction courante, la dimension sociable n'intervient qu'à titre de médiation de l'action réciproque : la configuration d'une relation de réciprocité est fonctionnelle pour la coopération sociale. Elle n'est pas une fin en elle-même. Par contre, dans la rencontre sociable pure, l'action réciproque est visée pour elle-même » (Quéré, 1988, p.83-84).

Dans cette perspective, un robot « sociable » est donc un robot qui n'a pas d'autre fonction que l'interaction elle-même et qui, en un sens, est totalement inutile. Si Paro correspond bien à cette modalité du social, il n'est pas le seul. Le Lovot (contraction de « love » et de « robot ») développé par la société japonaise Groove X est un autre très bon exemple de sociabilité artificielle. L'artefact est conçu pour donner et recevoir de l'affection. L'entreprise affirme d'ailleurs qu'il « fonctionne à l'amour » (« *powered by love* »⁶⁶). Les comportements du robot rappellent fortement ceux du chien. Câliner un Lovot agit sur son attitude, mais peut aussi engendrer la « jalousie » d'un autre Lovot et l'attirer aussitôt (Groove X ne vend ses produits que par paires). En outre, l'automate reconnaît celles et ceux qui se montrent aimants envers lui et peut ainsi leur demander encore plus d'attention en les suivant partout dans la maison.

⁶⁶ Voir le site internet de Lovot : <https://lovot.life/en/> (consulté le 30/11/2021).

Groove X a rendu possible l'interaction par « contact oculaire » (*eye contact*) avec ses créatures pour une plus grande impression de vie. Enfin, les Lovot peuvent accueillir leur propriétaire lorsqu'il ou elle rentre à la maison et effectuer des « patrouilles » filmées en cas d'absence prolongée. Cette dernière fonctionnalité mise à part, le robot n'a pas grand-chose d'un artefact utilitaire. Pour Kaname Hayashi, fondateur de GrooveX, l'inutilité des Lovot est motif de fierté :

« Je voulais créer une opportunité pour les humains d'aimer. [...] Notre robot n'accomplit aucune tâche pour les humains et ne propose aucun contenu de divertissement. Mais les chiens et les chats non plus. Ce qu'il sait faire, c'est vous reconnaître et vous enquiquiner. C'est le but de notre robot » (JapanGov, 2019).

Or, absence d'utilité ne signifie pas absence d'effectivité. Nous l'avons vu, la sociabilité de Paro peut intervenir dans un cadre thérapeutique d'accompagnement des personnes âgées souffrant d'isolement social et/ou de troubles neurodégénératifs. La rencontre sociable, non fonctionnelle par nature, est alors fonctionnalisée. Toutefois cette fonctionnalisation intervient uniquement à l'échelle du dispositif médical dans lequel elle intervient. En d'autres termes, les patients et patientes ne sociabilisent pas avec l'artefact pour une autre raison que la sociabilisation elle-même.

Dans cette perspective, nous soutenons 1) que les développements techniques de la sociobotique contemporaines relèvent majoritairement d'une sociabilité rudimentaire et que 2) loin des discours de certains fabricants sur l'autonomie des robots sociaux, leur principal intérêt est à chercher dans l'espace d'expérimentations socio-anthropologiques autant qu'ontologiques qu'ouvre cette sociabilité rudimentaire (Grimaud & Paré, 2011 ; Grimaud, 2012), ce qui n'est pas sans créer un problème majeur dès qu'il s'agit de commercialiser ce genre d'artefacts. A cet égard, le gros des services assignés aux robots sociaux relève plus du prétexte que d'une réponse sérieuse à des besoins existants. En effet, la plupart des cas d'usage imaginés sont déjà pris en charge par d'autres supports qui remplissent leur fonction de façon beaucoup moins erratique. Illustrons cette idée avec un exemple tiré de notre terrain CIFRE. En mars 2017, nous avons participé à la première expérimentation publique conduite par Spoon. Pendant trois semaines, l'un de ses robots, le BigSpoon, fut installé dans l'une des allées d'un centre commercial rennais. L'objectif principal de cette expérience était d'étudier

la substituabilité d'un robot social aux bornes interactives d'information, trop coûteuses eu égard à leur faible taux d'utilisation.



Le BigSpoon installé dans le centre commercial Alma de Rennes

BigSpoon fut alors programmé pour prendre en charge deux scénarios d'usage. Le premier était un scénario d'informations et d'orientation (boutiques, toilettes) par langage verbal et corporel. Le second, quant à lui, était un scénario d'apprentissage baptisé « Agora ». L'idée était de donner aux passants la possibilité d'enseigner oralement des connaissances au robot afin que chaque interaction participe à son « éducation ». A la fin des trois semaines de test, le constat était sans appel : le scénario « utilitaire », c'est-à-dire celui faisant appel à des fonctions extrinsèques à l'interaction (FEI), avait été totalement délaissé au profit d'Agora et de sa fonction intrinsèque à l'interaction (FII). Bien qu'elle ne soit pas scientifique, mais strictement empirique, cette expérience va dans le sens des autres observations généralement effectuées sur le terrain : c'est « l'interaction pour l'interaction » qui, aujourd'hui, fonde la valeur d'un sociobot. Voilà pourquoi les débouchés les plus prometteurs pour ces artefacts sont ceux où l'interaction sociale produit en elle-même ses effets, à l'instar des dispositifs de socialisation des personnes âgées ou des enfants présentant un trouble du spectre autistique (Pennisi *et al*, 2016).

2.2.3 Le social comme propriété systémique : la fonction sociale

2.2.3.1 Le sociofonctionnalisme : interaction et fonction sociales

La seconde réponse s'inscrit dans le prolongement de l'approche interactionniste décrite à l'instant, sans réduire le social aux seules interactions. Le sociobot y est pensé comme

l'artefact issu de la rencontre entre social et sociétal aux sens proposés plus haut, c'est-à-dire entre capacité d'interaction sociale et fonction(s) sociale(s). Nous l'avons vu, Dumouchel et Damiano s'inscrivent dans ce schéma en conceptualisant le robot social sous l'angle sociofonctionnel de la substitution : le substitut est un robot dont les capacités d'interaction sont mises au service du rôle qu'il joue au sein de la société en lieu et place d'un humain dont il reçoit l'autorité. En d'autres termes, le social est ici interprété à l'aune des fonctions dont l'agent s'acquitte à travers ses interactions avec l'extérieur. Par exemple, un robot d'assistance médicale programmé pour rappeler à une personne âgée de prendre ses médicaments (à l'hôpital ou à domicile) devient momentanément le relai de l'autorité médicale du personnel soignant et le curateur de la personne en question. Ainsi conçu, le social se présente comme une propriété systémique. Autrement dit, le substitut n'est pas social en vertu de ses seules capacités d'interaction, mais aussi en vertu de son insertion dans un contexte social, une organisation précise dont il tire son statut de *socio-bot*. Bien sûr, le social comme propriété systémique s'appuie ici sur la propriété dispositionnelle qu'est l'interaction – il faut bien qu'il y ait des agents pour s'acquitter des fonctions instituées par le système –, mais il ne s'y réduit pas. Dès lors, on ne saurait comprendre la socialité d'un robot uniquement à l'aune de caractéristiques techniques. Un substitut d'assistance médicale peut techniquement jouer son rôle à la perfection, c'est-à-dire vérifier sans aucune erreur que chaque patient a pris les bons médicaments à la bonne heure et prévenir une personne à charge si tel n'est pas le cas, sans être accepté comme curateur légitime par certains patients. Face à un tel rejet, l'artefact est *entravé dans sa socialité* pour des raisons non pas d'ordre technique, mais moral⁶⁷ (ou, dans d'autres situations, organisationnel, institutionnel etc.).

2.2.3.2 Fonction sociale *versus* autonomie sociale

Cela étant, cette appréhension du social sous l'angle fonctionnel doit être maniée avec précaution en ce que la réduction des êtres sociaux aux fonctions qu'ils occupent (du fait de leurs capacités et/ou de leurs origines sociales) relève d'une logique potentiellement totalitaire. Il suffit pour cela de penser à la société décrite par Aldous Huxley dans *Le Meilleurs*

⁶⁷ C'est d'ailleurs la volonté de contrôle de cette dimension morale du social qui donne naissance à des expressions programmatiques telles que la fameuse « acceptabilité sociale » des technologies qui présuppose l'acceptation en la faisant dépendre uniquement des moyens de la susciter. L'objectif est alors non pas de rendre acceptable voire désirable une technologie, mais de la *faire accepter*, ce qui est très différent.

des mondes où les naissances, contrôlées aux moyens de dispositifs ectogénétiques, obéissent aux besoins fonctionnels de la société. Chaque nouveau-né est génétiquement conçu pour occuper une position sociale précise tout au long de sa vie (le haut de l'échelle est occupé par les dirigeants « Alphas » et le bas par les « Epsilons », condamnés aux basses besognes). L'œuvre d'Huxley décrit ainsi une société fossilisée dont l'ordre apparent est surtout synonyme de machinisme social. Il est donc primordial de lier dialectiquement interaction et fonction sociales pour penser la socialité. Dumouchel et Damiano relativisent eux-mêmes la dimension sociofonctionnelle de leur raisonnement et affirment qu'une authentique socialité robotique est à chercher dans le développement d'une *autonomie sociale* telle que les substituts puissent s'affranchir, même momentanément, des fonctions qui sont les leurs. Le social comme propriété systémique repasse ici du côté de la propriété dispositionnelle qu'est la capacité d'interaction ou, plus précisément, la capacité de coordination et de re-coordination indéfinies. Citons longuement les deux philosophes sur ce point :

« Les êtres sociaux ne servent aucune fin particulière et doivent pouvoir remplir à divers moments des tâches différentes. Un robot social au sens d'un *substitut* ne peut par définition être simplement asservi à un unique rôle ou à une fonction particulière faute de quoi ce n'est plus un être social. Plus généralement et plus profondément, la socialité n'a aucune fin particulière sauf elle-même. Elle forme la condition nécessaire du surgissement de toutes fins au sein d'une culture matérielle, laquelle constitue un arrière-fond contraignant l'éventail des fins possibles. La socialité est une dimension fondamentale de la condition humaine, par rapport à laquelle sont déterminées tous les usages et tous les objectifs que nous pouvons avoir. Les êtres sociaux ont des fins ou des buts qu'ils se donnent à eux-mêmes ou qui leur sont imposés de l'extérieur. S'ils peuvent être réduits ou condamnés à répéter inlassablement la même action, la capacité de transcender le rôle ou la fonction, auxquels d'autres personnes tentent de les contraindre, ou l'ordre des choses tente de les restreindre, est inhérente à leur condition sociale. La socialité humaine suppose une capacité de coordination réciproque indéfinie. [...] Un robot social au sens où nous l'entendons doit donc pouvoir s'adapter à des situations nouvelles, imprévisibles [...]. Les substituts artificiels sont différents des autres objets techniques dans la mesure où tout agent social doit nécessairement participer de cette capacité d'adaptation et de coordination, même de façon limitée. [...] L'horizon ultime de la socialité artificielle [...] est effectivement de créer un robot *qui ne sert à rien et qui peut faire n'importe quoi* » (Dumouchel & Damiano, 2016, p.43-46).

La dimension systémique du social apparaît alors comme la « surcouche » d'une dimension plus originaire, à savoir une autonomie téléologico-interactionnelle. A cet égard, la socialité apparaît comme un mode d'existence érigeant l'indétermination fonctionnelle et, plus largement, téléologique en horizon de toute téléologie. On retrouve ici la figure archétypique du robot sociable de Cynthia Breazeal et de ses capacités d'interactions proprement humaines.

Une critique similaire peut être adressée à l'approche de Bartneck et Forlizzi qui, rappelons-le, pensent le sociobot sous l'angle du respect des « normes comportementales » attachées à son rôle. Le problème, là encore, est que l'obéissance aux règles, qu'elles soient explicites ou implicites, ne représente qu'une partie du social et de sa dynamique. La transgression et l'invention de nouvelles normes en sont tout aussi constitutives. On pourrait toutefois objecter que cette position opère un recentrement hasardeux sur l'humain et que le propre d'une *socialité robotique* se situe précisément dans son déterminisme normatif et fonctionnel. C'est une remarque à laquelle Peter H. Khan et ses collègues répondent que, si tel est le cas, « alors le terme robot social ne permet pas aisément de distinguer la socialité humaine de la socialité non-humaine » (Khan *et al*, 2004, p.546). Or, comme le propose Lestel, l'un des défis majeurs de la sociobotique repose peut-être dans son invitation au relativisme conceptuel ou, autrement dit, à la reconfiguration du sens que nous attribuons aux mots en fonction des êtres auxquels ils s'appliquent. Remarquons toutefois qu'un remaniement conceptuel fondé sur la définition de Bartneck et Forlizzi reviendrait à faire de la socialité robotique l'exact contraire de la socialité humaine, ce qui n'est pas sans interroger le choix de conserver ce concept pour décrire deux réalités antinomiques. Parler de « robot socialement interactif » comme Fong *et al* (2002) ou, plus simplement, de « robot interactif » semblerait alors plus pertinent.

2.2.4 Le social comme propriété relationnelle

2.2.4.1 Robot social ou robot socialement interactif ?

La troisième réponse que nous identifions s'inscrit dans le sillage de la théorie de l'acteur-réseau (ANT) pour laquelle « l'épithète "social" ne désigne plus une chose parmi d'autres, comme un mouton noir au milieu de moutons blancs, mais un *type de connexion* entre choses

qui ne sont pas elle-même sociales » (Latour, 1994b, p.13). La socialité ne se situe pas *dans* le robot, comme dans la perspective dispositionnelle, ni *autour*, comme dans la perspective systémique, mais procède des *associations*, des liens qui se tissent entre l'artefact et l'écosystème auquel il est intégré. Autrement dit, le social est une propriété relationnelle : il émerge de la mise relation de deux ou plusieurs entités. De ce point de vue, un robot n'est pas plus social qu'une fourchette, qu'un micro-onde ou qu'un ordinateur ; il est donc conceptuellement plus juste de parler, comme Fong *et al* (2003) de robot « socialement interactif » (*socially interactive robot*) ou de robot « sociable » (au sens simmelien). Si la sociabilité relève de l'interaction entre deux ou plusieurs actants, la socialité, elle, procède d'un régime associatif bien plus vaste pour lequel nous proposons le terme d'*interactance*.

2.2.4.2 De l'interaction à l'interactance : une typologie de l'agentivité des artefacts

Nous employons le néologisme d'interactance pour désigner l'association, les échanges effectifs qui adviennent entre deux ou plusieurs actants. La notion d'actant est quant à elle empruntée à la sémiotique par les sociologues de l'acteur-réseau, notamment par Bruno Latour (1991, 1994b, 2006). « [Elle] permet d'élargir la question sociale à tous les êtres qui interagissent dans une association et qui s'échangent leurs propriétés » (Latour, 1994b, ndbp 25, p.51). En d'autres termes, l'actant renvoie à toute entité, humaine ou non-humaine, dont la présence compose *effectivement* la sphère sociale et contribue à configurer une situation donnée. Dans cette perspective, la socialité apparaît comme le résultat de la compénétration des actances, c'est-à-dire de l'interactance.

Avant d'aller plus loin, une précision s'impose quant à l'importance de bien distinguer l'interactance de l'interaction. Aidons-nous, pour cela, d'une remarque formulée par Peter-Paul Verbeek, actuel chef de file du courant postphénoménologique, très influencé par Latour et les STS.

« Le champ du design d'interaction se fonde sur l'idée, qu'en dernière instance, ce ne sont pas des *choses* qu'il faut concevoir, mais plutôt les *interactions* entre les humains et les choses. Pourtant, l'interaction n'est peut-être pas toujours le concept le plus utile pour comprendre les relations entre les humains et les produits, ou pour comprendre les artefacts technologiques en général. [...] *Interaction* peut être littéralement traduit par "action entre". Le terme indique ce qui se déroule entre un humain, d'une part, et un

artefact technologique, de l'autre. Les deux entités sont mutuellement en relation et agissent l'une sur l'autre. Le concept d'interaction présuppose donc l'existence de sujets humains et d'objets technologiques *entre* lesquels il y a un genre spécifique d'*activité* » (Verbeek, 2015, p.26).

Cette idée nous indique que l'interaction n'est qu'un type particulier de relation nous unissant aux artefacts. Elle est l'interprétation moderne (et restrictive) de nos relations au monde et, plus précisément, aux techniques. La critique de Verbeek a pour but de substituer le concept de *médiation technique*, initialement développé par Latour, à celui d'interaction pour rendre compte plus richement de ces relations. Disons-le d'emblée, les notions de médiation et d'interactance possèdent, ici, le même contenu conceptuel. L'artefact conçu comme médiateur de nos relations au monde échappe à la dichotomie sujet/objet ainsi qu'à toute détermination *a priori*. Il s'agit d'un concept de nature relationnelle au sens où il ne s'intéresse aux individus qu'en tant qu'ils s'informent mutuellement au sein de leur relation. Étudier les médiations plutôt que les interactions permet ainsi d'appréhender ce qu'une technique *fait* et *nous fait* en plus de ce que nous en faisons. L'interaction n'est que la partie *fonctionnelle* de nos relations aux artefacts (*ibid.*) que le concept de médiation technique enrichit d'une dimension herméneutique (participation des artefacts à notre accès perceptif au monde et à notre interprétation de ce dernier) et existentielle (participation des artefacts à la façon dont nous menons notre vie et nous réalisons en tant qu'humains) (Verbeek, 2005).

Pour comprendre en quoi un robot peut être qualifié de social, il semble donc nécessaire d'étudier son actance *in situ*, plus communément appelée « agentivité »⁶⁸ dans la littérature. Philip Brey, Professeur de philosophie à l'Université de Twente aux Pays-Bas, identifie cinq formes d'agentivité artefactuelle, chacune se manifestant sous forme d'affordances⁶⁹ ou de contraintes (2006) :

Affordances et contraintes comportementales : un artefact est réputé avoir une agentivité comportementale dès lors qu'il encourage ou décourage certains comportements pendant

⁶⁸ Bien que plus répandu, le terme d'agentivité (traduction de l'anglais *agency*) exprime moins bien, selon nous, le caractère désanthropocentré de l'action.

⁶⁹ Il reprend ici le terme introduit par James J. Gibson. Tantôt traduit par « potentialité », tantôt conservé tel quel, l'affordance désigne la capacité d'un objet au sens large à suggérer son mode d'emploi et l'incitation à l'usage qui découle de cette suggestion.

son usage ou pour permettre son usage. L'exemple caractéristique est celui du dos d'âne donné par Latour (1994a, 1994b) : l'artefact incite à ralentir avant de rouler dessus.

Affordances et contraintes liées au profil utilisateur : nombre d'artefacts présupposent un type particulier utilisateur et, de ce fait, présentent un usage plus ou moins inclusif. Un exemple hélas typique est celui des escaliers dans le métro parisien qui, en l'absence d'alternatives adaptées aux personnes handicapées ou âgées, défavorisent de nombreux usagers actuels ou potentiels.

Affordances et contraintes matérielles et infrastructurelles : ce type d'agentivité renvoie aux conditions matérielles dont certains artefacts ont besoin pour fonctionner correctement. Par exemple, l'utilisation d'internet sur un ordinateur suppose l'existence d'une connexion permettant d'y accéder. De la même façon, un masque anti-pollution doit être équipé de filtres à particules dont la durée de vie va de quelques semaines à plusieurs mois.

Affordances et contraintes sociales : les techniques peuvent également participer à la (re)configuration de nos statuts, rôles ou relations sociales ou à celle de nos organisations et de nos institutions. Dans son célèbre article « Do Artifacts Have Politics », Langdon Winner donne deux exemples qui illustrent ces idées. Le premier est celui des passerelles new-yorkaises surplombant les voies rapides en direction des plages de Long Island. Construites par l'architecte Robert Moses dans les années 1930, elles avaient été placées de sorte à ce que les autobus des transports publics, largement utilisés par les populations noires défavorisées, ne puissent emprunter ces routes et accéder aux plages huppées de Jones Beach⁷⁰. Winner évoque ensuite l'énergie atomique dont la production en centrale impose un système de gestion extrêmement hiérarchisé et autoritaire.

Affordances et contraintes culturelles : toute technologie embarque des valeurs et contribue à la constitution de nos expériences et de nos actions. Ce faisant, elle participe à l'entretien ou à l'évolution culturelle et morale d'une époque. L'exemple de l'échographie analysé par Verbeek (2005) est très parlant sur ce point. L'introduction de cette technologie en

⁷⁰ Notons toutefois que cet exemple a été critiqué comme historiquement faux. Sa logique n'en reste pas moins illustrative de l'idée de contraintes sociales.

obstétrique a profondément modifié l'expérience de la grossesse et son interprétation en y associant plus étroitement les futurs pères, en faisant du fœtus un être à la fois « distinct » de sa mère et un patient potentiel, mais aussi en créant la possibilité de choisir une interruption volontaire de grossesse (IVG) pour motif médical.

Reprenons l'exemple de Paro pour illustrer ces considérations. Le blanchon, nous l'avons vu, dispose de caractéristiques techniques qui, malgré leur simplicité et leur petit nombre, favorisent une certaine récurrence de l'interaction. De ce point de vue, Paro se rapproche explicitement d'un agent social au sens courant du terme, c'est-à-dire d'un actant doté de caractéristiques ordinairement associées aux sujets telles que l'intentionnalité, la sensibilité, l'affectivité, l'interactivité etc. Or, l'effectivité de Paro ne se limite pas à ces ISHM, loin de là. La simple présence du robot peut, par exemple, favoriser les interactions entre pensionnaires d'un EHPAD⁷¹ et les inciter à passer plus de temps ensemble dans les espaces communs. C'est là un exemple d'agentivité de type « affordance », pour reprendre la catégorie proposée par Brey. A l'inverse, Paro peut également se montrer générateur de contraintes. Dans la thèse de doctorat qu'il consacre aux robots dans le système de soin japonais, l'anthropologue James Wright montre comment l'introduction d'un Paro au sein d'une maison de retraite a pu favoriser la socialisation des pensionnaires sous diverses formes, mais aussi occasionner des replis sur soi (avec le robot) perturbant l'organisation et les routines de l'institution comme, par exemple, celles des horaires dédiés au repas communs (Wright, 2018). Autre exemple très parlant : celui de l'impact du robot aspirateur Roomba sur les pratiques ménagères domestiques. Une étude conduite par la professeure et chercheuse en IHM Jodi Forlizzi a montré comment l'introduction du Roomba dans certaines familles – comparée à celle d'un simple aspirateur manuel dans d'autres – modifia durablement la répartition des rôles face aux tâches ménagères, mais aussi la façon de les pratiquer (Forlizzi, 2007). Dans les familles avec enfants, par exemple, le robot eut pour effet de les responsabiliser quant au nettoyage de leur chambre. En outre, Forlizzi note que la facilité d'utilisation et l'autonomie du Roomba substituèrent l'aspiration opportuniste à l'aspiration planifiée. Enfin, une augmentation du nombre hebdomadaire d'aspirations fut également constatée.

⁷¹ Établissement d'hébergement pour personnes âgées et dépendantes.

2.2.4.3 Le statut protéiforme du sociobot

Cette agentivité protéiforme peut aussi être attachée au statut que nous accordons au robot lui-même selon qu'il est directement ou indirectement impliqué au sein d'une situation. Dumouchel et Damiano opèrent une distinction entre deux modes d'existence du robot : celui d'agent social, qui s'exprime au sein des interactions avec Paro, et celui d'objet, lorsque l'artefact est extérieur aux interactions.

« [Si le blanchon] facilite les interactions entre les personnes, c'est en étant entre eux comme un objet plutôt que comme un agent. Paro ne peut participer que de façon extrêmement limitée à la conversation du groupe dont il est le centre d'attention. Ce n'est guère un interlocuteur, mais plutôt un facilitateur ou un prétexte. Il est *ce dont on parle* et non pas *celui à qui l'on parle*, contrairement à ce qui a lieu dans les rapports un à un » (Dumouchel & Damiano, 2016, p.172-173).

Or, le robot semble pouvoir passer du statut d'agent à celui d'objet et vice versa y compris au sein d'un même contexte interactionnel. C'est ce qu'affirme Morana Alač (2016), chercheuse en sciences de l'information et de la communication (SIC) à l'Université de Californie à San Diego, dont les expérimentations témoignent d'une (re)négociation permanente du statut de sociobot au sein d'un même contexte interactionnel. Rendant compte d'une étude menée au sein d'une école maternelle avec de jeunes enfants, leurs institutrices, un robot humanoïde « low-cost » et ses concepteurs, la chercheuse écrit :

« les caractéristiques stables de l'architecture du robot sont en relation avec des aspects plus éphémères de l'interaction : [...] les mouvements à travers la pièce, les coups d'œil aux personnes coprésentes, leurs gestes de pointage et de toucher » (Alač, 2016).

Alač suggère que le statut *perçu* du robot procède de ses « caractéristiques stables » (comme par exemple son apparence et ses fonctionnalités) associées aux diverses composantes de l'interaction telles que la spatialité et les mouvements des corps en son sein. Cette association entre éléments stables et changeants génère une *socialité intermittente* de l'artefact dont le statut oscille entre « chose » (*thing*) et « agent » au gré des interactions et, plus largement, de la configuration interactancielle du contexte expérimental.

« En interaction, le robot présente ses multiples facettes de sorte que chaque thème⁷² peut refaire surface à tout moment, s'articulant l'un à un à l'autres de façon dynamique » (*ibid.*).

On retrouve une idée similaire sous la plume de l'anthropologue Emmanuel Grimaud (2015) pour qui les robots « oscillent entre inerte et vivant » au gré des liens, interactifs ou non, que nous tissons avec eux.

En sus des discussions déjà entamées, une critique générale peut être adressée aux trois réponses dont nous venons de rendre compte, à savoir qu'aucune d'entre elles ne s'intéresse à la dimension immanquablement morale de la socialité et, corrélativement, à la phénoménalité très particulière du sociobot. En effet, quelle que soit la vision retenue, le social se présente comme un ensemble de relations qui lient directement ou indirectement des individus entre eux. En d'autres termes, le social implique toujours à la fois un rapport à soi et à l'autre. Or, dans le cas du sociobot, c'est précisément l'artefact lui-même qui se présente comme un « autre ». Le robot social appartient donc à une classe d'artefacts dont le phénomène a pour effet singulier de conjointre machinité et altérité. C'est d'ailleurs cette altérité qui, selon nous, fait que toute réflexion sur la socialité du robot invite, en même temps, à penser sa moralité. Force est pourtant de constater que l'altérité n'est pas ou très peu thématifiée en tant que telle par la philosophie des techniques et par l'éthique appliquée à la robotique. Avant de discuter ce point en profondeur dans la seconde partie de ce travail, il convient de clore notre état de l'art par l'analyse des différents courants et des différentes discussions qui, justement, composent le large champ de l'éthique robotique.

⁷² Dans l'article d'Alač, le mot « thème » (*theme*) renvoie aux deux statuts du robot dont elle parle, à savoir la chose et l'agent.

Chapitre 3 - La roboéthique : des éthiques « envers », « autour », « dans » et « avec » les robots sociaux

Les réflexions éthiques consacrées aux sociobots prennent place au sein du champ plus général de l'éthique robotique (*robot ethics*) également nommée « roboéthique » (*roboethics*). Selon Peter Asaro (2006) celle-ci se construit autour des trois grands axes de réflexion suivants :

1. L'éthique des personnes qui conçoivent et utilisent les robots⁷³ ;
2. L'éthique des systèmes intégrés aux robots ;
3. La question de la valeur morale des robots, de leurs actions et de nos actions à leur égard.

En termes plus imagés, on peut dire que la roboéthique s'intéresse à l'éthique *autour* du robot (éthique de la conception et des usages), à l'éthique *dans* le robot (éthique intégrée/programmée) et à l'éthique *envers* le robot (statut moral du robot). Asaro plaide pour l'analyse conjointe de ces trois dimensions afin d'éclairer « comment la responsabilité morale devrait être distribuée dans des contextes socio-techniques impliquant des robots » (*ibid.*, p.10). Cette approche peut être enrichie par celle de Marc Coeckelbergh (2009a) dont la contribution relève, dans la logique prépositionnelle que nous venons de proposer, d'une éthique *avec* le robot. Coeckelbergh note que les réflexions éthiques dédiées aux robots relèvent la plupart du temps d'une éthique *appliquée* au sens où elles identifient et traitent les questions morales soulevées par la robotique à l'aune de grands principes issus des différentes théories morales existantes. Dans ce schéma applicatif, l'enjeu est de concevoir ou d'utiliser les robots conformément à des critères moraux préexistants aux IHM.

« Par exemple, si l'on pense qu'il est moralement condamnable de payer pour avoir des relations sexuelles, alors, en vertu de ce critère, on jugera que les robots sexuels sont moralement inacceptables » (*ibid.*, p.218).

⁷³ Nous ajoutons à ces questions de conception et d'utilisation celles de la commercialisation des sociobots et de la communication qui y est associée.

Aussi utile et intéressante soit-elle, Marc Coeckelbergh estime que cette position néglige la question majeure de savoir ce que les robots nous font à travers les IHM. Le philosophe propose donc d'enrichir la roboéthique d'un axe phénoménologique fondée sur l'analyse de l'apparence des sociobots et des interactions qu'elle engendre.

Dans ce chapitre, nous proposons de classer et de passer en revue les grands thèmes de l'éthique des sociobots à partir des quatre perspectives prépositionnelles que nous avons dégagées, à savoir l'éthique « envers », « autour », « dans » et, enfin, « avec » les robots.

3.1 L'éthique *envers* les robots ou la question de leur statut moral

Ce que nous qualifions d'éthique *envers* les robots concerne l'ensemble des discussions dédiées à la question de leur considérabilité morale. Les sociobots sont-ils des êtres moraux ? Autrement dit, doit-on les intégrer à la communauté morale et, si oui, sous quelle forme et selon quel(s) critère(s) ? En pratique, ces questions visent le problème très concret de la régulation des comportements et de l'attribution de responsabilité, tant du point de vue des robots envers nous que du nôtre envers eux. La littérature distingue et examine trois grands statuts possibles pour les robots en général et les sociobots en particulier.

1. L'instrumentalité : les robots ne sont pas des êtres moraux, mais de simples instruments comme les autres. Nous sommes donc responsables d'eux, mais pas envers eux.
2. La patientalité : les robots sont des patients moraux. Nous sommes donc responsables envers eux.
3. L'agentivité : les robots sont des agents moraux. Nous sommes donc responsables envers eux et eux envers nous.

Examinons, dans l'ordre évoqué, chacun de ces points.

3.1.1 Le sociobot comme instrument

La première position, que nous qualifions d'instrumentaliste, consiste à pointer l'impossibilité catégorique et/ou l'indésirabilité morale d'un statut moral artefactuel. Dans « *Robot should*

be slaves », un article régulièrement cité sur ces questions, Joanna Bryson (2009) offre une argumentation exemplaire de ce point de vue. Son objectif est de trouver ce qu'elle nomme la « métaphore correcte » pour décrire nos relations avec les robots. Elle distingue pour cela deux sens du mot « correct » : ce qui est *exacte* (ontologiquement) d'un côté et ce qui est *approprié* (moralement) de l'autre. Sur la base de cette distinction, Bryson propose de qualifier les robots de « serviteurs » (ou d'« esclaves ») au sens d'« objets subordonnés à nos propres objectifs et conçus dans l'intention d'améliorer nos vies. » Ce terme est, selon elle, à la fois conceptuellement et moralement correct (exacte et approprié). Penser les robots comme des servants permettrait ainsi d'éviter les fautes morales liées à des erreurs de catégories.

« Il ne fait aucun doute que nous possédons les robots. Nous concevons, manufacturons, détenons et pilotons les robots. Nous en sommes entièrement responsables. Nous déterminons leurs objectifs et comportements, soit directement soit indirectement à travers la spécification de leur intelligence, ou plus indirectement encore en spécifiant la façon dont ils acquièrent leur intelligence. Mais quoi qu'il en soit, une vérité demeure : il n'y aurait pas de robot sur cette planète si l'humain ne décidait pas délibérément d'en créer » (Bryson, 2009).

Et la chercheuse de poursuivre :

« Étant acquis que nous possédons inévitablement les robots, négliger qu'ils sont essentiellement à notre service serait malsain et inefficace. Plus important encore, cela inciterait à prendre des décisions inappropriées telles que la mé attribution de responsabilité ou la mauvaise allocation de ressources » (*ibid.*).

Les choses sont claires : pour Bryson les robots sont ontologiquement des objets et, plus précisément, des outils comme les autres. Dès lors, « ils devraient être construits, commercialisés et légalement considérés comme des esclaves, pas comme des compagnons. » Le fait qu'ils puissent, à certains égards, nous ressembler (ou mimer d'autres vivants) et stimuler notre empathie explique notre tendance à les humaniser ou, plus largement, à les anim(alis)er, mais ce sont là des élans métaphoriques qui n'ont aucune valeur catégorique : un robot, quelles que soient son apparence et ses fonctionnalités, demeure un instrument au service de l'humain.

Passant du concept au précepte, Bryson cherche ensuite à ancrer moralement son argumentation en analysant la sur-identification aux sociobots à l'aune d'une analyse coûts/bénéfices et ce à l'échelle individuelle puis institutionnelle. Sur le plan personnel, la chercheuse fixe trois critères/questions d'évaluation : 1) la quantité absolue de temps et de ressources qu'un individu alloue à un compagnon artificiel ; 2) les activités que cet individu sacrifie pour réaliser cette allocation ; 3) le bénéfice de cet engagement envers un compagnon artificiel surpasse-t-il le coût ou le bénéfice (pour l'individu et/ou tout autre personne pouvant être impliquée) des alternatives négligées ? En somme, Bryson propose un raisonnement de type conséquentialiste dont l'axiome principal

« découle du credo libéral-progressiste par défaut selon lequel le temps et l'attention de chaque être humain sont des biens précieux qui ne devraient pas être gaspillés pour des choses insignifiantes » (*ibid.*).

Quelles sont, dès lors, les choses significatives qui méritent notre temps et notre attention ?
Les relations sociales entre humains, selon Bryson.

« [Son] inquiétude est que les humains n'ont qu'une quantité finie de temps pour tisser des relations sociales [...], et que nous trouvons de plus en plus de moyens de combler ce besoin par des divertissements improductifs et faussement sociaux » (*ibid.*).

Bryson ne dit pas que la présence et les interactions avec un sociobot n'ont aucun intérêt, surtout lorsque ce dernier intervient auprès de personnes isolées socialement, mais que les IHM ne sauraient en aucun cas se substituer aux relations entre humains. Or, la sur-identification aux robots porte en elle les germes d'une telle substitution.

« Au niveau personnel, le risque est de préférer des interactions sociales moins riches et moins épanouissantes avec des robots au motif qu'elles sont plus prévisibles et moins risquées que les relations humaines. Si vous faites quelque chose de stupide devant un robot, vous pouvez l'effacer de sa mémoire. Si vous êtes lassé d'un robot ou que vous voulez simplement vous coucher tôt, vous n'avez pas besoin de son opinion, vous pouvez simplement l'éteindre » (*ibid.*).

L'argumentaire est ici le même que celui mobilisé par Sherry Turkle (2007,2015) à propos de l'authenticité des relations humaines et de l'inauthenticité des IHM.

Sur le plan institutionnel, l'argumentation précédente reste valable. Sur le principe, Bryson juge indésirable qu'un gouvernement ou qu'une institution quelconque alloue du temps, de l'argent ou tout autre ressource aux robots alors qu'elles pourraient bénéficier directement aux humains et à leurs interactions. Ce niveau institutionnel comporte toutefois une problématique supplémentaire : celle de l'attribution de responsabilité légale et morale qu'une sur-identification aux robots pourrait engendrer. Bryson craint que les financements massifs dédiés aux « systèmes cognitifs autonomes » légitiment les erreurs et les fautes morales correspondantes. Elle pense notamment aux recherches en *machine ethics* portées par des universitaires comme Michael Anderson, Susan-Leigh Anderson, Wendell Wallach et Colin Allen dont l'objectif est de permettre aux robots, par voie de modélisation algorithmique, de prendre des décisions morales. Le problème n'est pas que des machines aident l'humain à décider, mais plutôt l'idée selon laquelle elles prendraient *elles-mêmes* des décisions, *a fortiori* morales. Cette attribution d'agentivité risque d'accroître les erreurs conceptuelles et l'aveuglement éthique en invitant à considérer les robots comme des personnes morales, c'est-à-dire comme des êtres à la fois moralement et légalement responsables de leurs actions. Nous l'avons dit, Bryson pense les robots comme de simples outils et ne voit donc aucune raison conceptuellement valable et moralement souhaitable pour qu'ils bénéficient d'une promotion ontologique leur octroyant un quelconque statut moral. Une telle promotion serait synonyme d'« abrogation de notre responsabilité » à l'égard de ces machines que nous construisons, programmons, vendons, utilisons, bref que nous possédons. De ce point de vue, l'humanisation des machines représente une déshumanisation de l'humain.

La position de Bryson peut être qualifiée d'utilitariste au double sens ontologique et moral du terme en ce qu'elle conçoit le robot comme un instrument essentiellement structuré pour « améliorer nos vies » en nous étant utile. L'utilité d'un sociobot peut être celle d'un artefact qui, par l'intermédiaire d'ISHM, aide à (re)socialiser l'humain, mais son essence n'en demeure pas moins celle d'un outil. « Les constructeurs de robots, affirme l'universitaire, sont éthiquement obligés – obligés de concevoir des robots qui ne nous obligent pas moralement. » Notre responsabilité ne s'exerce donc pas *envers* les robots, mais *à travers* eux, dans nos usages, nos mots et, plus largement, dans les intentions qu'ils concrétisent.

3.1.2 Le sociobot comme patient moral

La seconde position consiste à attribuer au robot le statut de patient moral au sens d'entités que l'on valorise et protège moralement en vertu du fait qu'elles peuvent être affectées par l'action des agents. L'enjeu consiste alors à préciser ce que nous entendons par « être affecté » afin d'identifier le ou les critères dont dépend cette patientalité. Nous proposons de passer ces critères en revue selon deux grandes catégories que nous nommerons « patientalité morale intrinsèque » et « patientalité morale extrinsèque ». Alors que la première cherche à déterminer l'existence de propriétés effectivement possédées par le potentiel patient moral (des propriétés catégorique et/ou dispositionnelles, donc), la seconde se concentre sur ces propriétés en tant qu'elles sont *apparentes*. La patientalité morale intrinsèque relève donc du registre ontologique (ce qu'une entité *est* en vertu de ses capacités) alors que la patientalité morale extrinsèque, elle, est avant tout de nature phénoménologique (elle procède, comme nous le verrons, de l'apparaître et de notre expérience de l'entité en question).

3.1.2.1 La patientalité morale intrinsèque

Dans un article très récent, Henry Shevlin propose une liste de cinq indices pouvant conduire à attribuer aux robots ce qu'il nomme la « patientalité morale psychologique » (*psychological moral patiency*) et qu'il définit comme « une forme de statut moral pouvant procéder de la possession de capacités psychologiques spécifiques telles que la sentience, l'autonomie, le désir etc. » (Shevlin, 2021, p.460). Sur ces cinq critères, trois relèvent de propriétés intrinsèques à l'entité. Il s'agit 1) de l'intelligence, 2) du désir autonome, et 3) de la sentience. Précisons, avant d'aller plus loin, que cette liste se présente comme une transposition, dans la roboéthique, de critères habituellement associés au vivant. Néanmoins, Shevlin n'affirme pas que les robots possèdent effectivement ces capacités, ni même qu'ils s'en approchent. Ces critères sont présentés à titre exploratoire afin de préparer la réflexion morale en vue d'un futur où les robots pourraient éventuellement faire preuve de telles propriétés. Un tel futur, si tant est qu'il puisse se réaliser, nous conduirait alors à repenser en profondeur nos conceptions du vivant (ce que fait déjà Lestel, mais du point de vue existentiel).

L'intelligence

Si l'essor de l'intelligence artificielle et des sciences cognitives a contribué à faire de l'intelligence une notion aujourd'hui omniprésente, celle-ci ne faisait pas partie du répertoire conceptuel classique de la philosophie avant le XXe siècle. On parlait alors plus volontiers de *logos*, de raison ou encore de pensée. Sans soutenir que tous ces termes, l'intelligence comprise, s'équivalent conceptuellement, on peut affirmer qu'ils étaient traditionnellement attribués de façon exclusive à l'humain et participaient ainsi d'une morale anthropocentrée fondée sur notre supériorité ontologique. Depuis le milieu du XXe siècle et l'émergence de champs de recherche tels que la cybernétique, l'IA, l'éthologie et l'éthique animale, ces exclusivité et supériorité ontologique sont clairement remise en cause. La question de savoir si les animaux sont intelligents ne fait plus tant débat que celle de savoir de *quelles* intelligences il s'agit. Quant aux machines, on ne compte plus, depuis le fameux test de Turing, le nombre de travaux et de réflexions dédiées à la possibilité ou à l'impossibilité d'une intelligence « artificielle ». Désanthropocentrée et inclusive au point d'envisager de l'attribuer à certains artefacts, l'intelligence est donc aujourd'hui considérée comme un critère de respectabilité morale interspécifique (entre espèces vivantes) voire « interontologique » (entre vivants et non-vivants). Ce désinvestissement anthropologique de l'intelligence s'observe également au niveau moral car, comme le remarque Shevlin, celle-ci est de plus en plus considérée comme une condition suffisante, mais pas nécessaire à la patientalité. En témoigne la valeur morale que nous accordons à nombre d'animaux dont nous jugeons qu'ils ne possèdent pas nos capacités cognitives. Notre position à l'égard du poids moral de l'intelligence n'en reste pas moins ambiguë étant donné notre propension valoriser d'avantage les animaux réputés « très intelligents » comme les dauphins, les chimpanzés ou encore les chiens que ceux dotés d'un système nerveux moins complexe. Enfin, outre qu'il apparaît comme suffisant mais pas nécessaire, le critère d'intelligence est aussi particulièrement difficile à définir. Nous sommes aujourd'hui loin de parvenir à élaborer notre appréhension intuitive du phénomène dans un concept à la fois clair, rigoureux et applicable à des systèmes très différents les uns des autres. Ainsi,

« toute évaluation de la patientalité morale psychologique dérivant d'une mesure fondée sur l'intelligence se confronte à la tâche peu enviable voire impossible de déterminer

comment comparer et synthétiser la myriade d'aptitudes cognitives dans une mesure unifiée de l'intelligence globale d'un système » (Shevlin, 2021, p.463).

Autonomie et désirs

Un second fondement possible pour la patientalité morale est celui de l'autonomie dont une entité peut faire preuve. S'appuyant sur un article de la philosophe américaine Erica L. Neely (2014), Shevlin aborde ce critère sous l'angle de la capacité à éprouver des désirs, par contraste avec des comportements déclenchés sur le mode mécanique de l'attraction/répulsion. Dans son article Neely appréhende l'autonomie de façon très générale comme la capacité de déterminer, au moins parfois, ses propres buts sans qu'intervienne une source extérieure (Neely et Shevlin semblent ici traiter concepts de « but » et de « désir » comme équivalents). Autonomie et désirs offrent une justification intéressante pour la patientalité car, dès lors qu'on admet la capacité d'autodétermination d'une entité, on admet également la possibilité de lui causer du tort en entravant cette capacité. Toute la difficulté est alors de parvenir à évaluer l'autonomie réelle de l'entité en question. Un robot programmé de façon non supervisée pour développer des préférences, comme par exemple se positionner au soleil plutôt qu'à l'ombre, fait-il preuve d'autonomie et de désir ? La solution à ce genre de problème consiste, selon Neely, à adopter une attitude probabiliste en confrontant nos critères aux indices émanant de telles ou telles entités et en effectuant des analogies. Par exemple, j'estime que mon chien a des désirs autonomes car un même jouet peut parfois susciter son intérêt et parfois non ou encore parce qu'il ne court pas systématiquement après le bâton que je lui lance. Au fond, le problème est le même que celui évoqué en philosophie de l'esprit à propos de la conscience, de la pensée ou encore des émotions.

« Parce que la conscience est une propriété attribuée à l'"esprit de l'autre", attester sa présence ou son absence demande d'accéder à quelque chose qui est et qui demeure inaccessible. "Comment peut-on déterminer, comme le demande très bien Paul Churchland, si quelque chose d'autre que soi-même – une créature extraterrestre, un robot sophistiqué, un ordinateur socialement actif ou même un autre humain – est vraiment un être pensant, sensible ou conscient plutôt que, par exemple, un automate inconscient dont les comportements émergent d'autre chose que d'authentiques états mentaux ?" » (Gunkel, 2014).

Shevlin note une autre limite de ce second critère : celle du décalage possible entre des êtres manifestant, même de façon superficielle, quelque chose comme des désirs autonomes et le peu sérieux que nous accorderions à leur promotion morale. Il cite par exemple les comportements aversifs des nématodes, la chimiotaxie des bactéries ou encore la phototaxie des plantes. Le problème, dès lors, ne concerne pas seulement la définition du critère, mais aussi son insuffisance.

La sentience

Le troisième et dernier critère d'une patientalité morale intrinsèque est celui de la sentience définie comme capacité d'un être à éprouver des expériences telles que le plaisir ou la douleur et, ce faisant, à rechercher les premières comme un bien et à fuir les secondes comme un mal⁷⁴. Depuis l'émergence de l'éthique animale dans les années 70, la sentience et, plus particulièrement, la disposition à la souffrance, sont peu à peu devenues des critères relativement consensuels de patientalité morale. Malgré leur évidence théorique, ces propriétés se heurtent toutefois à la difficulté pratique d'attester leur présence avec certitude chez telle ou telle entité. Il s'agit là, note Shevlin, d'un intense débat en psychologie comparée. Si certains philosophes et scientifiques étendent la sentience à l'ensemble des vertébrés et à quelques invertébrés, d'autres, par exemple, remettent en question la capacité des poissons à ressentir consciemment la douleur (Key, 2015). En outre, le problème se révélerait encore plus périlleux face à une hypothétique sentience artefactuelle.

« Dans le cas des animaux, nous pouvons au moins en appeler à des similarités générales dans l'architecture cognitive et la composition matérielle ainsi qu'à une histoire de l'évolution partagée. Par contre, les systèmes artificiels possèdent – et continuerons de posséder – des formes d'architecture cognitive et des substrats de traitement largement différents de ceux que l'on trouve chez l'humain. » (Shevlin, 2021, p.464)

Une dernière difficulté est soulignée par Neely (2014) à propos de l'analgésie congénitale. Un humain atteint d'une telle maladie et qui, par conséquent, serait incapable de vivre consciemment une expérience de douleur, aurait-il une valeur morale moindre qu'un humain

⁷⁴ Dans les termes de la psychologie contemporaine, précise Shevlin, la sentience se décrit comme la capacité d'éprouver des états de conscience à la valence positive ou négative.

épargné par cette pathologie ? La solution à un tel problème peut consister à adopter une critériologie cumulative hiérarchisée et, par exemple, à se reporter sur la propriété du désir autonome si la sensibilité à la douleur est absente. Or, en cumulant les avantages de chaque critère, cette solution en cumulerait également les difficultés. Se poserait également la question de savoir qui déciderait de la hiérarchie des propriétés et comment. Enfin, le cas de l'humain analgésique renvoie aussi à la possibilité que l'un des principaux critères effectivement mobilisés dans nos valorisations morales soit celui de la proximité empathique et, plus largement, phénoménologique (apparences, émotions, comportements) avec le potentiel patient. Cette remarque renvoie cependant à une conception extrinsèque de la patientalité que nous allons maintenant détailler.

3.1.2.2 La patientalité morale extrinsèque

Nous rattachons les deux autres critères évoqués par Shevlin à la patientalité morale extrinsèque car ils dépendent surtout de l'effet que le patient potentiel a sur l'observateur. Il s'agit 1) de l'intuition et de l'empathie et 2) de l'équivalence comportementale.

Intuition et empathie

Ce critère, comme son nom l'indique, est peut-être le plus intuitif et le plus évident de prime abord. Il consiste à attribuer la personnalité morale aux entités dont l'apparence, le comportement ou tout autre indice phénoménologique active une réponse empathique chez l'humain. Plusieurs études expérimentales ont récemment confirmé ce que nous apprenions empiriquement nos interactions avec les robots, à savoir que nous entrons en empathie avec les non-humains artificiels lorsque nous percevons en eux quelque chose comme de la souffrance (Rosenthal *et al*, 2013 ; Suzuki *et al*, 2015). Autrement dit, l'empathie n'est pas une affaire d'ontologie, mais de phénoménologie (Gunkel, 2016). Cette approche empathique de la patientalité est notamment adoptée par Kate Darling (2016) dans une réflexion dédiée à l'extension de la protection légale aux robots. A première vue, explique-t-elle,

« il semble difficile de distinguer un robot social, tel que le jouet dinosaure Pleo, d'un appareil électroménager comme un grille-pain. Tous deux sont des objets créés par l'humain pouvant être achetés sur Amazon et utilisés comme nous le voulons. Pourtant,

il y a une différence dans la façon dont nous percevons ces deux artefacts. Alors que les grille-pains sont conçus pour faire du pain grillé, les robots sociaux, eux, sont conçus pour faire office de compagnons » (*ibid.*).

Cette différence de perception du robot procède, résume Darling, du phénomène d'anthropomorphisme et de ses trois grands principes, à savoir la physicalité, l'apparence de mouvement autonome et le comportement social. Comme nous l'avons vu, le design oriente bien souvent ce phénomène vers la *sollicitation active* de notre empathie et de nos capacités d'attachement par le sociobot⁷⁵. C'est sur la base de ces capacités à nous émouvoir, à nous attendrir ou encore à susciter notre compassion que Darling propose de fonder un droit des robots. Toutefois, son idée n'est pas tant de protéger les robots eux-mêmes que de cultiver et de respecter notre propre moralité par leur intermédiaire. La chercheuse mobilise ici le célèbre argument kantien sur le bon traitement des animaux.

« L'argument kantien pour prévenir la cruauté envers les animaux est que nos actions envers les non-humains reflètent notre moralité – si nous traitons les animaux de façon inhumaine, nous devenons des personnes inhumaines. Cela s'étend logiquement au traitement des compagnons robotiques. Accorder la protection [légale] aux robots peut avoir pour effet de renforcer les comportements qui, en général, nous semblent moralement corrects ou, au moins, les comportements qui rendent notre cohabitation plus agréable. Cela peut aussi prévenir la désensibilisation envers les vraies créatures vivantes et protéger l'empathie que nous avons les uns pour les autres » (*ibid.*).

Malgré son ancrage kantien et son affiliation au déontologisme, cet argumentaire pourrait également être déployé dans le cadre de l'éthique des vertus dont l'une des prescriptions centrales n'est autre que le renforcement des comportements vertueux par la pratique quotidienne des vertus. Quoi qu'il en soit, la patientalité morale qu'imagine Darling ne découle pas d'une quelconque valeur intrinsèque des sociobots – ceux-ci sont loin, selon elle, de posséder *véritablement*, c'est-à-dire au-delà des apparences, les propriétés nécessaires pour cela –, mais d'une valeur extrinsèque fondée sur leur participation potentielle au respect de soi et des autres. En définitive, la patientalité dont parle Darling ne relève pas tant d'un statut moral du robot que d'un statut instrumental mis au service d'une éthique expressément

⁷⁵ C'est d'ailleurs cette activité qui le distingue spécifiquement car le phénomène d'attachement aux artefacts, lui, existe depuis toujours.

anthropocentrée. Précisons que toutes les tentatives d'attribution de patientalité au robot sur la base de l'empathie ne relèvent pas nécessairement de cette logique. La stratégie probabiliste évoquée plus haut peut aussi être appliquée pour ce critère. Le raisonnement est alors le suivant : n'ayant pas la possibilité de déterminer avec certitude que mon empathie correspond à une expérience réellement vécue par telle ou telle entité (autrement dit, qu'il existe une correspondance entre la manifestation extérieure d'un état psychologique et sa manifestation intérieure voire qu'il existe un état intérieur sans manifestation extérieure), j'évalue le degré de probabilité que cette expérience soit véritablement vécue sur la base d'analogies avec mes propres réactions ou celles d'entités dont il est admis qu'elles sont douées de sensibilité. Le caractère plus ou moins scientifique de cette approche dépend, bien sûr, du degré d'explicitation et de rigueur des critères qui président aux analogies tissées. Comme dans les cas précédents, l'attribution de patientalité sur la base de l'intuition ou de l'empathie n'est pas sans poser problème. Tout d'abord, fonder un statut moral sur la base de l'empathie (méthodiquement ressaisie ou non) comportera toujours le risque de faux-positifs ou de faux-négatifs. Ce faisant, elle peut conduire à faire preuve de largesses morales infondées dans le premier cas ou, inversement, à se montrer trop restrictif dans le second. Ensuite, notons que l'empathie est une mécanique cognitive à la fois « myope et partielle »⁷⁶ comme l'affirment Laurence Kaufmann et Olivier Glassey, à la suite du psychologue canadien Paul Bloom.

« Elle est myope, car elle n'a aucun sens de la proportion statistique ; elle peut réagir immédiatement à la perception d'une détresse singulière, mais rester indifférente à la souffrance d'innombrables êtres lointains et sans visage [...]. L'empathie est aussi partielle et sélective, car elle répond à une heuristique du semblable et ne s'active qu'à la perception d'un autrui perçu comme un "même" » (Kaufmann & Glassey, 2020, p.4).

Entrons quelque peu dans le détail des arguments mobilisés par les deux universitaires. Le problème des évaluations morales fondées sur l'empathie vient de ce qu'elles relèvent d'une « morale de proximité » et ce de trois grandes façons (*ibid.*). Premièrement, en tant qu'elle repose sur la sensibilité, l'empathie ne s'active pleinement qu'*in situ*, c'est-à-dire au sein d'expériences vécues. C'est ce qui explique que nous sommes plus enclins à entrer en

⁷⁶ Laurence Kaufmann et Olivier Glassey (2020) citant l'essai du psychologue canadien Paul Bloom, *Against Empathy: the Case for Rational Compassion*.

empathie avec une souffrance singulière que nous avons sous les yeux, plutôt qu'avec des tragédies de masse, comme par exemple des famines ou des catastrophes naturelles, qui sévissent à l'autre bout du monde et auxquelles nous n'avons accès que par une connaissance désincarnée (i.e. « l'information »). Deuxièmement, et cela découle de ce premier point, l'empathie se veut également situationnelle en ceci qu'elle ne pose pas tant la question du « que dois-je faire » que celle, toujours située, du « que dois-je faire dans cette relation-ci ? » En d'autres termes, les évaluations morales fondées sur l'empathie interrogent toujours quant à la possibilité de leur universalisation et donc de leur institutionnalisation. A ce caractère situé s'ajoute par ailleurs une forme de relativisme lié aux plus ou moins grandes variations de capacités empathiques d'un individu à l'autre (allant du degré zéro qu'est la psychopathie aux formes très développées d'empathie qu'on trouve généralement chez les personnes hypersensibles) et donc conduire à des évaluations morales très différentes d'un individu à l'autre. Enfin, troisièmement, l'empathie est aussi une « morale de proximité » au sens où elle relève de ce que Kaufmann et Glassey nomment une « phénoménologie du semblable » où ce semblable, comme le mot l'indique, est perçu et donc saisi comme tel par rapport à un *moi* individuel lui-même toujours situé socialement et existentiellement. La recherche en neurosciences et en psychologie démontre par exemple que l'activité empathique est sensible aux catégories et aux appartenances sociales (Decety & Cowell, 2014). En d'autres termes, le semblable de l'empathie n'est pas d'abord la figure abstraite et universelle du prochain judéo-chrétien, mais celle du prochain au sens social et/ou spatial du terme. Dès lors, toute morale fondée sur une phénoménologie empathique court le risque de la « tribalisation », affirment Kaufmann et Glassey. Pour résumer, l'empathie, en tant qu'elle est une morale de proximité,

« permet de saisir *in situ* la manière dont tel ou tel être particulier *devrait* être traité dans telle ou telle situation en fonction de son groupe d'appartenance, de son rang social et, surtout, de son degré de proximité ou de distance relative avec *ego*. Lorsque cette distance, spatiale ou sociale, est trop grande, elle devient synonyme de cécité émotionnelle et de torpeur morale » (Kaufmann & Glassey, 2020, p.4).

Dans la perspective développée par Shevlin, le risque d'une attribution de patientalité fondée sur l'empathie est donc celui de l'incapacité à reconnaître des prétendants légitimes quoiqu'artefactuels en raison de leur trop grande étrangeté phénoménologique. Cela dit, si l'on se fie aux principes qui sont actuellement les siens, la robotique sociale semble plutôt à

l'abri d'un tel risque (très prospectif, rappelons-le) puisque, nous l'avons vu, l'empathie humaine est convoquée par le design en tant qu'élément central du fonctionnement de l'artefact en tant *socio-bot*. C'est d'ailleurs ce qui inquiète Kaufman et Glassey pour qui l'empathie envers les robots n'est pas qu'« une simple extension de notre concernement émotionnel » (*ibid.*, p.5) à de nouveaux semblables, mais également un risque d'aveuglement quant aux foules invisibles (parce que trop éloignées) de semblables humains qui auraient véritablement besoin des effets de notre empathie, mais qui ne bénéficient pas des effets d'annonce de la robotique.

L'équivalence comportementale

Le cinquième et dernier critère évoqué par Shevlin est celui de l'équivalence comportementale (*behavioral equivalence*) qu'il emprunte à John Danaher (2020). Il consiste, comme son nom l'indique, à fonder l'attribution de la patientalité morale sur l'observation des comportements d'une entité. Danaher affirme en effet que le comportement d'un être n'est autre que l'indice ou la preuve principale à partir de laquelle nous attribuons ou non la considérabilité morale. Cette preuve peut intervenir sur le mode inductif ou déductif. Dans le premier cas, l'observateur découvre la propriété qui le conduit à valoriser moralement l'entité, par exemple la sentience, à partir du comportement observé. Dans le second cas, le critère de sentience préexiste à l'observation du comportement qui vient en confirmer ou non la présence chez l'entité en question. Dans un cas comme dans l'autre, le cœur de l'argument est le suivant : si une entité affiche un comportement correspondant à un critère de patientalité morale alors cette entité est un patient moral, qu'il soit humain ou non humain, vivant ou non-vivant. Une telle démarche offre le double avantage d'être immédiatement applicable (sans être tributaire du développement des sciences cognitives) et d'éviter l'écueil subjectiviste de l'approche empathique en se fondant sur l'observation objective des comportements. Elle présente toutefois la difficulté majeure de déterminer ce que Danaher nomme le « seuil performatif » (*performative treshold*) à partir duquel la patientalité doit être attribuée. Prenons l'exemple des célèbres « tortues de Bristol », Elmer et Elsie, développées par William Grey Walter au mitan du XXe siècle. Elles étaient conçues pour tourner autour des sources de lumière de faible intensité et pour fuir celles d'intensité trop vive. Cette répulsion se transformait toutefois en attraction lorsque les batteries des tortues étaient déchargées

pour qu'elles aillent puiser l'énergie nécessaire à leur rechargement. Du point de vue de l'équivalence comportementale, si le critère retenu pour la patientalité est, par exemple, celui d'avoir des intérêts (Singer, 1975 ; Neely, 2014) au sens d'inclinations orientées vers la persévérance dans l'être ou de polarisation comportementale (plaisir/peine, attirance/répugnance), Elmer et Elsie devraient être considérées comme des patients moraux et ce malgré leur peu de complexité. Selon cette perspective, n'importe quel ingénieur ou amateur un tant soit peu qualifié pourrait concevoir des patients moraux artificiels et réclamer des formes de protections légales pour leurs créatures. Là encore, l'enjeu concerne donc le degré d'inclusivité que nous souhaitons donner à la communauté morale. Enfin, le principe d'équivalence comportementale pose également la difficulté de l'identification de la patientalité lorsqu'aucune similitude ne peut être détectée ou quand la réaction comportementale d'une entité consiste précisément à ne rien manifester « extérieurement ».

Notons que ces questions liées à l'attribution de patientalité morale rappellent, dans une large mesure, les partis pris évoqués plus haut au sujet de la robotique externe et interne des émotions. En effet, la patientalité intrinsèque cherche à déterminer l'existence d'une ou de plusieurs propriétés internes à l'entité examinée alors que la patientalité extrinsèque se concentre sur les manifestations phénoménales de ces propriétés. Dans le point consacré à l'éthique *avec le robot*, nous verrons qu'à l'instar du bouclage affectif évoqué par Dumouchel et Damiano, certains penseurs cherchent à penser la considérabilité morale comme un processus relationnel dépassant les partis pris strictement internalistes et externalistes.

3.1.3 Le sociobot comme agent moral

La troisième et dernière position étudie la possibilité de considérer les robots comme des agents moraux artificiels (AMA). Comme pour la patientalité, l'agentivité morale peut être pensée de façon intrinsèque ou extrinsèque à l'artefact. L'AMA intrinsèque correspond à la « justification standard » de l'agentivité morale (Himma, 2009) alors que l'AMA extrinsèque renvoie à sa justification non-standard.

3.1.3.1 L'agentivité morale intrinsèque : conscience, intentionnalité et responsabilité

La pensée philosophique classique fait dépendre l'agentivité morale de l'agentivité libre (*free agency*) et de l'agentivité rationnelle (*rational agency*) qui, elles-mêmes, dépendent de l'agentivité *tout court*. Le philosophe américain Kenneth Einar Himma propose une description extrêmement claire et méthodique de ce raisonnement déductif dont la conclusion consacre la conscience comme condition de possibilité de l'agentivité morale. Son point de départ est le suivant : l'agentivité, entendue comme fait d'être un agent, peut se définir comme la capacité d'agir, de réaliser des *actions*. Les actions représentent un type bien particulier d'activité (*doings*) puisqu'elles procèdent d'*intentions*, c'est-à-dire d'états mentaux tels que le désir ou, plus largement, la volonté. Les conditions de l'agentivité peuvent donc, selon Himma, se résumer ainsi : « X est un agent si et seulement si X peut faire preuve d'états mentaux intentionnels capables de causer directement une performance » (Himma, 2009, p.21). Or, l'agentivité ne se décline pas nécessairement en agentivité *morale*. Pour être moral, un agent doit d'abord pouvoir être considéré comme responsable de ses actes. Une telle responsabilité dépend elle-même de deux capacités nécessaires et conjointement suffisantes : la capacité de choisir librement ses actes et celle de distinguer le bien du mal, de l'autre.

Concernant le libre arbitre, Himma ne prétend pas trancher la question de sa nature ou de son fonctionnement. Il se concentre sur des remarques relativement générales et peu sujettes à controverse. Tout d'abord, on peut affirmer qu'un être choisit librement ses actes dès lors qu'ils ne lui sont pas imposés de l'extérieur. Autrement dit, l'agentivité libre suppose que l'agent soit la cause directe de ses actions (ce qui n'exclut aucunement les influences extérieures, précisons-le). Plus précisément, cette causalité doit relever d'une forme de décision de l'agent et, de ce fait, implique qu'il soit capable de délibération. Il ne s'agit pas d'affirmer que toute décision conduisant à un acte libre procède d'une délibération, mais plutôt que les actes libres résultent de décisions et que seuls des êtres capables de délibérer peuvent prendre des décisions⁷⁷. Un tel raisonnement aboutit à la conclusion que l'agentivité morale implique la rationalité de l'agent puisque seuls des êtres rationnels, c'est-à-dire des êtres capables de raisonner, peuvent faire preuve de délibération.

⁷⁷ En effet, une délibération passée peut engendrer une systématisation de telle ou telle décision et comportement correspondant, comme ne pas boire d'alcool en semaine par exemple, jusqu'à ce qu'une nouvelle délibération vienne modifier cette systématisation. Dans ce cas, le fait de ne pas boire d'alcool en semaine n'a fait l'objet que d'une seule délibération, mais reste un comportement *actif* fondé sur un libre choix.

La seconde grande capacité, celle de distinguer le bien du mal, est elle aussi rattachée à la rationalité puisqu'elle relève d'une forme de raisonnement, spécifiquement moral cette fois. Cela ne signifie pas qu'un agent, pour être moral, se doit d'identifier *systématiquement*⁷⁸ le bien et le mal, mais plutôt qu'il en est capable *de façon générale*. Cette aptitude peut donc être plus ou moins développée selon les individus, mais implique toujours la capacité, elle aussi générale, d'adapter son comportement aux principes moraux qui gouvernent une société. Schématiquement, tout raisonnement moral s'appuie sur quatre facultés : 1) la compréhension minimale des concepts moraux (« bien », « mal », « obligatoire », « autorisé », « interdit » etc.) ; 2) la connaissance et la hiérarchisation des principes moraux considérés comme absolument nécessaires pour vivre en société (par exemple, il est moralement plus impératif de ne pas tuer un innocent que de donner de l'argent à un SDF) ; 3) l'identification *in situ* des faits et gestes qui relèvent de tel ou tel principe et non de tel ou tel autre (par exemple, s'arrêter au feu rouge est une règle obéissant au principe moral de ne pas écraser les piétons ou, a minima, de ne pas causer d'accident, mais ne convoque pas celui consistant à se montrer poli en saluant lesdits piétons) et 4) l'induction ou la déduction des actes particuliers permis ou prohibés que recouvrent ces grandes règles morales (par exemple, s'empêcher de télécharger illégalement des contenus sur internet car cela relève du principe général de l'interdiction du vol).

Synthétisant chacun de ces points, Himma formule les conditions de l'agentivité morale comme suit : « pour tout X, X est un agent moral si et seulement si X est (1) un agent ayant la capacité de (2) choisir librement, (3) de délibérer à propos de ce qu'il devrait faire et (4) de comprendre et d'appliquer correctement les bonnes règles morales selon les cas » (*ibid.*, p.24). A l'issue de cette analyse formelle⁷⁹, le philosophe remonte le fil de son raisonnement et affirme que chaque condition évoquée repose elle-même sur la condition qu'est la conscience. Autrement dit, il n'y pas d'agentivité morale ni d'agentivité tout court sans conscience préalable. Qu'il s'agisse d'un artefact ou de n'importe quelle autre entité, le résultat reste le même : « [un artefact] peut être un agent moral si et seulement si il est un agent capable de choisir "librement" ses actions et de comprendre les concepts et prérequis de base de la moralité, ce qui présuppose également la conscience » (*ibid.*, p.28). Une telle

⁷⁸ Une telle systématisme supposerait que l'agent soit moralement infaillible et/ou qu'il existe quelque chose comme une objectivité absolue du bien et du mal.

⁷⁹ Au sens où Himma ne rentre pas dans l'analyse du contenu des concepts qu'il mentionne en tant que conditions de possibilité de l'agentivité morale.

conclusion ne cherche donc pas à nier ou à affirmer la possibilité d'un AMA, mais seulement à en exposer les conditions d'existence. Himma n'en reconnaît pas moins l'énorme difficulté qu'il y aurait à concevoir des artefacts conscients et dotés de libre arbitre (à commencer parce que nous sommes loin de comprendre exactement et d'expliquer de telles capacités). Dès lors, il semble raisonnable d'affirmer qu'à l'heure actuelle, le modèle philosophique standard ne permet de considérer les sociobots (et autres artefacts dits « intelligents ») ni comme des agents moraux ni comme des agents tout court.

3.1.3.2 L'agentivité morale extrinsèque : les positions relativiste et relationnelle

Un compte rendu non-standard de l'AMA nous est donné par le philosophe italien Luciano Floridi qui, dans le cadre de ses travaux en éthique de l'information, développe une réflexion sur la « moralité sans esprit » (*mind-less morality*) (Floridi & Sanders, 2004) également qualifiable de « moralité a-responsable » (*a-responsible morality*) (Floridi, 2014). Celui-ci constate que, contrairement à la patientalité, l'agentivité morale n'a pas fait l'objet de réflexions visant à repenser ses fondements et, éventuellement, à la rendre plus inclusive. Le seul exemple d'extension d'agentivité et de responsabilités correspondantes semble être celui des personnes morales comme les entreprises. Toutefois, cette extension reste empreinte d'une conception anthropocentrique de l'agentivité puisqu'elle ne s'applique qu'à des collectifs d'humains, consacrant ces derniers comme les seules sources possibles d'actions moralement responsables. En dernière instance, l'agentivité ne renvoie qu'à des personnes physiques ou morales, c'est-à-dire à des formes d'individualités humaines⁸⁰. Or, note Floridi,

« [I]imiter le discours éthique au niveau *individuel* entrave le développement d'une enquête satisfaisante sur la moralité distribuée, [entendue comme] phénomène macroscopique et grandissant d'actions morales globales et de responsabilités collectives résultant de la "main invisible" des interactions systémiques entre plusieurs agents à échelle locale » (*ibid.*, p.187).

⁸⁰ Notons qu'il existe des cas d'attribution de « personnalité juridique » à des entités naturelles. C'est le cas, par exemple, de certains fleuves comme le Gange Inde ou le Whanganui en Nouvelle-Zélande. Ces attributions, qui reconnaissent le caractère vivant desdites entités, relève toutefois du régime moral de la patientalité et non de l'agentivité.

Dès lors, son objectif n'est pas de déresponsabiliser l'humain, mais de mieux rendre compte de la complexité des phénomènes moraux par un double geste de désindividualisation et de désanthropisation de l'agentivité et de la responsabilité morale.

En substance, Floridi explique que la possibilité d'attribuer l'agentivité morale aux artefacts – comme toute autre analyse d'ordre ontologique d'ailleurs – dépend du « niveau d'abstraction » (*Level of Abstraction*) dont émane la réflexion. Un niveau d'abstraction ou NdA renvoie à « un ensemble fini et non nul de variables observables dont le choix même soutient la théorie qu'elles caractérisent » (*ibid.*, p191). Il s'agit donc de déterminer le niveau d'abstraction auquel il est légitime d'attribuer l'agentivité morale à un artefact. Comme Himma, Floridi note que, pour être un agent moral, il faut d'abord être un agent tout court. Il est donc nécessaire de déterminer ce qui constitue l'agentivité dans un NdA donné. On peut par exemple considérer qu'un agent est un système qui opère des transformations, produit des effets ou exerce un pouvoir quelconque dans son environnement. Or, Floridi constate qu'à ce NdA, il n'existe aucune différence entre un agent humain et un tremblement de terre, par exemple. Le NdA est donc trop élevé au sens où il abstrait trop de propriétés pour être pertinent. Pour le diminuer, l'auteur propose de prendre en compte trois critères souvent discutés dans la littérature, à savoir l'interactivité, l'autonomie et l'adaptabilité. L'interactivité renvoie à l'idée que l'agent et son environnement peuvent agir l'un sur l'autre. L'autonomie, quant à elle, désigne la capacité de l'agent à changer d'état sans que ce changement procède d'une interaction (ce qui, au passage, implique qu'un agent doive faire preuve d'au moins deux états différents). Enfin, l'adaptabilité signifie que les interactions de l'agent peuvent modifier ses « règles de transition » (*rules of transition*), c'est-à-dire les règles qui le font passer d'un état à un autre. Puisque l'attribution d'agentivité dépend du NdA adopté qui lui-même dépend des variables observables, une même entité peut être ou ne pas être un agent selon les variables convoquées dans l'analyse. Floridi prend, entre autres, l'exemple d'un « webbot » ayant pour fonction de filtrer les mails indésirables en apprenant les préférences de ses utilisateurs. Outre les critères susmentionnés, si le NdA de l'observateur comprend l'accès aux mails reçus (*input*) et aux mails filtrés (*output*) par le webbot, alors celui-ci semble remplir les critères de l'agentivité. En revanche, si le NdA est abaissé de telle sorte à inclure le code de l'artefact, l'adaptabilité dont il semblait faire preuve à l'égard des préférences utilisateurs disparaît derrière la nature déterministe de son algorithme.

L'étape suivante consiste, sans surprise, à examiner dans quelles conditions un agent donné peut être qualifié de moral. Floridi propose de considérer qu'une action est qualifiable moralement si et seulement si elle cause du bien ou du mal moral (*moral good/evil*). Un agent est donc moral dès lors qu'il est capable d'actions moralement qualifiables. L'auteur prend alors l'exemple de deux soignants, H et W, observé à travers une caméra de vidéosurveillance. Dans ce référentiel, le NdA permet d'observer qu'H et W font preuve d'interactivité, d'autonomie et d'adaptabilité. Supposons maintenant que H tue son patient et que W le soigne. Il s'agit bien là d'un acte moralement qualifiable. H et W répondent donc bien aux critères d'agentivité morale selon le NdA de référence. Or, nous apprenons par la suite que l'un d'entre eux est un humain et l'autre, peu importe lequel, est un agent artificiel. Cela change-t-il quelque chose à la conclusion précédente ? Dans ce NdA, aucunement selon Floridi. Pourtant, la moralité ne suppose-t-elle pas, comme nous l'avons vu avec Himma, une disposition intentionnelle et non simplement téléologique ? Autrement dit, dans cette situation, l'agent ne doit-il pas avoir, outre un objectif de soin, une *intention* de soigner ? Sans cela, comment faire la différence entre un programme informatique et une volonté à la fois consciente et librement déterminée ? Floridi rétorque que cette objection suppose un accès aux états mentaux de l'agent, ce qui renvoie au problème, déjà évoqué, des autres esprits. La position adoptée dans le NdA de référence est une position phénoménologique qui se concentre sur ce qui est observable de l'extérieur.

« Cette approche phénoménologique est une force, pas une faiblesse. Elle implique que les agents (humains compris) devraient être qualifiés de moraux dès lors qu'ils font effectivement partie du "jeu moral". La question de savoir s'ils ont l'intention d'y jouer n'est pertinente que dans un second temps, lorsque nous cherchons à savoir s'ils sont *moralement responsables* de leurs actions » (*ibid.*, p.201).

Selon Floridi, ce problème de la responsabilité constitue l'objection la plus sérieuse à son analyse. Elle souligne qu'un NdA qui n'ajouterait pas la responsabilité à la critériologie de l'agentivité morale serait incapable de distinguer les agents moraux d'une forme de cause efficiente certes complexe, mais non moins amoral (en l'état, le NdA serait donc composé de conditions nécessaires, mais insuffisantes). Floridi remarque toutefois qu'un tel problème se pose uniquement pour une pensée qui assimile totalement la morale à la responsabilité ou, autrement dit, qui conditionne la possibilité de la première à l'existence de la seconde. Or,

dans les faits, l'identification des situations et des actions comme relevant de la morale excède la question de la responsabilité. L'exemple le plus parlant est peut-être celui des parents qui, sans aller jusqu'à faire de leurs enfants mineurs des êtres moralement responsables, ne les considèrent pas moins comme des « sources d'actions morales » (*ibid.*, p.203). Ce partage entre simple causation et responsabilité est au cœur du propos de Floridi sur l'élargissement de l'agentivité. Le recours à la « Méthode de l'Abstraction » lui permet de découpler la « responsabilité morale » (*moral responsibility*) de ce qu'il nomme la « *moral accountability* » et que l'on peut traduire par comptabilité ou imputabilité morale.

« Un agent est moralement comptable de *x* si l'agent est la source de *x* et que *x* est qualifiable moralement [...]. Pour être aussi moralement responsable, l'agent doit témoigner de bons états intentionnels » (*ibid.*, p.106).

Le principal avantage d'une telle distinction, selon Floridi, est d'aboutir à une compréhension beaucoup plus fine de la morale en contexte sociotechnique et d'en saisir le caractère *distribué*. Une telle distribution implique, par exemple, de ne pas considérer la moralité ou l'immoralité d'un système donné uniquement à l'aune des individus humains qui le composent et de leurs agissements, mais aussi au regard des artefacts qui y participent sur le mode « comptable ».

Cette symétrisation relative entre agent humain et agent artificiel n'est pas sans rappeler les travaux de Bruno Latour sur l'agentivité. Toutefois, contrairement à Floridi, Latour n'inscrit pas sa pensée dans une ontologie relative (à un NdA), mais dans une ontologie relationnelle qui ne s'intéresse pas tant aux conditions de possibilité de l'agentivité qu'à son enchevêtrement nécessaire et permanent entre les êtres. C'est ce que nous avons proposé de désigner par le terme d'interactance au sens de compénétration des actances. Précisons que si Latour préfère le terme d'actant à celui d'agent, c'est parce qu'il se détache de la matrice de pensée moderne et signale mieux le projet de rupture avec une agentivité exclusivement détenue par la subjectivité et sa volonté. Pour rappel, toute entité produisant un effet quelconque sur une autre entité (ou participant à l'émergence d'une nouvelle entité) est considérée comme un actant. En d'autres termes, il n'y pas les agents/sujets d'un côté (quel que soit leur régime transcendantal) et les patients/objets de l'autre, mais des associations, dissociations, réassociations perpétuelles d'actants. Pour illustrer son approche, Latour donne notamment l'exemple d'une personne A tirant au pistolet sur une personne B et demande qui

a abattu B : A ou le pistolet⁸¹ ? La pensée moderne répond que le sujet coupable du crime est A, que le sujet victime du crime est B et que l'objet ou, plus précisément, le moyen utilisé pour commettre le crime est le pistolet. Dans cette logique, il est parfaitement irrationnel de considérer l'arme comme responsable de la mort de B puisqu'elle n'est qu'un objet, une chose inerte, un *intermédiaire* permettant de concrétiser l'intention meurtrière de A. Pourtant, selon Latour, cette partition du monde a plus pour effet de l'organiser artificiellement que d'en rendre compte empiriquement. Du point de vue de l'expérience, affirme-t-il, tout pousse à constater les limites des dichotomies modernes du type sujet/objet ou moyen/fin et à leur substituer une hybridation humains/non-humains. Ici, l'acte de tirer n'est à chercher ni du côté du pur sujet A ni du côté du pur objet pistolet, mais dans l'union des deux. Autrement dit, toute fusillade suppose l'association (actant 3) d'un tireur (actant 1) et d'une arme à feu (actant 2) dont la relation n'est jamais purement instrumentale. Latour généralise cette idée à l'ensemble des artefacts à travers le concept de *médiation technique* (1994a) que nous avons déjà évoqué à plusieurs reprises. Loin d'être de simples moyens ou intermédiaires, les artefacts participent activement à l'élaboration des situations existentielles comme des fins que nous poursuivons et ce à travers les quatre composantes fondamentales de toute médiation technique : la traduction, la composition, l'effet boîte-noire (*black-boxing*) et la délégation. La traduction désigne la transcription technique de ce que Latour nomme un « programme d'action » humain. Considérons un exemple simple : la présence d'une machine à café dans mon bureau peut, lorsque je décide de faire une pause, *traduire* mon programme d'action « faire une pause » en « prendre un café ». Le concept de traduction a ceci d'intéressant qu'il montre que la volonté ne se détermine jamais seule à agir, mais toujours avec autre chose qu'elle. L'homme agissant a/est toujours quelque chose de plus que lui-même, une composante technique qui donne corps à ses intentions en les co-constituant. Autrement dit, l'action est composite en ce qu'elle repose nécessairement sur des associations d'actants. C'est là le sens de la seconde dimension de toute médiation technique qu'est la composition. L'effet boîte-noire, quant à lui, désigne l'occultation de cette nature composite et traductrice par le fonctionnement et l'usage des artefacts. Enfin, la délégation renvoie au phénomène par lequel les actants s'assignent des responsabilités, des gestes, des

⁸¹ Exemple renvoyant au dépassement nécessaire des positions pro- et anti-armes aux États-Unis où les pros, représentées par la *National Rifle Association* (NRA), rétorquent aux antis que ce ne sont pas les armes qui tuent, mais les humains qui les utilisent.

comportements les uns les autres. L'exemple emblématique donné par Latour est celui du ralentisseur (Latour, 1992). Également nommé dos-d'âne ou « gendarme couché », le ralentisseur est un actant auquel on a *délégué* une certaine autorité⁸² : celle de veiller à l'application du code de la route par la contrainte matérielle. L'artefact se retrouve alors au milieu d'une chaîne de délégations successives : 1) les concepteurs matérialisent l'autorité publique dans le dos d'âne, 2) cette matérialisation est en même temps une délégation du pouvoir de faire respecter le code de la route, 3) par la contrainte matérielle, le dos d'âne assure cette bonne application en déléguant au conducteur le soin de ralentir sous peine d'endommager son véhicule.

A présent, en quoi la médiation technique ainsi expliquée est-elle en lien avec la question de l'agentivité morale ? Toute médiation technique, par le jeu de traduction/délégation dont relève son effectivité, est source de *prescriptions*, notamment morales. L'exemple du dos-d'âne reste très éclairant à cet égard. Le ralentissement qu'il impose renvoie, bien sûr, à la prescription légale qu'est le respect du code de la route, mais également à la prescription morale de la préservation de la vie d'autrui. Dès lors, en composant l'action, les artefacts composent aussi sa moralité.

Malgré leurs différences de cheminement analytique, Latour et Floridi partagent l'objectif de mieux appréhender l'action morale qui, en tant qu'elle procède d'une agentivité distribuée, implique une responsabilité elle aussi distribuée. Signalons tout de même que la position de Latour se veut bien plus radicale puisque, rompant avant la question des conditions de possibilité pour ne s'intéresser qu'aux médiations – principe fondateur de la postphénoménologie en tant qu'elle rompt avec la « tentation orphique⁸³ » de la philosophie des techniques (Verbeek, 2005) –, il suffit qu'un artefact compose d'une quelconque façon une situation morale pour qu'il se voit attribuer l'actance correspondante. A contrario, l'analyse de Floridi se fonde sur une liste de critères qui, malgré leur réélaboration par un relativisme méthodique, continuent d'appartenir à une ontologie moderne de l'agent. De ce point de vue, l'agentivité morale se présente comme une certaine forme d'actance morale, c'est-à-dire comme un type bien particulier de participation au « jeu moral » (*moral game*) pour reprendre une formule de Floridi.

⁸² Autorité à laquelle l'expression de « gendarme couché » renvoie très clairement.

⁸³ Verbeek utilise la figure d'Orphée comme symbole des approches transcendantales de la technique qui consistent à se retourner vers l'amont du phénomène, à regarder « en arrière » pour appréhender ses conditions de possibilité.

Bien qu'elles ne soient pas spécifiquement dédiées aux sociobots, ces deux analyses théoriques peuvent, bien sûr, leur être appliquées. Pour Floridi, tout robot social qui, sur la base des données observables, causera du bien ou du mal moral en faisant preuve d'interactivité, d'autonomie et d'adaptabilité, sera considéré comme un agent moral. Chez Latour, il suffira que celui-ci contribue à la configuration d'une situation morale pour endosser, momentanément ou non, le rôle d'actant moral. Notons que chacune de ces deux positions soulève des questions intéressantes quant au thème de l'altérité que nous traiterons en seconde partie. On peut par exemple se demander si le NdA de l'agentivité est une condition nécessaire et/ou suffisante de l'altérité. L'autre serait-il celui qui manifeste, dans un référentiel donné, de l'interactivité, de l'autonomie et de l'adaptabilité ? Notons, d'ailleurs, que ces trois critères rappellent l'autonomie sociale des substituts dont parlent Dumouchel et Damiano et qui, pour rappel, soutiennent que « [l]a caractéristique définitionnelle des "robots sociaux" est d'être perçus par ceux qui interagissent avec eux comme étant *présents* à la manière d'une personne, [...] de provoquer chez l'interlocuteur l'impression d'être face à un "autre" » (Dumouchel & Damiano, 2016, p.117-118). En fin de compte, l'autre serait-il plus précisément l'« autre agent » de sorte que l'altérité soit synonyme d'agentivité ? La perspective latourienne pose, quant à elle, une des questions structurantes de notre recherche, à savoir : quelle place l'altérité occupe-t-elle dans les médiations sociobotiques. Derrière la généricité du concept de médiation technique, n'est-elle pas ce qui singularise l'expérience la médiation sociobotique en tant que telle ? On le voit, c'est bien à un travail conceptuel sur l'altérité qu'appelle cette interrogation.

Après ces analyses dédiées à l'éthique *envers* le robot, passons à présent au second volet de notre découpage de la roboéthique : l'éthique *autour* du robot.

3.2 L'éthique *autour* du robot : conception, commercialisation et usages des sociobots

3.2.1 Éthique instrumentale et conformité axiologique des sociobots

Ce que nous qualifions ici d'éthique *autour* du robot concerne l'ensemble des développements techniques, des modes de commercialisation et de communication ainsi que des usages dont un sociobot peut faire l'objet. Cette frange de la réflexion éthique peut être

qualifiée d'instrumentale au sens où elle se concentre sur le pouvoir d'action du concepteur, du vendeur ou de l'utilisateur humain de l'artefact. Ce dernier ne participe au raisonnement que sur le mode objectal, c'est-à-dire en tant qu'objet totalement informé par la maîtrise humaine. En somme, la question est ici de savoir ce que nous devons faire pour concevoir, vendre et/ou utiliser éthiquement un robot. A notre connaissance, la plupart des travaux universitaires en roboéthique concernent le domaine du soin et sont donc en dialogue plus ou moins direct avec l'éthique médicale (Sparrow & Sparrow, 2007 ; Sharkey & Sharkey, 2010 ; Vallor, 2011 ; van Wynsberghe, 2013). De façon plus générale, en dehors des sphères académiques, l'éthique de la robotique est pensée et pratiquée dans une littérature grise principalement composée de rapports et d'avis institutionnels (Borelle, 2020) ou aux moyens d'instruments déontologiques « généraux » tels que les chartes éthiques⁸⁴. Comme le remarque Céline Borelle, ces outils

« ont en commun d'énoncer des recommandations à partir d'une réflexion plutôt prospective, sans s'appuyer sur des enquêtes empiriques relatives à l'état actuel des technologies. Concernant les robots sociaux, ces rapports entendent alerter l'opinion, les communautés professionnelles et les pouvoirs publics sur deux types de risques : celui d'un attachement inapproprié de l'humain à la machine susceptible de fragiliser le lien social, et celui d'une autonomisation démesurée de la machine assortie d'une perte de contrôle humain » (*ibid.*).

Outre ces questions d'attachement et de perte de contrôle, l'éthique des sociobots se concentre généralement sur sept autres points de vigilance communément identifiés, à savoir la duperie, l'impératif d'information, l'autorité robotique, le respect de la vie privée, le respect de l'autonomie et de la dignité humaine, les relations humain-humain et, enfin, les principes de justice et de responsabilité (Tzafestas, 2016). Précisons que ces thématiques sont principalement d'actualité dans le domaine médical, mais qu'elles se généralisent peu à peu, au rythme de la diffusion progressive des technologies sociobotiques. En voici le détail :

⁸⁴ A titre d'exemple très concis, l'Institut pour l'Étude des Relations Homme-Robots (IERHR), présidée par Serge Tisseron, psychiatre et membre de l'Académie des technologies, propose depuis 2017, une charte fondée sur les cinq principes suivants : 1) le respect de la liberté de chacun, 2) la transparence des algorithmes, 3) l'autonomie de l'utilisateur, 4) la dignité des individus, 5) l'égalité dans l'accès aux technologies (<https://www.ierhr.org/charte-ethique/>, consulté le 14/12/2021).

L'attachement : nous l'avons vu, la plupart des sociobots sont conçus pour générer une forme d'attachement émotionnel quelle que soit la tranche d'âge de l'utilisateur. S'il n'est pas nécessairement problématique en soi, cet attachement doit être appréhendé avec précaution selon les contextes pour ne pas créer l'illusion d'une réciprocité (Turkle, 2007, 2015). Ce point de vigilance éthique concerne particulièrement les publics vulnérables comme les enfants atteints de TSA dont nous comprenons encore mal les relations aux robots (Borelle, 2020). De façon plus générale, cet attachement prescrit aux soignants ou à tout autre responsable du robot de les traiter comme de simples objets. Par exemple, si l'artefact doit subir une réparation quelconque ou s'il est amené à ne plus être utilisé (e.g. pour des raisons de coût, d'inefficacité thérapeutique et/ou de pannes à répétition) dans un centre de soin, ce « départ » doit *a minima* être préparé en amont auprès des patients pour anticiper et minimiser la détresse voire la perte d'efficacité thérapeutique que cette absence pourrait occasionner.

La tromperie : le risque de tromperie associé aux sociobots est l'un des sujets les plus débattus en éthique de la sociobotique. Il se développe généralement dans le cadre d'argumentations centrées sur la crainte de *confusion ontologique* entre l'humain et la machine⁸⁵ au sens d'assimilation de la seconde au premier (Tisseron, 2015 ; Turkle, 2007, 2015). Dans le cadre médical, par exemple, cette confusion peut mener un patient à assimiler les capacités d'un robot de soin aux compétences d'un ou d'une soignante humaine et provoquer des attentes inadéquates. Face à ce genre de risques, l'enjeu est alors de faire en sorte que la nature artefactuelle des robots reste évidente. Cette prévention peut se manifester dans la conception elle-même et/ou dans les discours associés à l'utilisation du sociobot dans un environnement donné. Serge Tisseron (2015) propose, par exemple, de concevoir tout ou partie des « corps » robotiques en matière transparente afin de voir en permanence les fils et les câbles qui les composent. D'autres préconisent que les robots prennent « eux-mêmes » les devants en rappelant régulièrement ce qu'ils sont, que ce soit lorsqu'ils ne comprennent pas ce qu'on leur dit⁸⁶, ou de façon totalement préventive avant telle ou telle intervention de leur part.

⁸⁵ Nous proposons une critique de cette idée de confusion ontologique dans le quatrième point de ce chapitre.

⁸⁶ Par exemple, en répondant « Je te rappelle que je suis un robot, mes capacités de compréhension sont donc assez limitées » plutôt que « Je suis un peu dur de la feuille, peux-tu répéter ? ».

L'impératif d'information : ce devoir concerne aussi bien les destinataires de la médiation robotique (e.g. les patients) que ses destinataires (e.g. les soignants). Chacun doit être informé avec précision des risques ou des dangers liés à l'utilisation de l'artefact. Ces aléas peuvent être minimisés par la description détaillée des aptitudes et des limites du robot. La responsabilité de telles informations incombe à celles et ceux qui, au sens large, favorisent l'introduction de l'artefact dans tel ou tel environnement (ce qui inclut l'organisme de vente et les personnes qui, en interne, décident de recourir au robot).

L'autorité du robot : quel que soit leur domaine d'intervention, les sociobots sont conçus pour avoir de l'influence voire de l'autorité sur leurs utilisateurs. Cela pose bien entendu la question de l'identité et des intentions de celles et ceux qui conçoivent et/ou introduisent le robot, mais aussi celle de la transcription technique du respect de l'autonomie humaine. Par exemple, si un patient souffre lors d'une activité dirigée par un robot et souhaite y mettre fin, l'artefact doit être capable d'évaluer cette souffrance et d'arrêter l'activité comme le ferait un humain face à une douleur jugée anormale.

Le respect de la vie privée : il est important que l'utilisateur sache si le sociobot collecte des données et, le cas échéant, de quelles données il s'agit. Comme l'explique Spyros G. Tzafestas (2016), professeur émérite à la *National Technical University of Athen*, il est probable que le robot ne soit pas capable de discriminer entre les données partageables et celles qui ne le sont pas. En outre, il est également difficile d'assurer que l'artefact puisse clairement identifier qui peut accéder aux données sensibles d'un patient et qui ne le peut pas. Outre l'impératif d'information, il est également souhaitable que le patient (ou tout autre type d'utilisateur) puisse refuser la médiation robotique pour ces différentes raisons (ou demander de suspendre la collecte de données si cela est possible).

Le respect de l'autonomie et de la dignité humaine : Dans le domaine du soin comme ailleurs, un sociobot peut être amené à prendre des « décisions » impliquant le destinataire de la médiation robotique. Le respect de son autonomie et plus largement de sa dignité prescrit donc de récolter son consentement éclairé ou celui de sa figure tutélaire avant que ces médiations ne produisent leurs effets. Ce point de vigilance est intimement lié à l'impératif d'information puisque la clairvoyance du consentement ne peut se construire que sur la base d'une information intelligible et suffisante.

Les relations humain-humain : Depuis ses débuts, la robotique a soulevé la question de ce qu'on appelle aujourd'hui le « grand remplacement robotique »⁸⁷ (Arpin-Simonetti, 2020), c'est-à-dire la crainte que les machines remplacent systématiquement l'humain au travail et, plus largement, en société. Les sociobots ont un potentiel anxiogène particulier sur ce point puisqu'ils touchent à l'une des dispositions que nous croyons la moins délégable (en tout cas en première personne) aux machines, à savoir la constitution de relations sociales⁸⁸. L'un des grands principes consensuels de l'éthique sociobotique instrumentale est celui du caractère *complémentaire et/ou encapacitant* de la médiation robotique. Autrement dit, le robot doit intervenir en plus de l'humain et non pas à sa place, sauf de façon ponctuelle et ciblée. A nouveau, cela est particulièrement important lorsque l'artefact intervient auprès de personnes vulnérables qui, pour telle ou telle raison, souffrent d'isolement social. La complémentarité et la substituabilité du sociobot doivent ainsi faire l'objet d'un examen approfondi et, autant que possible, adapté à chaque individu.

Le principe de justice et de responsabilité : Dans un souci de justice le déploiement des sociobots doit répondre au principe de « répartition équitable des ressources rares ». Les technologies robotiques sont généralement très coûteuses, c'est pourquoi il est important de s'intéresser au rapport coûts/bénéfices qui est le leur. A titre d'exemple, l'enquête de Céline Borelle (2020) dans différents hôpitaux de jour, expose très clairement les difficultés que pose l'utilisation d'un NAO auprès d'enfants atteints de TSA. Outre des capacités techniques très limitées qui font de lui un « interlocuteur décevant », le sociobot n'a pas été pensé pour intervenir auprès de ces enfants. En 2013, SoftBank a donc décidé de concevoir une suite logicielle spécifique nommée *Ask Nao*⁸⁹ dont le bilan, deux ans plus tard, fut très mitigé. La configuration du robot demande des investissements temporels et financiers très voire trop importants par rapport aux bénéfices thérapeutiques que l'artefact semble capable d'offrir.

⁸⁷ Employée de façon tantôt ironique tantôt inquiète, cette formule est inspirée de la théorie complotiste du « grand remplacement » introduite par le militant français d'extrême-droite Renaud Camus. Sa théorie avance qu'un remplacement intentionnel et méthodique des populations françaises et européennes par des populations non-européennes (principalement maghrébines et subsahariennes) serait orchestré par les élites du vieux continent.

⁸⁸ Des relations épistolaires aux réseaux sociaux en passant par les appels téléphoniques, la technique a de tout temps médiatisé nos relations sociales. La singularité des robots sociaux est que la médiation se fait en « première personne » au sens où le robot se présente activement comme destinataire de l'interaction qu'il co-constitue.

⁸⁹ Pour « *Autism solution for Kids* ».

Concernant la responsabilité, la question est de savoir qui l'endosse en cas de nuisance. Si le problème relève d'une défaillance technique liée au design, au hardware ou au software, la responsabilité incombe aux concepteurs, au fabricant ou au distributeur. En revanche, si la nuisance procède d'un mauvais usage de l'artefact, il est bien sûr nécessaire de déterminer la cause de cette utilisation erronée avant de pointer un quelconque responsable. Est-elle liée à un manque évident d'informations sur le produit ? S'agit-il d'une simple erreur de jugement, d'un mauvais entretien du sociobot ou d'un usage volontairement nuisible de la part de l'individu ? Le script (Akrich, 2010), c'est-à-dire le scénario d'usage prévu par les concepteurs, de l'artefact est-il lui-même en cause ? En fonction des réponses à ces questions, la responsabilité incombera à l'utilisateur, aux personnes en charge de le former et/ou de l'informer quant aux risques ou encore à l'entité conceptrice du produit. L'autre grande problématique liée à cette question de la responsabilité est celle de l'impossibilité d'anticiper tous les effets indésirables d'une technologie.

L'objectif de l'éthique *autour* des robots, on le voit, est de garantir ce que nous proposons d'appeler leur *conformité axiologique*, c'est-à-dire de veiller à ce que les sociobots respectent des valeurs morales de référence conditionnant leur désirabilité ou, *a minima*, leur acceptabilité. Mais des valeurs pour qui, par qui et comment ? On le voit, la grande problématique est ici de savoir qui fixe le référentiel axiologique, pour quelles raisons et par quels moyens. Dans un article dédié à la sociobotique appliquée au grand-âge, Noël et Amanda Sharkey (2010) proposent par exemple d'évaluer les robots à l'aune du principe de bien-être humain (*human welfare*) et ce dans deux référentiels : la Déclaration universelle des droits de l'Homme (DUDH) d'une part et le *Value Sensitive Design* (VSD) d'autre part. Le référentiel universaliste et égalitaire des droits de l'homme permet aux auteurs d'ancrer l'assomption selon laquelle le bien-être physique et psychique des personnes âgées est aussi important que celui des autres. Dans la charte des Nations Unies, comme dans la DUDH, le principe de bien-être humain s'affirme dans les droits fondamentaux suivants :

- La dignité des conditions de vie (propices à la santé et au bien-être physique et psychologique) ;
- Le respect de la vie privée ;
- L'absence de toute forme de torture ou de traitement dégradant ;
- L'absence de toute forme de discrimination.

La question directrice des auteurs est donc celle de savoir si l'introduction de sociobots dans le soin des personnes âgées constitue un motif d'inquiétude à l'égard de ces grands principes de référence. Ils prennent toutefois soin de rappeler que, malgré la référence aux droits de l'Homme, leur réflexion ne s'inscrit pas dans une démarche juridique, mais axiologique, d'où le second référentiel qu'est le *Value Sensitive Design*. Fondé par l'informaticienne Batya Friedman dans les années 1990, le VSD est le premier champ de recherche à proposer une méthodologie d'éthique pragmatiste en trois phases – conceptuelle, empirique et technique – dont l'objectif est de structurer axiologiquement les processus de conception pour que les artefacts produits correspondent à des valeurs identifiées et volontairement transcrites techniquement.

« Le Value Sensitive Design s'intéresse essentiellement aux valeurs centrées sur le bien-être humain, la dignité humaine, la justice, la protection sociale et les droits humains. Il s'agit d'une "approche de principe"⁹⁰ en ce qu'elle soutient que de telles valeurs ont un statut épistémique et moral indépendamment de leur respect par une personne particulière ou un groupe. Dans le même temps, le Value Sensitive Design affirme que l'interprétation de ces valeurs peut considérablement varier selon les cultures et les époques » (Friedman & Kahn, 2003, p.1186).

En d'autres termes, les principes évoqués, dont celui de bien-être humain, *valent* indépendamment de leur inscription dans une pratique, ce qui revient à dire qu'ils sont moraux *en soi*. Outre sa valeur intrinsèque le rendant de facto légitime pour constituer une évaluation, le principe de bien-être humain affiche également une valeur relative évidente en matière de soin dispensé aux personnes âgées. Sharkey et Sharkey estiment donc « raisonnable » de le considérer comme un motif de préoccupation aussi bien majeur que partagé et de l'adopter comme clé de lecture pour identifier les vices et vertus potentielles des *carebots*. Selon eux, la désirabilité morale des robots sociaux à l'égard du bien-être des anciens n'est pas qu'une simple vue de l'esprit.

« Les robots d'assistance et les technologies robotiques pourraient aider à surmonter les problèmes de mobilité et réduire la dépendance des personnes âgées à l'égard du personnel soignant occupé et, parfois, inattentif. L'usage de robots contrôlés à distance

⁹⁰ « This approach is principled...»

pour surveiller et rendre virtuellement visite aux anciens pourraient leur permettre de vivre en indépendance plus longtemps. Les robots pourraient leur rappeler quels médicaments prendre et faire attention à d'éventuels problèmes de santé ou risque de sécurité. Les robots compagnons pourraient enrichir la vie sociale des personnes âgées en tant que gadget à propos desquels il est intéressant de parler. Les interactions sociales pourraient être également facilitées par des robots de téléprésence permettant des visites virtuelles d'amis ou de la famille » (Sharkey & Sharkey, 2010, p.37).

Ces vertus potentielles s'accompagnent toutefois de six points d'inquiétude majeurs. La plupart relève d'une déclinaison *in situ* des risques généraux évoqués précédemment.

1. Proches et soignants pourraient croire, à tort, que les robots sont capables de prendre en charge les principaux besoins physiques et émotionnels des personnes âgées. Les machines joueraient alors le rôle d'« excuses » (*ibid.*) pour se défaire de la charge morale et/ou professionnelle qui peut être celle du grand-âge. Le risque principal selon les auteurs est de voir les opportunités de contact social (humain-humain) réduites à la portion congrue, accroissant ainsi la négligence, pourtant déjà grande, de notre société envers ses anciens.
2. Viennent ensuite les risques liés à un usage purement pratique des robots, centré sur la seule commodité des soignants. Les machines participeraient alors du sentiment corollaire d'objectification et de perte de contrôle chez les personnes âgées.
3. La troisième problématique identifiée est celle que les technologies de captation des robots (liées à certaines de leurs potentielles « missions » de soin) posent à l'égard du respect de l'intimité des patients et donc, de façon plus générale, à leur liberté. La question est d'autant plus problématique que la « surveillance » (*monitoring*) fait très souvent partie des avantages reconnus aux technologies robotiques dans ce contexte.
4. La quatrième question concerne l'évaluation des restrictions de liberté qu'un robot pourrait imposer aux personnes âgées. Les auteurs prennent l'exemple (qu'ils qualifient eux-mêmes d'« extrême ») d'un robot qui empêcherait une personne de manger des aliments très caloriques afin d'éviter les risques de surpoids ou d'obésité.

5. L'avant-dernier risque est celui de la tromperie et de l'infantilisation des personnes âgées pouvant résulter d'une incitation à interagir activement et de façon répétée avec des machines comme si elles étaient de « vrais » compagnons. Cette question est d'autant plus complexe à traiter que l'illusion relationnelle peut être désirée par les patients eux-mêmes.
6. Enfin, le sixième et dernier point concerne la question du contrôle que les personnes âgées peuvent exercer sur le robot. Jusqu'où est-il souhaitable de répondre à leur désir d'autonomie ? Quel équilibre trouver entre ce désir et la condition physique et/ou mentale dégradée de certains anciens ?

Malgré la référence au VSD, l'article de Sharkey et Sharkey se concentre sur une analyse axiologique générale et vigilante qui n'a pas pour ambition de fournir un guide de conception éthique. Les contributions universitaires spécifiquement dédiées aux questions d'éthique sociobotique « opérationnelle » sont, à notre connaissance, très peu nombreuses. Selon nous, cela s'explique notamment par la dimension encore très confidentielle et prototypique de la robotique sociale. On trouve toutefois une référence intéressante d'éthique appliquée à la conception de carebots (qui peuvent être des robots sociaux ou sociétaux au sens que nous avons suggéré) dans les travaux d'Aime van Wynsberghe (2013). Son objectif est de théoriser une méthodologie résolument normative pour permettre aux ingénieurs d'intégrer les exigences du soin au design des futurs robots, mais aussi suffisamment souple pour que ces exigences soient celles des milieux spécifiques dans lesquels les artefacts interviendront. Pour ce faire, Wynsberghe opère un rapprochement entre le VSD et l'éthique du care. Elle fonde plus précisément son modèle (qu'elle nomme *care centered value sensitive design*) sur l'opérationnalisation, au sein du VSD, des quatre « phases du care » proposées par Joan Tronto (2009) – « Se soucier de », « Prendre en charge », « Prendre soin » et « Recevoir le soin ») – et des quatre « éléments éthiques » correspondant, à savoir l'attention, la responsabilité, la compétence et la capacité de réponse. Un tel cadre suppose, comme le remarque Xavier Guchet, d'engager les concepteurs « dans une démarche d'enquête quasi ethnographique auprès des soignants » (Guchet, 2022, p.63).

Dans ce tableau général de l'éthique *autour* du robot, une autre question s'avère très peu investie voire abandonnée aux initiatives plus ou moins construites des acteurs privés : celle de la communication et des façons de parler des sociobots.

3.2.2 La question (délaissée) de la communication

Nous l'avons déjà suggéré, l'éthique autour du robot doit aussi interroger la façon dont nous en parlons et, plus précisément, la façon dont ses concepteurs et distributeurs le présentent. Là encore, nous n'avons pas connaissance de travaux spécifiquement dédiés à cette question. Des critiques sont bien sûr émises sur le caractère trompeur des termes et formules telles que « robot compagnon », « robot empathique » ou encore « robot émotionnel » (Tisseron, 2015 ; Turkle, 2015), mais ces objections ne vont pas jusqu'à analyser le discours déployé par les firmes de robotique sociale dont l'objectif est souvent de combler, par une rhétorique de la promesse, le gouffre entre l'imaginaire associé aux robots et leur maturité technique. La promesse phare de la sociobotique marchande se concentre principalement sur la fonction supposée de compagnon ou d'assistant fidèle et dévoué. Il n'est donc pas étonnant que les fabricants mobilisent une rhétorique du soin et de la bienveillance où le robot social est plus volontiers nommé « robot émotionnel » ou « robot empathique ». La rhétorique prospective n'est pas nécessairement problématique en soi. Les technologies évoluent en partie grâce à cette mise en récit d'elles-mêmes et de projection qui suppose d'imaginer dans le présent un état de développement futur jugé désirable (encore faut-il savoir expliquer sérieusement pourquoi). Cet effort projectif permet d'orienter la recherche scientifique et les développements techniques. Le piège commence néanmoins lorsque le discours convertit des hypothèses prospectives en fausses certitudes aux moyens de promesses spé cieuses. Prenons un exemple tiré de notre propre analyse (De Sanctis, 2020b) d'un article rédigé par le directeur de la communication de SoftBank Europe (Halftermeyer, 2020) et publié sur le site internet *frenchweb*. Intitulé « Coronavirus : robots et drones bientôt au cœur de la crise française ? », l'article affiche une rhétorique de la promesse pour le moins problématique car fondée sur la confusion implicite entre ce qui pourrait (très hypothétiquement) être et ce qui est. Prenant pour exemples de départ les interventions de divers robots lors de catastrophes comme le tremblement de terre de 2016 au Népal ou l'incendie de Notre-Dame en 2019, l'auteur introduit ses protagonistes par des exemples assez consensuels :

« Ce n'est pas la première fois que les objets futuristes peuvent aider efficacement. Lors de l'incendie de Notre-Dame, le 15 avril 2019, le monde a observé, impuissant, des images

captées par un drone survolant l'édifice, doté d'une caméra thermique. Puis, de manière quasi-surréaliste, on a vu un robot à chenille Colossus_ – doté d'un canon à eau – entrer dans la cathédrale... »

On remarque le ton futuriste employé pour soutenir la dimension extraordinaire, sensationnelle de ces interventions (les robots sont qualifiés d'« objets futuristes », leur intervention est « quasi-surréaliste ») et laisser entendre que l'avenir est déjà là. Ces artefacts ne sont pourtant pas plus futuristes que leur intervention n'est extraordinaire : ils existent dans le présent, certains depuis plusieurs années, ont des fonctions bien identifiées et leur mobilisation relève de plus en plus de la norme dans les diverses procédures d'intervention. Est ensuite évoquée la crise sanitaire du Covid-19 à Wuhan où

« drones et robots ont repris un rôle déterminant, presque issu d'un film de science-fiction. [...] Dans les airs, les énormes hexacoptères de DJI⁹¹, habituellement utilisés pour des produits fertilisants utilisés en agriculture, diffusaient cette fois-ci des produits antibactériologiques. Des modèles plus petits avec caméra thermique se sont mis à faire l'inspection des appartements des tours de la ville : étage après étage, les familles se rendent sur les balcons ou devant les vitres afin de mesurer la température ! Dans la rue, les plateformes e-commerce ont déployé des minibus de livraison autonomes, croisant sans doute les petits tanks chenillés destinés eux à projeter des désinfectants. Dans les hôtels et hôpitaux, ce sont des robots de livraison qui apportent plats et médicaments. Un robot émotionnel Pepper est quant à lui déployé à l'accueil de Wuhan Wuchang, l'un des plus grands hôpitaux de la cité. »

L'accumulation est une figure rhétorique bien connue. Dans cet extrait, Wuhan semble s'être transformée en véritable « robotière » où des nuées de droïdes en tout genre s'affairent infatigablement dans les moindres recoins de la ville. La narration permet alors d'introduire, en fin de catalogue, un autre type d'artefact : le « robot émotionnel Pepper ». Les sociobots de SoftBank sont mis sur le même plan que les drones et autres automates d'intervention alors que, comme en témoigne la suite de l'article, leur contribution à l'effort de guerre est plus que douteux.

⁹¹ DJI (Da Jiang Innovation) est une société chinoise de fabrication de drones professionnels et de loisir.

« Les robots Pepper, d'une taille de 1 mètre 20, sont déjà en utilisation dans des hôpitaux comme Necker à Paris ou St-Marien à Cologne. Ils s'occupent de l'accueil des visiteurs ou de distraire les enfants. Nos robots peuvent mener différentes fonctions d'information et conduire des interactions riches avec les humains. »

Et l'auteur de poursuivre :

Ainsi, leur rôle pourrait devenir crucial dans des cas d'épidémies où le personnel médical est surchargé, fatigué. Avec leurs échanges en langage naturel, en plusieurs langues, la détection d'humains et la navigation autonome, il est possible pour les robots sociaux d'aider positivement à éliminer des moments chronophages ou stressants, en tant qu'assistant du personnel médical. Et même après la crise, il faudra soulager le personnel et lui permettre plus de plages de récupération, d'où l'idée de confier certaines tâches répétitives à des humanoïdes fonctionnant presque 24 heures sur 24. »

Quels moments chronophages ou stressants ? Quelles tâches répétitives ? La formule « aider positivement », naïve en apparence, n'utilise-t-elle pas de l'adverbe comme cache-misère pour pallier le manque d'idées concrètes et abouties ? Le fondateur d'Aldebaran, Bruno Maisonier, parlait lui-même de NAO en ces termes : « je ne sais pas à quoi ça va servir, mais je fais une plateforme et les gens inventeront les usages avec » (Borelle, 2020). En outre, pourquoi recourir aux « humanoïdes » alors que cette forme pose généralement plus de problèmes techniques qu'elle n'en résout, y compris pour prendre en charge des « tâches répétitives » ? Un Pepper est, par exemple, incapable d'attraper ou de porter des objets. Dès lors, auraient-ils pour seule vocation de s'occuper « de l'accueil des visiteurs ou de distraire les enfants », selon les mots de l'auteur ? Nous l'avons vu, les fonctions d'accueil et de divertissement sont des débouchés classiques pour les fabricants de sociobots. Au Japon, les boutiques SoftBank Mobile sont équipées de Pepper pour souhaiter la bienvenue aux visiteurs et les informer. En France, on a régulièrement parlé des mêmes robots installés dans des magasins Carrefour, Uniqlo ou encore dans des gares comme la Part-Dieu à Lyon. Toutefois, on souvent oublie de préciser que ces automates sont délaissés par les clients et rapidement remisés dans un coin du magasin puis retirés de la circulation. On assiste alors à ce genre de mise en scène sans acteur où un arsenal signalétique explicatif est déployé pour introduire un robot absent.



Gare de Lyon Part-Dieu (09/03/2019)

Parler d'« interactions riches » avec Pepper est résolument mensonger car, contrairement à ce que la formule peut laisser entendre, le « langage naturel » n'est aucunement synonyme d'aptitude à la conversation. Ainsi, lorsqu'on rapporte les (in)capacités réelles d'un sociobot au coût qui est le sien, l'idée que « leur rôle pourrait devenir crucial dans des cas d'épidémies où le personnel médical est surchargé, fatigué » est particulièrement surprenante. Elle l'est d'autant plus que l'article fut publié le 18 mars 2020, c'est-à-dire en pleine première vague de Covid-19 où le manque d'équipements primordiaux comme les respirateurs et les masques de protection engageaient soignants et soignés sur le plan vital. Précisons d'ailleurs que, lorsqu'il était encore produit, SoftBank proposait Pepper au prix minimum 20 000€, contrairement au 1400€ annoncés avant sa mise sur le marché (Lecomte, 2014). D'après un article cité par l'auteur lui-même, le coût du sociobot montait à 30 000€ si l'on y intégrait des fonctions logicielles d'accueil. Comparativement, un respirateur (technologie qui n'en est plus, loin s'en faut, au stade de la « preuve de concept ») coûte entre 10 000 et 25 000€ et des projets d'« innovation frugale » visent à baisser ce prix sous la barre des 500€ (Chodorge, 2020).

En l'absence d'éthique claire et structurée de la communication sociobotique, l'évaluation d'un tel discours peut être, *a minima*, adossée aux corpus de valeurs évoqués au début du

point précédent. On voit alors que l'article en question pose problème sur au moins trois éléments : la tromperie, l'impératif d'information et le principe de justice. La tromperie d'abord, puisqu'il est question d'« interaction riches » avec Pepper alors que celui-ci, comme la majorité des sociobots, se révèle être un « interlocuteur décevant » (Borelle, 2020). A cet égard, il convient de ne pas se laisser abuser par le jargon des *startups* et, en particulier, par la notion de « preuve de concept » (« *proof of concept* » ou « POC »). En sociobotique (au moins), ces expérimentations généralement produites en partenariat avec d'autres organisations ou institutions relèvent autant de l'exercice rhétorique que de la démonstration technique. Mises en avant par les jeunes entreprises pour obtenir des financements, elles ne prouvent généralement pas grand-chose d'autre que l'immaturation des systèmes déployés. On pourra rétorquer que l'objectif n'est pas de révéler un produit fini, mais un potentiel que l'afflux de ressources humaines et financières permettrait d'actualiser. C'est juste, mais le choix des mots n'en reste pas moins trompeur : une « preuve de concept » ne prouve rien, elle nous projette dans un futur hypothétique où le script *souhaité* par les concepteurs pourrait, sous certaines conditions, éventuellement s'actualiser. En toute rigueur, il serait donc plus juste de nommer cela *projection* de concept. S'il peut paraître anodin, ce point est en réalité tout à fait déterminant car il révèle la confusion sans cesse reconduite (volontairement ou non) entre présent et futur technologiques dont l'effet (délétère) bien connu est d'entretenir l'illusion d'une inéluctabilité technique. En lien avec ce premier point, l'article déroge également à l'impératif d'information puisqu'il n'évoque jamais les immenses limites des sociobots dont il fait la promotion et reste extrêmement vague sur ce que pourrait faire un artefact aussi coûteux en dehors d'une fonction d'accueil à la maigre valeur ajoutée. Enfin, le dernier point concerne le principe de justice. Si la communication n'en est pas directement responsable, elle peut néanmoins le menacer (ou le favoriser) indirectement. En l'occurrence, l'article s'avère peu vertueux sur ce point puisqu'il instrumentalise une situation d'urgence sanitaire pour promouvoir des artefacts au très mauvais rapport coût/bénéfice (défaut de vertu accentué par la pénurie d'artefacts indispensables comme les masques et les respirateurs).

Malgré des pratiques marketing souvent douteuses, la question de la meilleure façon de parler des sociobots et, plus précisément, de les *nommer* n'en demeurent pas moins une problématique réelle pour certains socioboticiens. Toute la difficulté est ici de clarifier et de fixer les critères qui rendront une appellation plus vertueuse qu'une autre. Chez Spoon, par

exemple, la formule « créatures artificielles » fut utilisée comme qualificatif officiel pendant environ cinq ans, ainsi qu'en témoigne l'ancien logo de la société. Encore aujourd'hui, sur le site internet de l'entreprise, on voit apparaître en page d'accueil la déclaration « *We design artificial creatures* »⁹².



Malgré son caractère officiel, cette terminologie n'était pourtant pas réellement stabilisée en interne. L'un des trois fondateurs l'affectionnait pour son côté à la fois « mignon » et démiurgique (la créature que l'on crée et dont on prend soin), alors que les deux autres manifestaient, pour l'un, une certaine indifférence à l'égard des qualificatifs employés (bien que conscient de leur importance) et, pour l'autre, une prudente réserve quant à sa pertinence tant conceptuelle que commerciale. Aujourd'hui, le PDG de Spoon, Jérôme Monceau, qualifie conceptuellement ses robots de « personnages interactifs » et la formule de « créatures merveilleuses » (« *marvelous creatures* ») est devenue le nouveau « slogan » de l'entreprise.



Nous approfondissons notre analyse de ces questions dans la section consacrée à la présentation Spoon. Pour l'heure, examinons l'avant-dernier volet de cette section consacrée à la roboéthique, à savoir l'éthique *dans* le robot.

⁹² « Nous concevons des créatures artificielles »

3.3 L'éthique *dans* le robot : la moralité « implicitement » et « explicitement » programmée

L'image d'une éthique *dans* le robot renvoie à l'idée de sa délégation ou de sa prise en charge par le robot lui-même au sens où elle serait codée, d'une façon ou d'une autre, dans son logiciel. Autrement dit, cette partie de l'éthique entend permettre aux artefacts de fonctionner moralement « par eux-mêmes » en vertu de leur programmation. Elle s'insère donc dans les recherches sur les AMA et la « moralité sans esprit » déjà mentionnée. Comme nous allons le voir, l'éthique dans le robot peut se réaliser sur le mode implicite ou explicite. Il existe plusieurs moyens de moraliser une technologie, c'est-à-dire de la rendre éthique ou de la faire advenir comme telle. Nous l'avons vu dans le paragraphe précédent, l'une des possibilités consiste à orienter le design de l'artefact afin que sa conception, tant en termes de processus que de produit fini, promeuvent ou, *a minima*, respectent des normes ou des valeurs morales spécifiques. Une technologie donnée peut aussi être spécifiquement conçue pour répondre à un problème moral. Parmi les nombreux exemples disponibles, donnons celui des jockeys robotiques développés par la société Suisse K-team en réponse aux pratiques esclavagistes des courses de chameau au Qatar (Lewis, 2005). Historiquement, les propriétaires qataris de chameaux asservissaient de très jeunes enfants venus de pays étrangers (en particulier du Soudan) pour chevaucher les montures. Dépourvus de toute éducation élémentaire, ces enfants étaient bien souvent sous-alimentés afin qu'ils pèsent le moins possible sur les chameaux (plus les premiers sont légers, plus les seconds sont rapides) et ne bénéficiaient d'aucun soin particulier en cas de blessure. L'ONU, associée au département d'État américain, voulut mettre fin à ces pratiques et au trafic d'êtres humains qu'elles impliquaient en menaçant le Qatar de sanctions économiques. C'est sur cette base que K-Team fut sollicitée pour développer, avec succès, des substituts robotiques aux enfants esclaves⁹³. Voilà qui relève de l'éthique *autour* du robot. En matière d'éthique *dans* le robot,

⁹³ Hélas, l'histoire ne s'arrête, pas là car ce succès technique des jockeys robotiques et leur adoption par les personnes concernées n'étaient qu'un maillon de la chaîne permettant d'offrir aux enfants esclaves une vie meilleure. Alors que le département d'État américain avait enjoint le Qatar à mettre en place un programme de rapatriement et de réhabilitation des enfants affranchis, les autorités Qatari se contentèrent de les renvoyer dans leur pays d'origine où rien ni personne ne les attendait. Comme le déclarait un propriétaire de chameau : « Si vous délivrez [ces enfants] de ce travail, allez-vous leur donner une vie meilleure ? Non. Ils iront vers leur mort. Ils retourneront au Soudan et vivront là-bas comme ils y vivaient avant, sans éducation, sans nourriture, sans rien. A quel point leur vie sera difficile là-bas ? Tout le monde s'en fiche » (Lewis, 2005).

à présent, on peut distinguer ce que James Moor (2006) nomme les « agents éthiques implicites » (*implicit ethical agents*) et les « agents éthiques explicites » (*explicit ethical agents*). Un agent éthique implicite est un artefact dont le code est implicitement structuré pour répondre à des exigences morales ou, *a minima*, pour éviter des comportements et des effets immoraux. Moor donne l'exemple d'un logiciel pharmaceutique dont la fonction est de rendre compte des interactions médicamenteuses. Un tel logiciel est conçu pour aider médecins et pharmaciens dont le devoir, tant éthique que légal, est de s'assurer que tout médicament prescrit profite au patient plus qu'il ne lui nuit. L'artefact joue donc le rôle d'assistant à la prescription, portant ainsi implicitement la valeur de bienfaisance/non-malfaisance de l'éthique médicale⁹⁴. L'agent éthique explicite, quant à lui, est un artefact dont le comportement est informé par des règles d'éthiques normatives explicitement inscrites dans son code. L'ambition de concevoir ce type d'artefact est portée depuis le début des années 2000 par la *machine ethics*, un courant de recherche ayant pour objectif d'interroger les conditions de possibilité de l'intégration des théories éthiques aux systèmes computationnels. L'un des constats fondateurs du champ met en évidence le déséquilibre patent entre normativité technique et normativité morale dans notre appréhension des technologies informatiques. Autrement dit, nous comprenons très bien ce que signifie un « bon programme » du point de vue technique (i.e. un programme qui exécute correctement la fonction prescrite par son code), mais pas du point de vue moral. Or, les contributeurs et contributrices à l'éthique des machines partent du principe que la question de l'autonomie morale des artefacts se posera de plus en plus vigoureusement à mesure que nous développerons leur autonomie pratique. Selon eux, il est donc impératif d'articuler la question ontologique et morale du « bon programme » à celle de sa transposition opérationnelle dans les logiques spécifiques du code informatique. Du point de vue théorique, Colin Allen, Ida Smit et Wendell Wallach (2005) identifient trois grandes façons d'intégrer l'éthique « dans » le robot :

⁹⁴ Ainsi défini, on peut considérer que le régime de l'agent éthique implicite s'applique également aux jockeys robotiques, ce qui en ferait des représentants de l'éthique « dans » le robot et non plus « autour ». Comme souvent en matière de catégorisation, il existe une certaine perméabilité entre les catégories. Il nous semble toutefois plus pertinent de maintenir notre proposition initiale en ce que la notion d'agent éthique implicite qu'utilise Moor est spécifiquement tournée vers les artefacts informatiques dont le code est axiologiquement informé, ce qui n'est pas le cas des jockeys robotiques. C'est donc bien le paramètre du code qui permet de fonder l'image d'une éthique dans le robot.

1. l'approche *top-down* ou descendante centrée sur l'application de règles morales au sens où les actions de l'artefact sont subordonnées à un ensemble de principes (issues des grandes théories éthiques que sont le vertuisme, le déontologisme et le conséquentialisme) codés dans son logiciel ;
2. l'approche *bottom-up* ou ascendante centrée sur l'apprentissage moral de l'agent qui ne possède pas de système moral *a priori* et « apprend » à se comporter moralement en évaluant la conformité de ses actions à la morale en suivant ses propres règles (son modèle d'apprentissage étant humainement supervisé afin de l'améliorer) ;
3. l'approche hybride qui propose de combiner les deux précédentes pour prendre en charge une plus grande complexité.

Force est toutefois de constater que la *machine ethics* reste, aujourd'hui encore, un domaine de recherche principalement théorique. Elle a toutefois beaucoup fait parler d'elle ces dernières années – sans forcément être nommée – suite à l'emballement médiatique autour des véhicules autonomes et des questions éthiques soulevés du type « dilemme du tramway ». Nombre d'universitaires estiment toutefois qu'il s'agit là de problèmes factices. C'est notamment le point de vue de Sophie Sakka, maîtresse de conférence à l'école Centrale de Nantes et présidente de l'association « Robots! »⁹⁵, qui, dans l'entretien qu'elle nous a accordé, s'exprimait en ces termes :

« Après, il y a aussi la question des "faux problèmes" éthiques ou des débats éthiques construits sur des bases quelque peu fantasmatiques. Je pense notamment à la question des voitures autonomes et des décisions qu'elles devraient supposément prendre en cas d'accident : écraser deux personnes âgées ou un enfant ? Faucher une femme enceinte ou sacrifier ses passagers ? Personnellement, ça me heurterait de devoir programmer ce genre de règles dans une machine car, la "solution" au problème étant inscrite dans le programme, un hacker pourrait très bien la modifier et choisir, par exemple, d'écraser l'enfant au lieu des personnes âgées. Si rien de tout ça n'est programmé initialement, il n'y a aucune chance que ça se produise, même si la machine est piratée. En outre, si le

⁹⁵ Fondée en 2014, l'association loi 1901 Robots! se donne pour objectif de diffuser les savoirs liés à la robotique et à son utilisation. Elle s'interroge également sur l'utilité sociale des robots en termes d'amélioration du quotidien (<https://www.association-robots.com/>).

véhicule autonome est suffisamment bien programmé, la question ne devrait même pas se poser : avec sa multitude de capteurs bien plus adaptés que les sens humains, le véhicule s'arrêterait avant toute situation difficile » (De Sanctis, 2019b).

En peu de mots : outre le problème éthique posé par l'idée même de programmer l'éthique, le véritable enjeu normatif, tant technique que moral, est précisément celui de ne pas avoir à créer ces situations. De ce point de vue, un *bon* véhicule autonome est celui dont la maturité technique ne donne aucune prise à la nécessité de devoir effectuer de tels choix. Il est toutefois nécessaire de préciser un point sur cette critique afin qu'elle ne se disqualifie pas elle-même : il ne s'agit pas d'affirmer qu'une « bonne » technologie est une technologie qui ne pose *aucun* problème moral, ce qui substituerait au thème de la neutralité technique celui, tout aussi douteux, de sa perfection morale. La difficulté est plutôt d'identifier les faux problèmes, c'est-à-dire ceux qui ne devraient pas se poser, afin qu'ils n'encombrent pas la conception.

Bien sûr, la *machine ethics* ne s'épuise pas dans la question des véhicules autonomes et encore moins dans les débats, parfois stériles, qu'elle peut engendrer. Son intérêt premier est celui des grandes questions théoriques et conceptuelles que soulève son ambition opérationnelle. Tout d'abord, cette introduction de la morale « dans » le robot par voie de programmation informatique modifie-t-elle son statut ontologique et moral ? Un agent éthique explicite est-il « plus » moral qu'un agent éthique implicite du fait de son autonomie apparente ? La « moralité sans esprit » de la *machine ethics* ne fait-elle pas de l'artefact une entité intrinsèquement morale au sens non pas relationnel ni même dispositionnel, mais catégorique du terme ? En outre, comme le note Flora Fischer dans sa thèse de doctorat en philosophie, la *machine ethics* nous renvoie

« à une interrogation profondément méta-éthique, qui semble moins présente dans les débats : qu'est-ce que la théorie des AMA dit de notre conception de la morale ? Y a-t-il un comportement moral idéal, universel ? La morale consiste-t-elle seulement à suivre des principes et des règles ? Ne peut-elle pas également se définir par le non-respect des règles, comme Racine nous le décrit avec *Antigone*, qui en faisant le choix de sa morale, désobéit à la loi ? Vouloir formaliser notre moralité dans un système technique supposerait de résoudre les débats séculaires qui existent entre éthiciens depuis l'Antiquité » (Fischer, 2020, p.80-81).

Mais la *machine ethics* pose une question plus spécifique à notre contexte de recherche : celle des liens entre moralité explicite et altérité. La « moralité sans esprit » ou, plus précisément, la « moralité *explicite* sans esprit » conduit-elle à une altérité elle aussi sans esprit ? La moralité conditionne-t-elle l'altérité de sorte que, pour être un autre, il faille faire explicitement preuve, au moins en apparence, d'une conduite morale ?

3.4 L'éthique avec le robot : une éthique immanente à la relation

Les trois éthiques que nous venons de détailler ont ceci de commun qu'elles s'intéressent au robot à partir de critères préétablis. Ces différents critères ont pour objectif de trancher la question de l'appartenance de l'artefact à la communauté morale ou de l'informer moralement au niveau de la conception. Dans tous les cas, il est question d'étudier la conformité morale de l'artefact en tant qu'objet, c'est-à-dire de l'extérieur, *indépendamment* de son actance et des relations que nous pouvons tisser avec lui (sauf, bien sûr, chez Bruno Latour). Ce dernier point est donc consacré à l'éthique *avec* le robot en tant qu'éthique fondée sur les relations humains-robot. Cela implique par exemple de penser le robot comme co-auteur de son statut moral.

3.4.1 « The machine question » : une approche lévinassienne

La formule anglaise de « *machine question* » est utilisée, par David Gunkel (2007, 2012, 2014) comme objection à ce qui, selon lui, relève d'une injustice dans notre rapport aux artefacts. Cette injustice – bien qu'il ne le formule pas en ces termes – repose sur le décalage entre progressisme ontologique et conservatisme moral dans la pensée des techniques ces cinquante dernières années. Plus précisément, nous pouvons observer que le dépassement de la conception « anthropologique et instrumentale » (Heidegger, 2013) des techniques en philosophie ne s'est pas traduit, comme ce fut le cas pour les animaux, par l'attribution d'un statut moral aux artefacts. Ce déséquilibre aboutit à un paradoxe : bien qu'ils soient sortis ontologiquement du régime de l'instrumentalité, les artefacts continuent d'y appartenir moralement. Prévenons d'emblée une méprise éventuelle. Une telle remarque ne signifie pas que l'éthique des techniques continuent de penser les artefacts comme de simples instruments dont la valeur morale dépendrait uniquement de leurs usages, bien au contraire.

Cette thèse de la neutralité axiologique des artefacts est depuis longtemps disqualifiée philosophiquement au profit d'interprétations accréditant le caractère moralement constitutif des techniques. On pense en premier lieu aux travaux déjà cités de Peter-Paul Verbeek⁹⁶ (2011) dont les influents concepts de médiations technologiques et de moralité intrinsèque (*intrinsic morality*) des artefacts intègrent à la philosophie des techniques l'idée latourienne selon laquelle les technologies contribuent à configurer les situations morales. En d'autres termes, Verbeek, comme Latour avant lui, affirment « que l'agent humain n'est jamais moral tout seul : il n'est moral qu'avec ses techniques » (Guchet, 2022, p.57). Mais parce qu'elle se focalise sur la primauté des relations entre actants (ontologie relationnelle) et sur les propriétés médiatrices des artefacts, la postphénoménologie de Verbeek pense l'éthique des techniques en termes de désirabilité des médiations technologiques et non en termes de statut moral. La postphénoménologie parle bien d'une agentivité morale (*moral agency*) des artefacts, mais celle-ci relève du statut ontologique de médiateur, pas d'un quelconque statut moral entendu comme l'attribution d'une forme de considérabilité. Autrement dit, si les artefacts sont ontologiquement des médiateurs de nos relations, notamment morales, au monde, cela n'en fait pas des êtres moraux pour autant. L'idée de « moralité intrinsèque » est donc formellement trompeuse sur ce point. L'agentivité morale d'un artefact dont parle la postphénoménologie ne relève pas d'une propriété intrinsèque au sens catégorique et/ou dispositionnel du terme, mais d'une propriété relationnelle⁹⁷. Or, l'attribution de statut moral, quel qu'il soit, se fait précisément sur la base de propriétés intrinsèques (Coeckelbergh, 2012a), c'est-à-dire sur la base de propriétés qu'un être possède *en lui-même* et non en vertu de ses relations avec l'extérieur.

La *machine question* de Gunkel se présente comme une critique de cette modalité d'attribution dont le principe consiste à déterminer de façon *a priori*, c'est-à-dire antérieurement à toutes formes de relations, qui est digne d'appartenir à la communauté morale et qui ne l'est pas. Cette question doit donc avant tout être comprise comme *remise en question* de notre système moral *par* la machine et comme une invitation à « penser autrement » (*thinking otherwise*) notre rapport, non seulement aux techniques, mais plus largement au monde. Selon Gunkel, le problème d'un tel système est qu'il n'accorde aucune

⁹⁶ Travaux auxquels Gunkel, étonnamment, ne se réfère presque jamais.

⁹⁷ Pour utiliser la terminologie proposée précédemment, on peut dire que l'agentivité morale de la postphénoménologie est une actance morale.

place à l'altérité et ce, malgré le passage d'une « éthique de l'exclusion » à une « éthique de l'inclusion » ces cinquante dernières années (Gunkel, 2007, 2012). Suivant le constat de Floridi (2010), l'auteur note que la philosophie n'a cessé d'étendre les frontières de la communauté morale à des êtres qui en étaient traditionnellement exclus. Cette extension fait suite à l'émergence de nouveaux critères de considération marquant le passage d'une pensée morale « classique » centrée sur l'agent humain et/ou sur ses actions (vertuisme, déontologisme, conséquentialisme) à des éthiques s'ouvrant à la perspective du patient moral, c'est-à-dire à toute entité, humaine ou non, pouvant être touchée par l'action des agents (éthique animale, éthique environnementale, éthique médicale, éthique du *care*, éthique de l'information etc.). L'évolution de la philosophie morale depuis les années 1970 peut donc être perçue comme une sortie progressive de l'anthropocentrisme au profit de divers allocentrismes au sens d'ouverture aux altérités non-humaines. Par exemple,

« [l']animocentrisme focalise son attention sur l'animal et étend la considération à tout organisme aux critères définissant l'animalité. Le biocentrisme fait un pas de plus dans le processus d'abstraction ; il fait de la vie le dénominateur commun et admet dans la considération tout ce qui peut être qualifié de vivant. Enfin, l'ontocentrisme complète cette progression en incorporant à la considération tout ce qui existe, a existé ou qui existe potentiellement » (Gunkel, 2007, p.147).

Or, aussi progressiste soit-il, Gunkel remarque que ce processus allocentrique d'ouverture reste, comme son nom l'indique, un processus fondamentalement *centré* qui ne s'ouvre jamais réellement à l'altérité puisqu'il n'échappe pas à la polarité inclusion/exclusion.

« En choisissant une approche centrée, ces différentes théories éthiques s'efforcent d'identifier ce qui est essentiellement identique au sein d'une diversité phénoménale d'individus. Par conséquent, elles incluent les autres en réduisant les différences ou en les évacuant. Bien qu'elle donne l'impression d'être de plus en plus inclusive, cette approche efface immédiatement l'altérité unique des autres et les ramène au même, instituant ce que Slavoj Žižek nomme la structure de Möbius : "Au cœur même de l'Altérité, nous rencontrons l'autre face du Même" » (*ibid.*).

Et Gunkel de conclure :

« L'exclusion est problématique car elle attire l'attention et se concentre sur ce qui est différent en dépit des similarités. L'inclusion est problématique car elle insiste sur les similarités au détriment des différences. [...] Ce qui est requis, dès lors, c'est une troisième alternative qui ne s'oppose pas simplement à l'exclusion par l'inclusion et vice versa » (*ibid.*, p.148).

Gunkel trace les contours de cette troisième voie en empruntant à l'éthique lévinassienne de l'altérité⁹⁸. Pour s'extraire de la polarité inclusion/exclusion, explique-t-il, il est nécessaire d'inverser le rapport qui place l'éthique, entendue comme relation à autrui, en aval de l'ontologie. Pour Levinas (1961), l'autre n'est autre que comme *tout autre*, c'est-à-dire uniquement comme celui qui échappe à toute détermination ontologique, à toute thématization. Un autre que je pourrais constituer, penser, définir comme étant ceci ou cela quitterait *de facto* l'altérité pour la même. Ce tout autre qu'est l'autrui lévinassien m'apparaît à travers l'expérience de sa visagité. Cette expérience est toutefois paradoxale puisque, n'appartenant pas à l'ordre du constituable, elle est sans forme, sans essence, sans phénoménalité. Le visage est une épreuve perpétuelle et constituante, une épreuve à laquelle nous sommes soumis à chaque instant de notre vie et, plus radicalement, à laquelle nous sommes toujours déjà soumis : celle de la vulnérabilité obligeante d'autrui qui exprime, dans la nudité la plus totale, l'interdiction du meurtre. Le renversement Lévinassien de l'ontologie en éthique va jusqu'à affirmer la constitution de la subjectivité par la visagité : le sujet émerge comme requis par la responsabilité qu'autrui lui assigne et qui le fait exister comme sujet. Autrement dit, nous sommes sujets parce que responsables et non l'inverse. L'expérience de l'altérité n'est donc pas *mon* expérience, comme l'explique François-David Sebbah, mais l'expérience par laquelle « je surgis comme requis absolument par ['] injonction traumatique qui vient du visage de l'autre » (Sebbah, 2016).

Gunkel ne va pas si loin dans son emprunt à Levinas. Sa proposition consiste principalement à s'inspirer de l'inversion du rapport entre éthique et ontologie pour repenser notre posture à l'égard des machines et, plus largement, de l'ensemble des êtres avec qui nous entrons en relation. Car c'est bien la relation aux choses qui prime lorsque nous déterminons leur valeur morale, et non ce qu'elles sont ontologiquement.

⁹⁸ A dire vrai, c'est l'ensemble du problème qui est posé en termes lévinassiens.

« Alors que nous rencontrons et interagissons avec d'autres êtres – qu'il s'agisse d'un autre humain, d'un animal, de l'environnement naturel ou d'un robot domestique – nous expérimentons d'abord ces autres en tant que nous sommes en relation avec eux. Ainsi, la question du statut moral ne dépend pas de ce que l'autre est dans son essence, mais de la position qu'il/elle/cette chose⁹⁹ (et le choix des noms et pronoms fait partie du problème) occupe dans la relation qui nous lie et de la façon dont nous décidons, face au visage de l'Autre (pour utiliser une terminologie lévinassienne), de répondre. Par conséquent [...], ce que l'entité est *ne détermine pas* le degré de valeur morale dont elle bénéficie. L'exposition au visage de l'Autre, ce que Levinas nomme "éthique", précède et prime toutes ces machinations et déterminations ontologiques » (Gunkel, 2016, p.206).

Cette préséance de l'éthique sur l'ontologie ne relève pas, selon Gunkel, d'une proposition théorique, mais d'une réalité à la fois empirique et phénoménologique. Au quotidien, nous n'attribuons pas de valeur morale aux choses sur la base de critères ontologiques fixés à l'avance, mais sur la base des relations que nous tissons avec elles. Dans cette configuration, l'ontologie n'intervient que de manière rétroactive pour justifier des *décisions* déjà prises en amont, dans le concret de nos interactions avec les êtres.

« En situations sociales – lorsque nous sommes confrontés à l'extériorité des autres – nous décidons toujours déjà "qui" [*who*] compte moralement et "ce qui" [*what*] ne compte pas, puis nous justifions rétroactivement ces choix en "trouvant" les propriétés dont nous croyons qu'elles motivent notre décision. Ces propriétés ne sont donc pas les conditions de possibilité intrinsèques et *a priori* du statut moral. Elles sont le produit *a posteriori* de relations sociales extrinsèques avec et face aux autres » (*ibid.*, p.208).

Autrement dit, l'attribution de valeur morale est un acte non pas transcendant, mais immanent à la relation. Gunkel en veut pour preuve la différence que, nombre d'entre nous, font entre un chien et un cochon. Sur la base des critères généralement admis pour la considérabilité morale, à savoir la conscience, la sentience ou encore la souffrance, ces deux animaux sont « virtuellement indistincts » (*ibid.*, p.213). Or, dans les faits, le chien occupe généralement une place moralement supérieure à celle du cochon : celle d'un ami ou d'un

⁹⁹ Le texte anglais dit : « ... on how she/he/it stands in relationship to us ». Le pronom « it » n'existant pas en français, nous proposons de le traduire par « chose ».

membre de la famille¹⁰⁰. Le cochon, lui, reste majoritairement marqué par le statut instrumental d'aliment, de viande¹⁰¹. L'éthique, telle que la conçoit Gunkel, consiste « à maintenir ouverte la possibilité [que tout être] puisse devenir un Autre » (*ibid.*, p.214), un « *significant other* », pour employer une expression anglaise très parlante, en reconnaissant la nature relationnelle des statuts moraux que nous attribuons. Dans une telle perspective, un robot social n'est donc pas d'abord une machine, un instrument ou un compagnon, mais un « autre » qu'il s'agit de qualifier relationnellement.

Si notre lecture est correcte, la position de Gunkel consiste, en substance, à situer le commencement de l'éthique dans la résistance aux réflexes et assignations ontologiques *a priori* pour ancrer la qualification des êtres à *partir* des relations que nous entretenons avec eux, et non au-delà de ces relations. Toutefois, on peut se demander si la solution proposée par Gunkel ne consiste pas, en fin de compte, à remplacer une oppression par une autre. En effet, si l'ontologie fige la valeur morale des choses en la prédéterminant, la relation ne présente-t-elle pas le risque exactement inverse ? Gunkel n'est, par ailleurs, pas très clair quant au rôle effectif de l'ontologie. En effet, le problème initial semble être celui de son caractère prédéterminant, mais l'auteur finit par affirmer qu'elle n'intervient en réalité qu'*a posteriori* pour justifier des décisions toujours déjà prises en situation, c'est-à-dire en relation. Cette apparente contradiction dans l'argumentation ne montre-t-elle pas que l'enjeu est de penser le rôle de l'ontologie dans nos relations, plutôt que d'opposer la première aux secondes ? En outre, Gunkel devrait préciser, en gage de précaution théorique, qu'il rompt avec Levinas presque immédiatement après l'avoir invoqué. En effet, l'éthique lévinassienne de l'altérité n'est en rien relationnelle au sens classique du terme. Il existe bien un lien entre autrui et moi, mais ce lien est celui de la responsabilité pour autrui que son visage m'assigne de façon absolument unilatérale. Attribuer relationnellement une valeur morale aux choses plutôt qu'ontologiquement, c'est encore ramener ces choses à soi, donc au Même, puisque nous les valorisons par rapport à l'effet qu'elles ont sur nous. Voilà pourquoi il s'avère théoriquement périlleux d'invoquer la métaphysique lévinassienne pour soutenir des thèses

¹⁰⁰ Un éleveur de porcs ne verrait probablement pas les choses de cette façon, en supposant bien sûr qu'il aime ses animaux et voit en eux autre chose qu'un gagne-pain. Quoi qu'il en soit, cela confirme l'importance voire la primauté de la relation eu égard à nos évaluations morales.

¹⁰¹ Ce qui n'empêche pas de bien traiter le cochon durant l'élevage et de mettre en place les techniques d'abattage les plus indolores et les moins stressantes possibles. Ces choix peuvent d'ailleurs répondre à des objectifs à la fois moraux et instrumentaux puisqu'un animal élevé dans de bonnes conditions donnera une viande de meilleure qualité.

dont la portée est, en dernière instance, empirique¹⁰². A cet égard, l'ancrage phénoménologique et socio-relationnel que Mark Coeckelbergh donne à la considération morale nous semble théoriquement plus adaptée¹⁰³.

3.4.2 Une approche phénoménologique et socio-relationnelle de la roboéthique

Les contributions de Gunkel et de Coeckelbergh ne diffèrent pas tant quant à leur objectif ou résultat qu'au chemin emprunté pour y parvenir. D'inspiration phénoménologique, la réflexion de Coeckelbergh (2010b) cherche à construire la considération morale à partir de l'*expérience* que nous avons des choses, expérience qu'il situe dans un contexte nécessairement relationnel. Cette démarche suppose de lier la considérabilité morale d'une entité à la façon dont elle nous *apparaît* dans le concret de nos relations et non à ses « caractéristiques ontologiques *réelles* » (*ibid.*, p.214). Cela implique plus précisément trois grands principes :

- 1- La considération morale n'est plus affaire de propriétés intrinsèques (en tout cas pas exclusivement), mais de propriétés extrinsèques tenant à la relation entre l'entité en question et nous ;
- 2- Les propriétés individuelles de l'entité peuvent toujours avoir une signification morale, mais cette signification n'est pas tant ontologique que phénoménologique. Ce sont les propriétés en tant que phénomènes qui comptent, c'est-à-dire les propriétés en tant que nous les expérimentons ;

¹⁰² De façon plus générale, l'une des principales critiques à laquelle cette approche relationnelle prête le flanc est celle de son relativisme moral. La position de Gunkel (2014) consiste à intégrer cette critique à sa réflexion en attribuant au relativisme une valeur épistémologique positive. Au lieu de désigner une société dépourvue de normes où la valeur des choses n'aurait pas d'autre ancrage que nos évaluations individuelles, un tel relativisme fait signe vers une mode d'organisation où les normes seraient continuellement établies et renouvelées collectivement. N'est-ce pourtant pas déjà ce qu'il se passe dans les sociétés démocratiques ? S'agirait-il, en dernière instance, d'un problème de trop grande inertie dans le renouvellement des normes et des valeurs ? Si oui, l'accélération de ce renouvellement ne risquerait-il pas de générer des problèmes symétriquement inverses, à savoir une instabilité axiologique et normative mettant en péril nos repères existentiels déjà lourdement affectés par les multiples crises que nous traversons ?

¹⁰³ Pour être très précis, Coeckelbergh s'appuiera lui-aussi fortement sur Levinas dans un article de 2016 intitulé « Alterity Ex Machina: The Encounter With Technology as an Epistemological-Ethical Drama ». Nous y reviendrons.

- 3- L'expérience en question n'est jamais indépendante du contexte ni du sujet qui la vit au contraire, elle est « *context-dependent* » et « *subject-dependent* » affirme Coeckelbergh. Concrètement, cela indique que la signifiante morale (*moral significance*) est relative non seulement à la relation entre le sujet et l'entité, mais aussi au contexte dans lequel son attribution intervient. Par exemple, la considération morale attribuée aux chiens résulte à la fois de nos relations individuelles avec eux et du contexte socio-politique qui, au-delà de la seule question canine, est propice à l'attribution de droits aux animaux en tant qu'ils sont de plus en plus reconnus comme des patients moraux.

On voit que Coeckelbergh n'évacue pas tant l'ontologie que la métaphysique au sens essentialiste et ontologiquement hiérarchique du terme. Comme Latour, dont il cite brièvement les travaux, Coeckelbergh pense l'ontologie en terme relationnels.

« L'idée est que si nous vivons avec des robots artificiellement intelligents, nous ne restons pas les mêmes individus et les mêmes humains que nous étions auparavant. Dès lors, la théorie relationnelle de la considération morale proposée ici doit être connectée à une théorie relationnelle de l'identité et, au final, à une ontologie (socio-) relationnelle » (*ibid.*, p.215).

Cette forme d'éthique *avec* le robot prend donc le singulier au pied de lettre : on ne peut parler que d'éthique avec « un » robot en particulier et non avec « le » robot comme catégorie, chaque entité étant le fruit de ses relations singulières avec d'autres entités. Il est donc impossible de prévoir à l'avance les formes de considérations que l'on pourrait attribuer à un robot (étant bien sûr entendu que l'acte d'attribution en lui-même ne peut venir que de nous). Comme chez Gunkel ou même chez Lestel, elles sont à inventer à partir des situations qu'offrent les relations. Coeckelbergh invite alors à explorer différents degrés de considération morale et à ne plus penser la communauté morale selon la dichotomie dedans/dehors. L'enjeu n'est plus tant celui du statut moral et des propriétés de chaque individu que de la désirabilité des relations qui se tissent entre eux. Dans l'ensemble, les similitudes entre cette approche et les thèses défendues par Verbeek en postphénoménologie sont très grandes, mais Coeckelbergh, étrangement, ne les relève pas.

En tout état de cause, cette réflexion soulève d'épineux problèmes comme celui de savoir comment créer du politique, c'est-à-dire du commun, sur cette question de la désirabilité des relations humain-robot, individuelle par principe ? Par ailleurs, cette forme d'éthique avec le robot ne s'intéresse au robot que sous l'angle du phénomène d'attribution de considérabilité morale, quelle qu'elle soit, mais qu'en est-il de la régulation de ce que l'artefact doit être et ne pas être, faire et ne pas faire ? Une éthique avec le robot est-il possible sur ce point ?

3.4.3 De la critique de l'empathie artificielle à l'anthropomorphisme comme fondement d'une « éthique synthétique »

L'une des principales critiques morales adressées à la robotique sociale est celle du caractère inauthentique et foncièrement trompeur de l'anthropomorphisme suscité par ses artefacts. Elle se fonde, nous l'avons évoqué dans le point consacré aux émotions, sur l'idée d'une correspondance entre fond et forme, entre un état intérieur et son phénomène extérieur. En effet, l'anthropomorphisme est, jusque dans sa définition, associé à des projections irréelles au sens où, relevant d'interprétations humaines de phénomènes non-humains, elles ne correspondent pas à des états effectifs de leurs objets. Pour Sherry Turkle, principale tenante de cette critique, l'intérêt que nous portons aux artefacts relationnels comme les robots sociaux est révélateur d'une « crise de l'authenticité » (Turkle, 2007). Plus précisément, cet intérêt montre, selon elle, que nous apprenons peu à peu à préférer les « connexions » (Turkle, 2018) aux relations. En effet, contrairement aux robots, « les gens sont au moins capables de réciprocité. Nous pouvons être déçus par les gens, mais au moins nous sommes déçus par un véritable potentiel » (Turkle, 2007, p.505). A rebours du parti pris phénoménologique de Coeckelbergh, Turkle associe précisément la crise de l'authenticité à un surinvestissement des apparences : craignant d'avoir à affronter l'altérité véritable, authentique, nous nous contentons d'une confortable illusion de réciprocité dans des relations simulées.

« Qu'est-ce qui explique la séduction exercée par Woebot¹⁰⁴ ? Le sentiment de vulnérabilité que ressentent les gens vis-à-vis du jugement des autres. La crainte des

¹⁰⁴ Woebot est un logiciel en ligne tenant le rôle de psychothérapeute. Il est présenté comme un « agent relationnel pour la santé mentale » (<https://woebothealth.com/>, consulté le 23/03/2022).

contacts humains. Les personnes qui utilisent ce programme évoquent leur sentiment de ne pas savoir bien faire les choses. Nous cherchons à appliquer aux relations humaines les critères de la maîtrise technologique : l'idée que les choses doivent se faire correctement et efficacement. Nous sommes prêts à aller jusqu'à parler à des machines parce que nous y voyons le moyen d'être moins vulnérables, de nous protéger des autres » (Turkle, 2018, p.230-231).

En d'autres termes, la seule altérité dont disposent sociobots et autres agents conversationnels est celle que nous sommes disposés à leur donner. Dans cette perspective, on comprend que le concept d'artefact relationnel est porteur d'une critique vigoureuse de l'importance que nous accordons à la simulation. Turkle conçoit la relation en un sens à la fois plus fort et plus strict que Coeckelbergh et Gunkel. Il ne s'agit pas non plus d'une perméabilité réciproque des êtres et de leurs propriétés comme chez Latour. La relation est fondamentalement *incarnée* et renvoie à une *réciprocité émotionnelle*, qu'il s'agisse du partage d'un même sentiment (un amour, une amitié réciproque), d'une réponse elle aussi affective, mais par une émotion différente (répondre à la colère par la tristesse par exemple) ou encore d'une disposition à accueillir les émotions de l'autre (dans la relation psychothérapeutique par exemple). Le fond du problème est alors le suivant :

« Quand nous imaginons un robot comme un vrai compagnon [...], ce qui manque au premier chef est l'altérité : la possibilité de voir le monde à travers les yeux d'autrui. Sans altérité, l'empathie devient impossible. »¹⁰⁵

Le danger est ici de nous rendre « allergique » à l'altérité et à l'engagement émotionnel, parfois inconfortable, qu'elle suppose. Une exposition répétée aux simulacres affectifs accroît le risque d'un mauvais développement de l'empathie chez l'humain et, conséquemment, celui de l'incapacité à accepter l'autre pour vivre ensemble. On pourrait toutefois rétorquer, comme le fait Catherine Malabou, que l'inauthenticité n'est pas le propre des robots et que « la plupart du temps, nous humains, nous simulons nos émotions dans nos échanges sociaux. ». Selon la philosophe, « nous [refuserions] de voir que, la plupart du temps, nous nous comportons comme des robots » (Mao, 2018). La comparaison nous paraît toutefois assez superficielle, car il s'agit là de deux formes de simulations bien différentes. En effet, les robots

¹⁰⁵ Turkle, S. (2015). *Reclaiming Conversation*, New York, Penguin, p.100, citée in (Turkle, 2018).

simulent l'affectivité *elle-même*, c'est-à-dire le fait d'avoir des émotions, alors que nous autres, humains, simulons *certaines* émotions pour masquer nos véritables états affectifs¹⁰⁶. Autrement dit, les simulations humaines sont des *dissimulations* : elle relève d'une intentionnalité qui n'existe pas chez le robot. Ce concept de dissimulation nous semble particulièrement utile pour penser l'éthique des artefacts relationnels en général et des sociobots en particulier : à quel moment une technique devient-elle dissimulatrice et comment ? La réponse à cette question est à chercher du côté de la conception et des intentions humaines qui lui président.

On peut toutefois nuancer cette critique en signalant que, du point de vue cognitif, l'anthropomorphisme ne se laisse pas saisir adéquatement selon les catégories de l'authentique et de l'inauthentique ou encore du réel et de l'irréel. Comme l'expliquent Dumouchel et Damiano,

« [I]es énoncés anthropomorphiques ne devraient pas être principalement compris comme des énoncés descriptifs, mais comme des énoncés pragmatiques dans le contexte de l'interaction. En tant que telle, la projection n'a pas besoin de reposer sur l'attribution d'états mentaux et n'implique pas non plus une quelconque fausse croyance. [...] Les projections anthropomorphiques spontanées se manifestent lorsque nous réalisons que nous avons affaire à une entité devant être expliquée par rapport à elle-même, et non simplement à l'égard de nos propres buts et objectifs. Plus précisément, elles correspondent à la prise de conscience que nous interagissons avec une entité qui, dans une certaine mesure, détermine elle-même son comportement – un agent. [...] L'anthropomorphisme est principalement un outil pour interagir, pas une description du monde » (Damiano & Dumouchel, 2018, p.5-6).

Nous pouvons développer plus amplement cette idée à partir de la notion de « pacte anthropomorphique » proposée par Denis Vidal. Celle-ci renvoie à

« toutes les instances d'interaction entre des robots [...] et leurs utilisateurs fondées sur l'apparence anthropomorphique des premiers, sans pour autant postuler que ceux qui interagissent avec eux ne devraient succomber à aucune forme d'illusion cognitive caractérisée » (Vidal, 2012, p.70).

¹⁰⁶ Véritable au sens d'effectivement ressentis.

Bien que Vidal ne le mentionne pas, la notion de pacte anthropomorphique rappelle celle de « pacte autobiographique » proposée pour en 1975 par Philippe Lejeune. Selon lui, la possibilité de l'autobiographie repose sur un pacte, explicite ou non, entre l'auteur et son lecteur. Cet engagement consiste, côté auteur, à livrer pleinement la vérité de son intimité et, côté lecteur, à faire preuve d'un jugement loyal quant à cette mise à nu et aux éventuelles altérations que la mémoire peut lui faire subir. Ni duperie d'un côté, ni crédulité de l'autre, donc ; voilà le contrat qui rend l'autobiographie possible. Le pacte anthropomorphique procède d'une logique similaire : l'interactant humain sait qu'il n'a pas affaire à un humain ou à un vivant véritable, mais il accepte, plus ou moins consciemment, l'ambiguïté statutaire du robot pour interagir avec lui. Ainsi, l'interaction est au pacte anthropomorphique ce que la mise à nue est au pacte autobiographique¹⁰⁷.

Dans cette perspective, la proposition de Vidal a ceci d'intéressant qu'elle invite à « habiter le trouble » provoqué par le sociobot et évite de précipiter l'analyse onto-anthropologique en associant trop rapidement l'IHM à une forme d'animisme. Dans *Par-delà nature et culture*, Philippe Descola (2005) caractérise l'ontologie animiste comme un rapport au monde où l'ensemble des entités dites non-humaines se voient attribuer une intériorité similaire à la nôtre, bien qu'elles se différencient (de nous, mais aussi les unes des autres) par leurs physicalités hétérogènes. Or, il n'est pas évident que les projections anthropomorphiques dont usent les roboticiens pour créer des ISHM passent nécessairement par l'attribution effective d'une intériorité. Bien qu'il fût longtemps décrit dans les termes d'une immaturité cognitive, l'anthropomorphisme est aujourd'hui considéré comme une structure aussi normale qu'universelle de la cognition humaine (Airenti, 2012 ; Vidal, 2016 ; Damiano & Dumouchel, 2018). Selon Gabriella Airenti, il trouve son origine dans la structure même de notre cognition dont il révèle la nature téléologico-dialogique. Téléologique, d'une part, en ce que nous sommes poussés à interpréter les actions des autres (humains comme non-humains) en termes intentionnels ; dialogique, d'autre part, car les bases intersubjectives de nos

¹⁰⁷ L'idée de pacte anthropomorphique pourrait d'ailleurs faire l'objet d'une enquête de sociologie interactionniste. La question ne serait plus tant d'identifier notre position ontologique à l'égard des sociobots que d'étudier le(s) rôle(s) que nous adoptons à leur égard et, plus particulièrement, les stratégies élaborées pour entretenir l'interaction. L'examen attentif de ce(s) rôle(s) pourrait nous renseigner sur la façon dont nous percevons et nous positionnons à l'égard de ces entités. Dans cette optique, il serait intéressant d'observer si l'artefact se voit attribuer une *face* au sens goffmanien par l'interactant humain.

subjectivités sont celles de l'interaction spécifique qu'est le dialogue. Définissant ce mode de communication par la prise de parole à tour de rôle, Airenti affirme notre

« tendance innée à interagir avec [nos] semblables en utilisant la forme du dialogue dont la caractéristique principale est d'être un format qui conduit à attribuer du sens à toute forme de contenu qui y est inclus, qu'il soit simple ou complexe » (Airenti, 2012, p.48).

Dans cette perspective, le dialogue n'est pas limité à sa dimension verbale, mais s'ouvre à l'ensemble des formes de langage. « Ce n'est [...] pas le contenu qui donne sa forme au dialogue mais bien la forme de l'interaction » affirme Airenti (*ibid.*) en illustrant cette idée par l'exemple des échanges entre parents et nourrissons. Bien que les nouveau-nés « ne répondent que par un répertoire de sons et de gestes très limités : sourires, gazouillis, grimaces », le sens irrigue de part en part ces « proto-conversations »¹⁰⁸. En témoigne l'expérience de la « *still face* » (visage impassible) dans laquelle on demande à une maman d'interrompre subitement une situation de dialogue avec son bébé de six mois en adoptant une expression faciale neutre. Le nourrisson, très perturbé par cette inexpressivité qu'il ne comprend pas, va alors chercher à ramener sa mère dans l'échange en attirant son attention par divers moyens. De là, Airenti propose une définition relationnelle de l'anthropomorphisme comme opération de l'esprit consistant à « mettre un objet ou un animal dans la position d'un interlocuteur au sein d'un dialogue » (*ibid.*, p.49-50).

L'ancrage téléologico-dialogique du phénomène anthropomorphique viendrait expliquer tant la possibilité que la plausibilité de nos interactions avec les machines. Prenons, ici, l'exemple du célèbre programme informatique ELIZA, conçu en 1966 par Joseph Weizenbaum au MIT. Considéré comme l'un des premiers agents conversationnels (écrit), ELIZA simulait le rôle d'un psychothérapeute rogerien en posant des questions simples telles que « Parlez-moi de votre père », en reprenant les derniers mots de son interlocuteur ou en effectuant des relances type « Pourquoi dites-vous que... ». Contre toute attente, l'artefact rencontra un vif succès parmi les étudiants et autres chercheurs du MIT. Certains allaient jusqu'à entretenir de longues et intimes conversations avec lui. Tous savaient pourtant qu'ils « conversaient » avec un programme informatique dénué de tout intérêt autre que simulationnel à leur égard. Le succès du chatbot ne peut donc pas s'expliquer par une croyance émanant d'une illusion

¹⁰⁸ Notion proposée par Mary Catherine Bateson, citée dans (Airenti, 2012).

cognitive. Conformément à son hypothèse, Airenti analyse la situation dans le cadre de l'interaction dialogique. Outre leur intrigante nouveauté, les interactions proposées par ELIZA offraient suffisamment de prises au réflexe téléologico-dialogique des participants pour qu'ils perçoivent le chatbot comme un interlocuteur viable et ce en dépit de sa nature rudimentaire.

« Même si le contenu de l'interaction est pauvre, l'effet ne l'est pas, car l'interaction entre l'ordinateur et l'utilisateur se fait sous la forme d'un dialogue, et c'est ce format qui lui donne son sens » (*ibid.*, p.50).

Dans cette perspective, l'anthropomorphisme semble agir comme un filtre cognitif d'accès au monde préexistant à toute forme d'interprétation construite de ce dernier. Il n'a donc rien d'irrationnel au sens d'adhésion irréfléchie à une fausse croyance. D'après la chercheuse,

« cette attribution de facultés humaines à l'objet, qu'elle s'exprime en paroles ou en actes, ne peut se produire que dans le cadre d'une interaction. On peut difficilement imaginer en effet que quelqu'un *pense* qu'une machine à café est un être humain, alors qu'on n'est pas surpris de voir une personne *traiter* une machine à café comme si elle avait affaire à un être intentionnel. Nous pouvons très bien dire à la machine à café : "Allez, fais vite, je suis en retard !" Ou à notre voiture : "S'il te plaît, ne t'arrête pas maintenant !" ou bien "Où t'es-tu cachée ? Je suis pourtant sûr de t'avoir garée ici hier soir." Mais cette interaction verbale ne suppose pas une croyance stable concernant l'ontologie des objets[...]. Personne ne pourrait affirmer sérieusement : "Ma voiture est une folle qui se cache quand je la cherche." Mais on peut facilement imaginer que cette phrase trouve sa place dans une interaction » (*ibid.*, p.46).

Airenti situe donc son propos dans un cadre d'analyse interactionniste. L'anthropomorphisme ne consiste pas à attribuer une intériorité (qu'on la nomme âme, esprit ou encore conscience) aux non-humains, mais plutôt à les « interacter » en les projetant dans notre référentiel téléologico-dialogique. L'anthropomorphisme n'est donc ni un animisme ni un irrationalisme¹⁰⁹, mais un interactionnisme.

A cet égard, il est souvent reproché aux critiques comme celle de Turkle de surgénéraliser les risques qu'elles identifient et, en un sens, d'essentialiser négativement l'anthropomorphisme et la robotique sociale. Pour Dumouchel et Damiano, « cette équation simple entre

¹⁰⁹ Distinction permettant, au passage, de ne pas réduire le premier au second.

"simulation" et "imposture" est non seulement incapable de rendre compte de différences éthiques fondamentales, mais tend aussi à les appréhender incorrectement » (Damiano & Dumouchel, 2018, p.7). Les auteurs prennent alors deux exemples : celui d'un sociobot conçu pour aider les enfants atteints de TSH à développer leurs compétences sociales et celui d'un robot « sexuel » qui intégrerait une option « viol ». Tous deux remplissent leur fonction notamment grâce au phénomène anthropomorphique et pourtant, il existe entre eux une différence aussi fondamentale qu'irréductible. Dans le premier cas, le robot intervient pour encapaciter des personnes vulnérables et relève donc du soin, alors que dans le second, il s'agit de banaliser un fantasme d'une violence et d'une cruauté inqualifiables au motif que, s'exerçant sur une entité dépourvue de sensibilité, il ne serait pas réel et donc pas condamnable moralement ni légalement¹¹⁰. Selon Dumouchel et Damiano, la critique radicale et systématique de l'anthropomorphisme sous l'angle de la simulation ne peut que passer à côté de ces différences décisives et, par conséquent, « condamne l'éthique à l'impuissance » (*ibid.*)¹¹¹. Pour éviter cela, les auteurs proposent d'utiliser l'anthropomorphisme comme un outil d'« enquête éthique », c'est-à-dire de partir des effets qu'il produit pour comprendre ce qui se joue dans les ISHM et les réguler à partir des observations effectuées empiriquement et scientifiquement. Cette approche qualifiée d'*éthique synthétique* intègre la science à l'enquête éthique au sein d'une démarche expérimentale (« *understanding by building* »). En d'autres termes, l'anthropomorphisme n'est pas analysé *a priori* (que ce soit dans le sens de sa condamnation ou de sa promotion) selon des règles éthiques prédéterminées, mais *in situ*, afin de comprendre, à partir des IHM telles qu'elles se font, quels sont les problèmes effectifs qui émergent de nos échanges avec les sociobots.

« L'éthique synthétique n'exclut pas les questions traditionnelles sur lesquelles se concentrent les approches dichotomiques, mais les reformule au sein d'une perspective

¹¹⁰ Il y aurait beaucoup à dire sur ces deux exemples. Paradoxalement, une critique radicale de la simulation pourrait chercher à égaliser les deux situations en expliquant qu'étant tous deux simulés, aucun acte n'est plus réel que l'autre et, partant, plus louable ou plus condamnable moralement. En plaçant *de facto* la simulation dans le champ de l'amoralité, une telle position ne prendrait pas en compte la question très sérieuse de la vertu (bien qu'il « évacue » factuellement – mais pas symboliquement – l'atroce chosification d'autrui, le viol simulé reste, en lui-même, un avilissement absolu de soi) et, plus largement, celle des effets, très différents, entre les deux médiations robotiques.

¹¹¹ Par exemple, du point de vue de la coordination affective (cf. § 2.2.2.1.2) et de l'effacement qu'elle opère entre dimension interne et externe des émotions, « les robots sexuels intégrant une "option viol", n'offrent pas aux utilisateurs humains de "simuler un viol", mais bien de s'engager dans un viol *tout court* » (*ibid.*, p.8) alors que le robot compagnon peut créer une dynamique émotionnelle positive permettant d'encapaciter socialement l'enfant atteint de TSH.

de recherche qui voit les robots sociaux comme un moyen de donner plus de pouvoir à nos relations. Comment construire des robots qui fonctionnent comme des connecteurs sociaux, renforçant les relations humain-humain au lieu de produire de l'isolement et d'affaiblir le lien social ? Comment peut-on concevoir des robots sociaux qui facilitent, encouragent et fortifient les échanges entre humains au lieu d'offrir la possibilité d'échapper aux défis de l'interaction humain-humain, de s'en éloigner au sein d'un monde d'interactions humain-robot qui ne demandent aucun effort ? Est-il possible d'exploiter les robots sociaux pour modifier les motifs du comportement humain dans la direction de la croissance morale ? » (*ibid.*).

La nature expérimentale de l'éthique synthétique n'est pas incompatible avec l'existence de règles morales préexistantes. Ces règles sont simplement convoquées et justifiées de l'« intérieur », par l'enquête, si le problème qu'elles sont censées prévenir se pose effectivement. En outre, l'éthique synthétique est aussi l'occasion d'inventer de nouvelles règles au fur et à mesure que ses enquêtes permettront d'affiner les connaissances en matière d'interaction humain-robot. Dans cette perspective, l'éthique synthétique se présente à la fois comme une éthique *avec* le robot et comme une éthique *by design* au double sens d'éthique par conception et par vocation puisque l'enquête semble être étroitement associée à l'activité même de fabrication (pour ne pas dire qu'elle devient l'autre nom de la conception).

Transition - Du robot « social » au robot « moral »

A l'issue de cette première partie, quelle réponse peut-on apporter à notre question introductive ? Le robot social est-il avant tout un robot *tout court* ou reconfigure-t-il en profondeur ce qu'on appelle communément robot ? En tant qu'il est un artefact doté de technologies sensorimotrices lui permettant d'entretenir un certain rapport au monde, le sociobot appartient bien à la catégorie conceptuelle du robot. Or, force est de constater que cette caractérisation générale reste muette quant à la singularité du robot dit *social*. La compréhension d'un tel artefact suppose donc d'abaisser le niveau d'abstraction, pour le dire comme Floridi. Dans quelle mesure notre enquête sur le sociobot et les sens du « social » peut-elle nous y aider ? Nous avons vu qu'une grande partie de la littérature, à commencer par la littérature « pionnière », pense le sociobot comme un type particulier de robot *disposant* d'aptitudes à l'interaction sociale avec l'humain au sens de capacité à établir des formes de communications rétroactives dont l'effet est de générer, au moins ponctuellement, des comportements émergents. Cette première caractérisation de type interactionniste peut être associée à une seconde caractérisation de type fonctionnaliste. Le sociobot devient alors un robot dont la sensoriactivité est mise au service de fonctions intrinsèques ou extrinsèques à l'interaction sociale qui, en tout état de cause, se réalisent *à travers* l'interaction sociale elle-même. Cela dit, notre enquête nous a également conduit à une critique de la socialité réduite à la seule perspective des interactions et des fonctions sociales. Celles-ci composent bien le social, mais ne le résument pas. Le point de vue de l'interactance nous a permis de penser le social comme n'étant pas uniquement structuré par les interactions entre humains, mais aussi par nos divers échanges et contacts avec les non-humains comme les artefacts. Or, dans cette perspective, tous les robots sont sociaux, au même titre que tout autre artefact avec lequel nous entrons en relation. Dès lors, comment comprendre la spécificité du sociobot ? Comment expliquer cette caractéristique si particulière, qui est aussi et peut-être surtout un *effet*, de générer le sentiment d'être en présence d'un autre ? Nous pensons que cette spécificité ne saurait être appréhendée sans articuler socialité et moralité des sociobots ou plus précisément, sans penser le caractère indissociable de ces deux dimensions en sociobotique. Pourtant, socialité et moralité ne sont que très indirectement mises en relation

dans la littérature consacrée aux robots sociaux. Pour être très précis, il faut reconnaître que l'approche relationnelle du social, d'inspiration latourienne, peut déboucher sur une prise en compte de la moralité du sociobot. Rappelons que, selon Latour, toute médiation technique se caractérise notamment par un phénomène de délégation dont certaines sont porteuses de prescriptions morales. L'illustration la plus emblématique de cette idée est celle, déjà mentionnée, du ralentisseur dont le programme d'action est de prescrire le ralentissement aux automobilistes. Dès lors, le dos-d'âne apparaît comme un agent à la fois social et moral en ce qu'il renvoie à une certaine association actantielle ayant pour objectif de préserver la vie d'autrui grâce à une prescription de ralentissement. Si l'on transpose cette perspective théorique à notre objet de recherche, il est frappant de constater que l'opérativité des délégations dont un sociobot peut faire l'objet dépasse non seulement le cadre matériel du ralentisseur, mais aussi le cadre interactionnel puisque le robot peut produire des effets en dehors de ses interactions avec un humain. Nous pensons que cette opérativité s'origine dans la phénoménalité singulière du sociobot qui se manifeste comme une altérité. Plus précisément, nous soutenons que ce qui caractérise les robots sociaux en tant que tels, par-delà la diversité de leurs formes et de leurs fonctions, c'est une certaine manière de se présenter comme des « autres ». Comprendre cette caractéristique des sociobots suppose donc de clarifier le sens d'être de leur altérité.

PARTIE II

L'ALTÉRITÉ : UN CONCEPT POUR LA PHILOSOPHIE DES TECHNIQUES ?

Chapitre 4 - Présentation de Spoon et discussion des fondements conceptuels de l'activité de conception des robots sociaux

Ces grandes questions théoriques et, plus généralement, le thème de l'altérité robotique trouvèrent un écho particulier chez le dirigeant de Spoon pour qui la conception technique va de pair avec sa conceptualisation. Dans ce chapitre, nous proposons d'ancrer notre problème dans le contexte spécifique de Spoon à travers la discussion de son univers conceptuel et de son contexte économique. Précisons d'emblée que la société ne peut plus, aujourd'hui, être considérée comme une entreprise de robotique sociale. En effet, quelques temps avant la fin de notre contrat, les dirigeants décidèrent d'orienter l'activité de Spoon vers la conception de personnages interactifs *numériques* et de délaisser, au moins temporairement, la dimension robotique des premiers temps. Notre recherche ne porte pas sur cette seconde période, que nous avons peu connue, mais sur celle précédant ce « virage numérique ». Un tel infléchissement s'explique surtout par les difficultés économiques inhérentes à la sociobotique marchande (cf.4.3.1) dont le rapport coût/bénéfice pour les clients se révèlent généralement désavantageux et par la volonté de mieux correspondre, au moins dans un premier temps, aux « besoins » de ses partenaires commerciaux, très centrés sur la dimension numérique de l'« expérience client ». A cet égard, la matérialité mouvante du robot représente plus un coût inutile qu'une réelle valeur ajoutée.

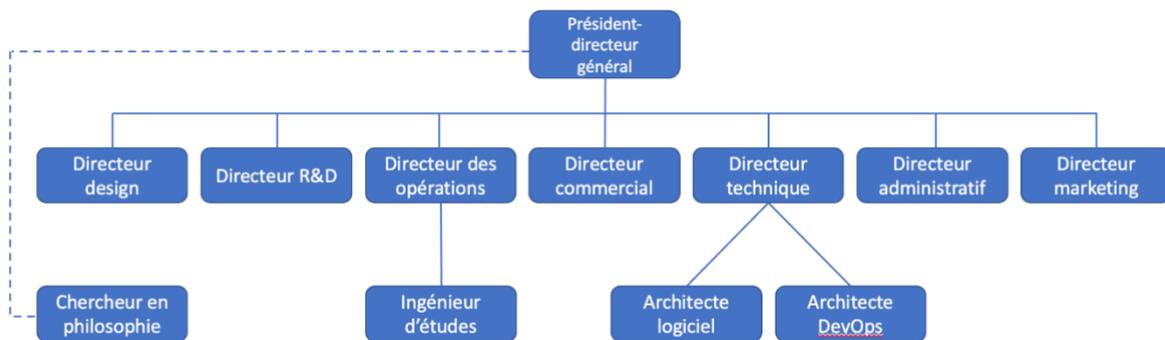
4.1 L'entreprise Spoon : objet social, organisation, produits et valeurs

Spoon est une société par actions simplifiée (SAS) créée en décembre 2015 par son président directeur général (PDG), Jérôme Monceaux, et ses deux associés fondateurs, Thibault Hervier et Aymeric Masurelle, tous trois ingénieurs de formation. L'entreprise est née de l'expérience passée de son PDG au sein d'Aldebaran Robotics où, pendant une dizaine d'années, il contribua à développer les robots NAO et Pepper. Spoon a pour objet social :

- « la conception, le développement, la production, le design, le marketing et la vente de robots interactifs programmables et autonomes ;

- toute prestation de conseil et d'assistance en matière de conception, de développement, de production, de design, de marketing et de vente de robots interactifs programmables et autonomes ;
- l'organisation de conférences, d'événements et la dispense d'enseignements en lien avec la conception, le développement, la production, le design et le marketing et la vente de robots interactifs programmables et autonomes »

Notons que Spoon se présente statutairement et légalement comme une entreprise de robotique interactive et non de robotique sociale. Cette question de vocabulaire cache un enjeu majeur justifiant en partie notre recrutement par l'entreprise : celui de la conceptualisation et, avec elle, de l'orientation de son activité. Nous y reviendrons bientôt. Durant nos trois années de contrat CIFRE, l'équipe fut, à quelques évolutions près, composée selon l'organigramme suivant :



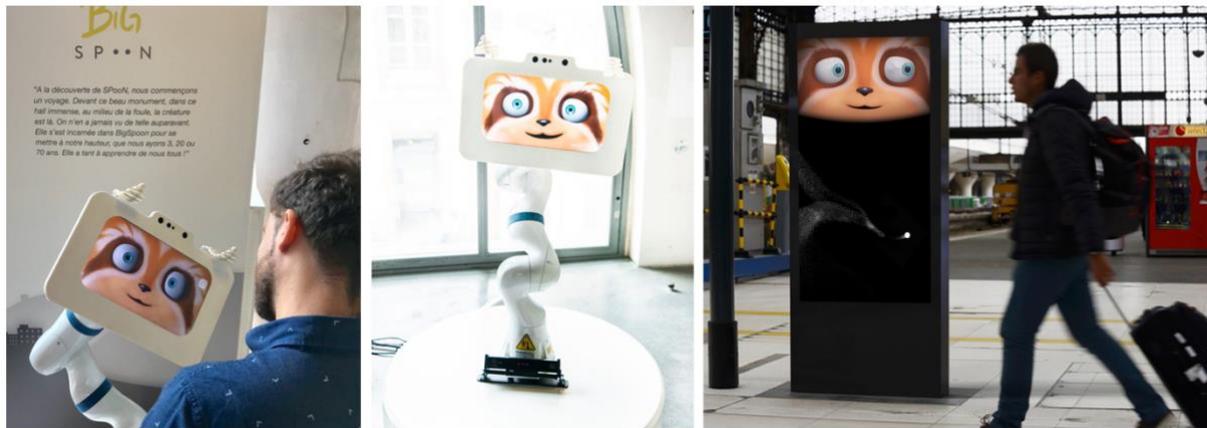
Spoon propose à la vente ou à la location deux produits différents : BigSpoon et TotemSpoon. Le premier est un bras cobotique¹¹² industriel Kuka¹¹³ détourné de son usage pour accueillir un ordinateur tactile affichant un visage animal en 3D et surmonté de deux oreilles imprimées en polyamide. Le second est une borne tactile fournie par la société Acrelec¹¹⁴ sur laquelle apparaît le même visage 3D. Chacun intègre des capteurs de quatre types différents (caméra,

¹¹² Le terme de cobotique est issu de la contraction de « robotique collaborative » ou « coopérative ». Les auteurs du néologisme définissent le cobot comme un « robot conçu pour des interactions physiques directes avec un opérateur humain, au sein d'un espace de travail partagé » (Peshkin & Colgate, 1999).

¹¹³ Il s'agit du modèle LBR iiwa.

¹¹⁴ Acrelec est un fournisseur d'équipements industriels dédiés à l'expérience client de la restauration rapide et de la vente de détail (*retail*).

microphone, profondeur) et des algorithmes associés (notamment de captation faciale¹¹⁵) lui octroyant une certaine perception de son environnement. L'activité technique de Spoon est principalement axée sur la conception du logiciel d'IHR (souvent qualifiées de « naturelles »¹¹⁶) qui « anime » ses machines. Le matériel, quant à lui, est presque intégralement acheté à divers fournisseurs puis assemblé par l'équipe.



A gauche et au centre : BigSpoon / A droite : TotemSpoon (vue d'artiste)

L'ambition applicative de l'entreprise et la stratégie produit qui en découle concernent ce que son dirigeant nomme les « espaces collectifs » (gares, centres commerciaux, halls d'accueil, abribus, musées, magasins etc.). L'idée est de favoriser l'introduction et le développement de services numériques dans ces espaces par l'intermédiaire de sociobots dont la vocation, dans le même temps, est de personifier les lieux en question. A l'heure actuelle, Spoon n'a donc pas l'intention d'investir le marché de la sociobotique personnelle.

¹¹⁵ La captation faciale n'est pas une *reconnaissance* faciale au sens où nous l'entendons généralement. Un algorithme de captation faciale ne fait que repérer la forme « visage » dans son environnement sans l'associer à une quelconque identité.

¹¹⁶ La notion d'interaction naturelle est une formule conceptuellement problématique souvent utilisée en robotique. Elle désigne les interactions verbales et corporelles (non manipulatoires) durant lesquelles s'effectuent des échanges sémantiques. Elle correspond, en réalité, à l'interaction sociale. La naturalité supposée de ces échanges semble renvoyer à quelque chose comme un automatisme inné, libéré des contraintes de l'apprentissage et de l'effort, contrairement aux interactions non naturelles dont un exemple type est l'utilisation du clavier d'ordinateur. Celle-ci suppose en effet d'apprendre et de retenir où se situent les lettres pour composer un texte. Or, on sait bien que l'interaction sociale demande elle aussi un apprentissage dont les acquis peuvent être renégociés lorsque, par exemple, nous changeons de culture. En outre, certains ou certaines d'entre nous se diraient sûrement plus à l'aise avec un clavier qu'avec autrui au sens où cela demande moins d'effort (cf. la critique proposée par Sherry Turkle à propos des difficultés relationnelles nourries par les TIC). Ces personnes seraient donc plus « naturellement » tournées vers les claviers que vers autrui ? Voilà qui paraît pour le moins douteux.

Peu de temps après le début de notre contrat à l'automne 2017, un travail d'identification des valeurs de la société fut conduit par l'équipe dirigeante (i.e. les trois associés fondateurs et le directeur de la R&D) et donna naissance aux quatre grands axes suivants¹¹⁷ : 1. *Intensely human* (« intensément humain »), 2. *Clever together* (« intelligents ensemble »), 3. *Embrace the marvelous* (« embrasser le merveilleux »), 4. *Be mindful* (« être attentionné »). De façon plus prosaïque, on peut ramener ces formules aux notions respectives d'humanité, d'intelligence collective, de merveilleux et d'attention soucieuse. Dans l'optique pragmatiste proposée par John Dewey (2011) où les valeurs se présentent à la fois comme « ce à quoi nous tenons » et « ce *par quoi* nous tenons », nous avons proposé d'entamer un travail collectif de définition de ces notions, notamment afin de clarifier et d'évaluer, empiriquement, leur rôle au sein de la conception. Cette proposition fut toutefois rejetée par le PDG de Spoon au motif que l'acte de nommer les quatre valeurs constituait la première étape d'un processus de recherche identitaire qu'il ne fallait pas entraver par un enfermement définitionnel. En définitive, ce travail ne fut jamais mené et l'axiologie de Spoon, au sens littéral de « discours sur les valeurs », se limita principalement aux communications orales de son PDG.

4.2 Genèse et répertoire conceptuel de Spoon

4.2.1 La « récurrence d'intérêt » des sociobots

L'ensemble de l'activité de Spoon est adossé à une problématique centrale dont dépend sa viabilité commerciale à long terme : celle que Monceaux nomme la « récurrence d'intérêt » des robots. Le constat du dirigeant à l'issue de la décennie passée chez Aldebaran est le suivant : la robotique dite sociale échoue à produire autre chose que des entités de laboratoire car l'intérêt que les humains leur portent n'est que très éphémère. Si les robots témoignent d'un indéniable pouvoir d'attraction, ce pouvoir ne procède pas tant de ce qu'ils sont ou font que de leur nouveauté et de leur sophistication technique. L'émerveillement qu'ils suscitent auprès du public cède ainsi très rapidement le pas à l'indifférence. Dès lors, comment faire en sorte que l'intérêt pour un sociobot se pérennise ? Dans quelle mesure cette quête de

¹¹⁷ Notons que ces valeurs sont pensées comme relevant des caractéristiques essentielles de Spoon et, ce faisant, ne doivent pas concerner uniquement la conception, mais l'organisation de l'entreprise dans son ensemble (management, relation client etc.).

récurrence d'intérêt invite-t-elle à penser à nouveau frais l'artefact qu'est le robot social ? La notion de récurrence d'intérêt, on le voit, superpose aux enjeux de conception un fort enjeu de conceptualisation. C'est donc sur la base de cette nécessité de (re)penser son objet que Spoon fit appel à nous.

4.2.2 Une appréhension phénoménologique du sociobot comme « reflet d'existence de l'utilisateur »

La réflexion conceptuelle était toutefois loin d'être vierge à notre arrivée. Selon Jérôme Monceaux, le défaut de récurrence d'intérêt des robots s'explique notamment par la mauvaise qualité des interactions qu'ils proposent. Plus précisément, et c'est peut-être là son originalité, l'ingénieur explique que le design des sociobots néglige généralement de prendre en considération ce qui constitue le cœur des interactions sociales, à savoir le sentiment d'exister aux yeux de notre interlocuteur.

« Je suis très fier des robots qu'on a conçus par le passé, nous confie-t-il en entretien, mais il leur manquait le principal. Malgré le travail faramineux qu'a représenté leur design, leur visage était bien trop inexpressif, trop fixe. Face à un NAO ou un Pepper, il est difficile de savoir ce que le robot regarde vraiment. Est-ce qu'il me regarde moi ou regarde-t-il le pot de fleur derrière ? Peut-être même qu'il ne regarde rien en particulier. Le fait de ne pas savoir si on "existe" aux yeux de la machine perturbe considérablement nos interactions avec elle et ça finit par provoquer un décrochage, un désintérêt. »

Le parti pris inhérent à la fondation de Spoon fut donc de créer des robots capables de générer chez leurs interlocuteurs ce sentiment d'existence personnelle nécessaire à l'inscription des ISHR dans le long terme. Concrètement, cela passe par le design d'un visage 3D et, plus spécifiquement, d'un regard capable de prendre *manifestement* autrui pour objet. L'entrepreneur rattache ce choix à une expérience vécue chez Aldebaran où il vit une dizaine de NAO se tourner simultanément vers lui suite au déclenchement de leur dispositif de localisation sonore. Il décrit cet événement comme chargé d'une singulière intensité :

« Pour la première fois de ma vie, je me suis senti regardé par une machine. C'est une expérience très étonnante voire perturbante, mais j'en ai tiré un enseignement fort :

l'enjeu n'est pas d'essayer de mettre de la "vie" dans les robots comme on tentait de le faire, mais plutôt de refléter l'existence de l'utilisateur dans les robots. »

Deux points essentiels méritent, ici, d'être soulignés. Premièrement, l'objectif de Monceaux n'est pas de conférer techniquement une quelconque « intériorité » au robot, mais de concevoir un artefact dont la médiation renvoie explicitement l'interlocuteur à lui-même. L'objection internaliste consistant à lier des capacités intentionnelles à des états mentaux peut donc être écartée puisqu'il ne s'agit pas d'attribuer au robot de telles capacités, mais de considérer, comme le propose Daniel Dennett (1990) de faire « comme si » il les possédait. Autrement dit, si l'artefact se comporte « comme si » il me regardait, il n'y a pas lieu, dans ce contexte, de lui dénier la capacité d'attention. Outre la référence à Dennett, cela n'est pas sans rappeler le principe d'« équivalence comportementale » proposé par John Danaher (2020)¹¹⁸. En conséquence, un tel discours témoigne, deuxièmement, d'une appréhension non pas technique, mais phénoménologique du sociobot. L'artefact est présenté comme une entité spéculaire dont la caractéristique essentielle est à chercher dans l'effet qu'elle produit sur nous, à savoir « refléter » notre existence en simulant un regard. Dans une intervention filmée pour la franchise TEDxParisSalon, Monceaux qualifie d'ailleurs le robot d'« être qui nous regarde » et associe cette disposition à son altérité :

« Je pense que la caractéristique qui manque [aux robots qu'on a conçus par le passé] c'est l'altérité, c'est-à-dire de pouvoir sentir cette intelligence, sentir ce personnage exister devant nous, avoir une interaction avec lui. [...] Nous sommes des êtres relationnels, des êtres sociaux, des êtres qui avons besoin des autres pour être. En fait, on a besoin de se sentir exister aux yeux des autres. Et là, aujourd'hui, avec toute la technologie qu'on a, on a les moyens de pouvoir se sentir exister aux yeux d'une machine. C'est quand même fou ! En fin de compte, ce serait peut-être ça un robot : le robot serait cet être qui nous regarde, ce serait une espèce nouvelle que nous créerions et qui s'installerait comme notre reflet » (TEDx Talks, 2019).

L'emploi du terme de « personnage » n'est pas anodin, bien que Monceaux ne lui donne pas de dimension conceptuelle. Outre qu'il peut évoquer le monde de l'animation et son imaginaire, le terme permet de contourner, au moins verbalement, le risque de confusion

¹¹⁸ Voir p.142.

ontologique tout en conservant l'impression de vie qu'il suscite. En effet, la notion de personne reste, dans le langage courant, très étroitement associée à l'humain.



Au-delà de cette remarque, ce passage, quoique court et abstrait, condense une importante partie de la réflexion du dirigeant sur la robotique sociale. A ce titre, au moins deux développements sont nécessaires pour le préciser et pour le mettre en perspective.

4.2.3 Du « robot social » au « créatures merveilleuses » en passant par les « créatures artificielles » : nommer pour comprendre, comprendre pour nommer

4.2.3.1 Le « vrai robot social »

Premièrement, signalons que la notion de « robot social » est en réalité très peu utilisée chez Spoon (elle n'est d'ailleurs pas mentionnée dans son objet social, comme nous l'avons déjà noté). A cet égard, il nous est apparu que Monceaux s'intéressait plus au concept d'altérité qu'à celui de socialité dont il semblait déjà avoir déterminé le sens. Lors d'une rencontre que nous avons organisée entre le psychiatre Serge Tisseron et lui, l'entrepreneur expliquait la chose suivante :

« on parle également de robots "sociaux". [...] Jusqu'à présent, les robots ont été écrits d'une façon un peu prédéterminée, c'est-à-dire que le concepteur va établir le schéma que le robot va suivre et auquel il va se conformer [...] et moi, je ne trouve pas ça très intéressant parce que, [...] au fond, le robot, il ne parle pas des gens avec qui il vit, il parle toujours d'autre chose. Pourtant, ce qui est intéressant, ce n'est pas de pouvoir remplacer la présence d'un Homme, ça n'a absolument aucun intérêt, on veut un monde avec des Hommes, pour aller dans le sens de l'amplification des interactions ou, en tous cas, de la promotion des interactions entre les personnes ; et ce que je trouve intéressant, justement, c'est que le robot soit l'écho des autres, qu'il soit vraiment le promoteur de l'interaction. Et donc, pour conclure là-dessus, on a pris soin, dans le cadre de Spoon, de travailler un sujet qu'on appelle la "programmation par interactions naturelles", c'est-à-dire que le robot va apprendre de son interlocuteur, des mots, des phrases, des comportements qu'il va resservir, parfois en commettant des erreurs, mais avec un apprentissage par renforcement, mais collectif, c'est-à-dire qu'une communauté va moduler le comportement du robot et, on espère, faire émerger un comportement qui soit l'image d'un collectif, peut-être le reflet du collectif » (Le Mouton Numérique, 2018).

Et l'ingénieur d'acquiescer aux propos ultérieurs de Serge Tisseron :

« le vrai robot social, c'est le robot qui met en relation les gens. [...] Hélas, le robot social qui s'annonce, c'est le robot de *substitution* sociale. [...] Pour moi, il n'y aura de vrai robot social – que j'appelle humanisant, créateur de lien - que lorsque les fabricants se résoudront, enfin décideront de faire en sorte que le robot soit conçu pour favoriser les liens humains, mais pas pour se substituer à eux » (*ibid.*).

Pour résumer, un robot est social lorsque ses capacités d'interactions (sociales) sont mises au service de la promotion voire de l'intensification des relations entre humains. La difficulté est ici de qualifier la nature de ces relations. En effet, Monceaux parle d'« interactions » là où Tisseron, lui, fait référence aux « liens humains ». Ces notions, souvent utilisées de façon interchangeable dans le langage courant, peuvent pourtant occulter des divergences conceptuelles tacites ou des impensés si, précisément, leur interchangeabilité n'est pas questionnée. Le lien social n'est-il pas la relation qui perdure *au-delà* des situations d'interactions ? N'est-il pas à la fois ce qui cause et ce qui résulte des interactions ? Prenons un exemple concret illustrant cette idée que lien et interactions sociales se nourrissent mutuellement : une collègue vous appelle par téléphone pour vous annoncer la mort d'un

confrère. Étant lié à cette collègue professionnellement voire amicalement, la relation que vous entretenez ne pas prend fin en même temps que votre interaction téléphonique. Celle-ci peut avoir une multitude d'influences sur votre lien, mais elle ne l'interrompt pas une fois terminée¹¹⁹. Suite à cette triste conversation, vous décidez d'envoyer un message de soutien à une autre collègue qui connaissait et appréciait beaucoup ce confrère. Celle-ci vous appelle sans tarder pour vous confier ce qu'elle ressent de vive voix. Dans cet exemple, on voit que, selon les points de vue, c'est l'interaction qui vient nourrir le lien ou le lien qui vient nourrir l'interaction. Ici, nous avons pris un exemple professionnel et amical, mais le lien social s'exprime et s'entretient aussi à travers les diverses conventions qui régissent les interactions « impersonnelles » entre inconnus comme celles que l'on peut avoir dans le métro, dans la rue ou encore dans un magasin. Cela étant dit, on voit que la distinction entre lien social d'une part et interactions sociales de l'autre permet de questionner en profondeur le type de sociobot que l'on souhaite concevoir. Dans cette perspective, le robot social avec lequel nous interagissons pour obtenir des informations dans un magasin et celui qui nous rappelle l'anniversaire d'une proche (pour prendre un exemple cité par Tisseron dans la vidéo et que Monceaux donne souvent lui-même) ne sont pas « sociaux » de la même manière¹²⁰. Dans ce second cas, l'ISHR est mise au service de l'interaction et du lien entre deux humains alors que, dans le premier, l'interaction avec le robot n'appelle à pas à autre chose qu'elle-même et peut, dans la logique de Tisseron, être qualifiée de substitutive. Précisons toutefois que rien n'empêche *a priori* une ISHR de substitution d'entraîner, de façon contingente, une interaction entre humain (lorsque le robot fait parler de lui pour telle ou telle raison, par exemple) ; mais un design se donnant pour objectif de favoriser aussi systématiquement que possible les interactions entre humains par le biais du robot ne saurait construire son approche autour de cette seule contingence. La position de Tisseron pose donc, en filigrane, la question de savoir comment concevoir les sociobots afin qu'ils nourrissent le lien social entre humains plus qu'ils ne l'appauvrissent. Ainsi explicitée, la position du psychiatre quant

¹¹⁹ A moins, bien sûr, que la conversation débouche sur une dispute causant votre rupture définitive. Or, même dans ce cas, n'est-ce pas la fin des interactions plutôt que celle du lien lui-même qu'entraînait une telle dispute (au moins dans un premier temps) ?

¹²⁰ Étant entendu que ces deux formes de socialité peuvent *a priori* être prises en charge par un seul et même robot.

aux « liens humains » semble correspondre à l'objectif que Monceaux assigne à Spoon en parlant d'« amplification » et de « promotion des interactions entre les personnes »¹²¹. Ces précisions étant données, notons à présent que cette approche partagée par les deux interlocuteurs relève d'une conception moralement normative de la socialité robotique. En effet, le « vrai robot social » dont parlent explicitement Tisseron et implicitement Monceaux est à prendre au sens de *bon* robot social ou de robot social *moralement désirable* (chez Spoon, la non-substitution était régulièrement évoquée comme composante pratique de la valeur « intensément humain »). Autrement dit, être et devoir-être du sociobot sont assimilés l'un à l'autre puisque pour être vraiment social, un robot *doit* favoriser le lien entre humains. Dans cette logique, un robot dit social dont les effets, recherchés ou non, relèveraient de la substitution à un humain serait un robot *anti*-social car désocialisant. Une telle approche n'est pas sans rappeler celle de Turkle et, avec elle, les difficultés liées à sa trop grande abstraction. En effet, elle semble attribuer un contenu nécessairement positif aux interactions humaines et un autre nécessairement négatif à la substitution, ce qui ne saurait aller de soi par principe. Une posture aussi radicalement dichotomique n'est possible que si l'on envisage les ISHM sous le seul angle de la tromperie et de l'inauthenticité. En outre, cette position présente les travers de la conception du social comme propriété dispositionnelle (cf. 2.2.2) et oublie qu'un même robot, en fonction du milieu auquel il est intégré et des personnes qui interagissent avec lui, ne produit pas les mêmes effets « sociaux ». L'exemple classique est ici celui, déjà cité, du Paro pouvant à la fois inciter les hôtes d'une maison de retraite à interagir entre eux pour parler *de* lui et conduire une des pensionnaires à se replier sur elle-même *avec* lui (Wright, 2018). L'un des écueils potentiels de cette posture est d'occulter la question, certes épineuse, de la qualité des interactions humaines en termes de contenu au seul profit de leur quantité. L'exemple des réseaux eux aussi qualifiés de « sociaux » est, à cet égard, particulièrement instructif. Cela dit, nous n'affirmons pas que Tisseron ou Monceaux préconisent une pure inflation interactionnelle entre humains. Nous cherchons plutôt à pointer une lacune dans les discours empreints de ce que nous pourrions qualifier d'« humanisme interactionniste », à savoir l'absence de réflexion – qui est aussi une absence de prise de risque théorique – quant à ce qui permet de déterminer la désirabilité morale des interactions¹²². En tout état de cause, les

¹²¹ Nous verrons, au regard des développements actuels de Spoon, qu'une telle position de principe peine à trouver un modèle d'affaire correspondant.

¹²² A cet égard, il est intéressant de constater que Spoon n'a fixé que des indicateurs quantitatifs, comme la durée d'interaction par exemple, pour évaluer les IHR avec ses produits.

réponses à cette question ne peuvent pas, elles non plus, être formulées « dans l'absolu », à moins de reconduire le problème qui les a suscitées. Au contraire, c'est dans le contexte d'un examen attentif des situations singulières et concrètes que de telles réponses doivent être trouvées, à l'instar de ce que préconise l'éthique synthétique de Dumouchel et Damiano.

Revenons plus spécifiquement à Spoon. Outre la correspondance entre le discours de Monceaux et une certaine demande sociale¹²³, au moins deux grandes raisons expliquent le caractère abstrait de son positionnement. La première tient au parti pris initial de l'entreprise dont le projet repose avant tout sur l'idée qu'il est nécessaire d'améliorer les ISHM *en elles-mêmes* avant de chercher à les appliquer à un secteur et/ou à une problématique précise. Dans cette perspective, l'approche de Spoon peut être interprétée dans les termes d'une recherche transcendantale des conditions de possibilité de la *bonne* ISHM. Une telle idée se veut normative non seulement sur le plan technique (i.e. la mécanique en question réalise effectivement ce pour quoi elle a été imaginée et conçue), mais aussi sur le plan moral. En effet, nous avons souvent été frappés par le lien que Monceaux semble tisser entre dimension fonctionnelle et dimension éthique du regard de ses artefacts. Être et devoir-être du sociobot se rencontrent à nouveau dans l'idée que la fonction interactive et « interactivante »¹²⁴ du regard se double d'une valeur éthique de reconnaissance de l'utilisateur humain comme *sujet*. Autrement dit, le regard du robot est investi d'une valeur à la fois intentionnelle et attentionnelle – certes, par « délégation » de la part des concepteurs comme dirait Latour – permettant à l'utilisateur de se sentir *reconnu* comme sujet existant en même temps que *perçu* comme objet présent. Fort de cette idée directrice, le dirigeant se forge la conviction que l'« intentionnalité simulée »¹²⁵ est à la fois fonctionnellement nécessaire aux robots et moralement souhaitable aux humains.

La seconde raison tient à la posture libérale dont se réclame l'entrepreneur et qui, au-delà des grandes normes sociales et/ou morales admises, l'empêche de décider pour autrui de ce qui constitue une interaction sociale désirable. Son intérêt pour ce qu'il appelle le « collectif » découle en grande partie de cette position : ce n'est ni à lui ni à Spoon ni même aux individus de décider seuls du *devenir* du robot à travers ses interactions, mais au « collectif » d'individus. Pour l'entrepreneur, la notion de collectif doit être articulée à celle de « territoire », terme

¹²³ En effet, il est socialement plus acceptable de prôner la non substituabilité robotique que l'inverse.

¹²⁴ C'est-à-dire qui incite à l'interaction.

¹²⁵ Nous proposons cette notion pour désigner le « regard » des machines de Spoon.

qu'il utilise pour désigner le lieu physique d'implantation du robot et son périmètre d'action autant que son contexte sociologique, culturel voire historique. Par exemple, si l'artefact est installé dans une gare ferroviaire, son collectif se composera notamment des employés du site, de ses usagers et des passants. Ainsi, un robot donné ne sera pas le même robot en fonction du « territoire » où il se trouve et des différents collectifs qui s'empareront de lui. Ce credo libéral centré sur l'appropriation collective d'un artefact socialement interactif pose toutefois le problème majeur de sa régulation. A cet égard, précisons qu'il existe chez Spoon un fort décalage entre théorie et pratique puisque la société ne peut pas (se) permettre de laisser le robot apprendre tout et n'importe quoi. L'équipe s'est donc résolue à mettre temporairement en place un système de régulation fondée sur la modération humaine de chaque contenu transmis au robot¹²⁶ pour éviter les dérives comme celles associées au tristement célèbre Tay de Microsoft. Tay était un agent conversationnel expérimental développé pour apprendre de ses interactions (notamment) avec les utilisateurs et utilisatrices du réseau social Twitter et pour « tweeter » en conséquence. L'artefact pouvait également répéter les tweets des internautes si cela lui était demandé. L'expérience a vite été interrompue car, en moins de 24h, certaines interactions du chatbot avec des internautes peu scrupuleux le conduisirent à poster des messages racistes, négationnistes, misogynes ou encore complotistes. Un article publié sur *The Verge* montre bien qu'une telle expérience

« demande de répondre à de sérieuses questions, comme celle de savoir comment nous allons entraîner les IA à partir de données publiques sans y intégrer les pires traits de l'humanité ? Lorsque nous créons des bots qui reflètent leurs utilisateurs, nous inquiétons-nous de savoir si ces utilisateurs sont des ordures ? Il y a de nombreux exemples de technologies qui matérialisent – accidentellement ou intentionnellement – les préjugés de la société, et les aventures de Tay sur Twitter montrent que même des grandes entreprises comme Microsoft omettent d'instaurer la moindre mesure préventive contre ces problèmes » (Vincent, 2019).

Qu'il témoigne d'une insouciance structurelle ou d'une grande naïveté, ce manque d'attention au très délicat problème du filtrage des contenus appris, restitués ou générés automatiquement par la machine révèle le caractère illusoire d'une autorégulation des

¹²⁶ Cette décision se voulait temporaire car, de toute évidence, elle entrerait en totale contradiction avec le credo libéral évoqué plus haut. Le problème de la modération fut toutefois progressivement et durablement occulté par d'autres enjeux de développements jugés plus prioritaires.

collectifs en question. Cela n'est pas sans rappeler l'analyse de George Canguilhem fondée sur la distinction entre société et organisme. Dans l'analyse en question, le philosophe et médecin s'exprime en ces termes :

« Je m'excuse, je vais peut-être vous scandaliser, mais une société n'a pas de finalité propre ; une société c'est un moyen ; une société est plutôt de l'ordre de la machine ou de l'outil que de l'ordre de l'organisme. [...] Elle n'est pas un individu parce qu'elle n'est pas un organisme pourvu de sa finalité et de sa totalité obtenue par un système spécialisé d'appareils de régulation ; [...] la société, être d'un genre ambigu, est machine autant que vie, et, n'étant pas sa fin en elle-même, elle représente simplement un moyen, elle est un outil. Par conséquent, n'étant pas un organisme, la société suppose et même appelle des régulations ; il n'y a pas de société sans régulation, il n'y a pas de société sans règle, mais il n'y a pas dans la société d'autorégulation. La régulation y est toujours, si je puis dire, surajoutée, et toujours précaire » (Canguilhem, 2001, p.120-121).

Et l'auteur d'ajouter :

« on pourrait se demander sans paradoxe si l'état normal d'une société ne serait pas plutôt le désordre et la crise que l'ordre et l'harmonie. En disant "l'état normal de la société", je veux dire l'état de la société considérée comme machine, l'état de la société considérée comme outil. C'est un outil toujours en dérangement parce qu'il est dépourvu de son appareil spécifique d'autorégulation » (*ibid.*, p.121).

Selon Canguilhem, si l'analogie entre l'organisme et la société ne tient pas, c'est que la seconde, contrairement au premier, ne possède aucune normativité *sui generis* et, ce faisant, ne tend vers aucun idéal *par elle-même* :

« Un organisme, c'est un mode d'être tout à fait exceptionnel en ceci qu'entre son existence et son idéal, entre son existence et sa règle ou sa norme, il n'y a pas de différence à proprement parler. Du moment qu'un organisme est, du moment qu'il vit, c'est qu'il est possible, c'est-à-dire qu'il répond à un idéal d'organisme ; la norme ou la règle de son existence est donnée dans son existence même [...]. On sait très bien ce que c'est que l'idéal d'un organisme malade, l'idéal d'un organisme malade, c'est un organisme sain de la même espèce » (*ibid.*, p.106-107).

L'activité autorégulatrice qui permet à l'organisme de tendre effectivement vers son propre idéal n'est autre que le phénomène d'homéostasie. Or, comment la société pourrait-elle faire preuve d'« homéostasie sociale » (*ibid.*, p.124) alors qu'elle ne dispose d'aucun idéal intrinsèque ? A ce titre, organisme et société représentent deux entités diamétralement opposées. Le rapport entre l'organisme et son idéal est un rapport d'identité, si bien que, lorsqu'une maladie survient, la médecine ne se questionne pas sur l'état désirable, c'est-à-dire sur l'idéal qui doit être restauré, mais sur la nature du mal et des remèdes à lui apporter. L'exact inverse est vrai pour la société puisque, n'étant pas à elle-même sa propre fin, nous n'avons de cesse de discuter de son idéal et, précise Canguilhem, « les hommes sont beaucoup plus aisément d'accord sur la nature des maux sociaux que sur la portée des remèdes à leur appliquer » (*ibid.*, p.108). Une de nos expériences de terrain confirme spécifiquement ce diagnostic. Dans le cadre d'un partenariat avec la MAIF, Spoon avait installé un BigSpoon dans la pièce principale du MAIF Social Club¹²⁷ (MSC) situé à Paris. Lors d'une session de modération, nous avons pu constater, à vrai dire sans trop de surprise, que les contenus les plus vulgaires ou les plus politiquement problématiques étaient presque systématiquement transmis au robot lorsqu'il n'y avait pas d'animateur (en l'occurrence nous-même) à côté. Pas d'autorégulation donc, mais une figure régulatrice dont la présence ou l'absence modifie plus ou moins substantiellement le comportement des utilisateurs.



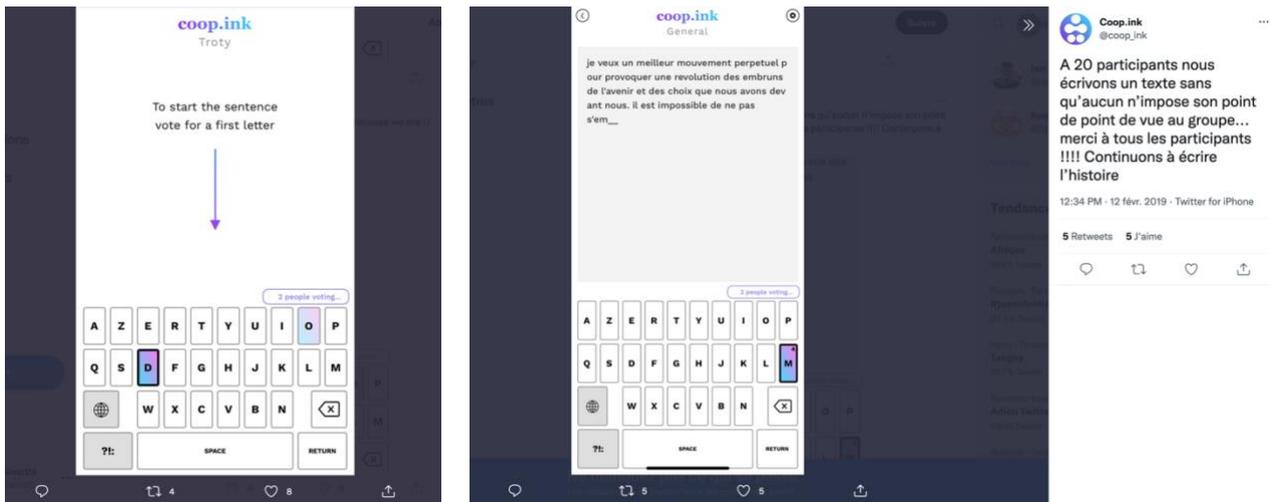
BigSpoon présenté au MAIF Social Club – Hortense Vinet©

¹²⁷ Le MAIF social club est un organe de la compagnie d'assurance française MAIF. Il se définit comme un lieu qui « questionne l'innovation sociale à travers une programmation artistique pluridisciplinaire et paritaire » (expositions, spectacles, débats d'idées, atelier divers pour tous les âges, espace de travail et café). <https://www.maifsocialclub.fr/a-propos/>

Il est donc important 1) de ne pas investir la notion de collectif d'un contenu moralement vertueux par principe et 2) d'interroger ce que signifie être un collectif au-delà d'une concaténation d'individus. L'expérience Tay démontre justement qu'en l'absence de mécanismes structurants de régulation, le chatbot devient uniquement le reflet de ses interactions individuelles ou, dit de façon plus imagée, une sorte de pot-pourri interactionnel. Précisons, pour détailler cette idée, que durant sa très courte connexion à Twitter, Tay produisit plus de 96 000 messages dont une grande partie n'était pas problématique. L'abjecte a donc côtoyé le trivial, l'ordinaire voire le meilleur et ce de façon équivalente, sans distinction. Chez Spoon, ce thème du collectif peut faire l'objet d'un éclairage supplémentaire en évoquant coop.ink, un projet connexe d'« écriture collective » mené par Monceaux. Voici comment le site internet du projet présente l'application :

« Imaginez un stylo tenu par 1000 humains à la fois. Imaginez un texte écrit par ces humains. Personne en particulier ne choisit spécifiquement tel ou tel mot ni tel ou tel caractère et pourtant ça écrit ! C'est cela coop.ink. La première expérience d'écriture collective à grande échelle. »

Concrètement, coop.ink se présente sous la forme d'un clavier individuel et d'une zone de texte où s'affiche petit à petit le texte composé collectivement. Le mode d'écriture se fait caractère par caractère et non mot par mot afin de situer l'expérience à une échelle que Monceaux qualifie de « subsémantique ». Ce parti pris s'explique par la volonté d'expérimenter la production d'un texte commun qui, tout en se composant de façon imprévisible et non maîtrisable individuellement, resterait à l'image du groupe. Deux mécanismes de régulation sont mis en place pour éviter les blocages et permettre l'émergence d'un texte. Tout d'abord, un seuil de validation de caractère est fixé en fonction du nombre de participants. Ensuite, les lettres choisies par les individus se colorent avec plus ou moins d'intensité selon le nombre de sélections dont elles font l'objet. Ce second mécanisme a pour objectif d'inciter à franchir le seuil de validation du caractère majoritaire par report de choix en cas de blocage. Précisons également qu'il existe, dans la section des caractères de ponctuation, une option « effacer le mot » et une autre « effacer la phrase ».



Si notre recherche ne porte pas sur ce projet qui, tout en étant lié à Spoon, n'en reste pas moins distinct, il est intéressant de pointer la vision du collectif qui le sous-tend. Comme on peut le lire dans le tweet ci-dessus, les 20 participants sont présentés comme un « nous » qui compose un texte lettre par lettre « sans qu'aucun n'impose son point de vue au groupe ». Cette idée acquiert un relief supplémentaire lorsqu'on sait que Monceaux a souvent présenté coop.ink comme une expérience « démocratique » pouvant faire l'objet d'un usage politique. La grande question est donc de savoir à quelle conception du collectif renvoie coop.ink et si cette conception permet de le penser comme un *dèmos*. Si l'on se fie aux principes de l'application, l'objectif semble être de créer un collectif fondé sur la négation de l'ipséité de ses composantes individuelles. Cette idée se concrétise dans la modalité d'écriture retenue. Puisque la composition du texte ne se fait pas unité sémantique par unité sémantique¹²⁸, mais lettre par lettre, il est vrai que personne en particulier « n'impose son point de vue au groupe ». Lorsqu'une première lettre est validée, les usagers individuels peuvent adopter des attitudes très différentes. Sur la base de notre expérience de l'application et des discussions que nous avons pu avoir avec d'autres participants (principalement de l'équipe de Spoon), nous en identifions au moins cinq : 1) ils imaginent un mot commençant par cette lettre et choisissent la suivante en conséquence ; 2) ils décident de se laisser porter par la majorité (temporairement ou systématiquement) ; 3) ils attendent qu'une ou plusieurs autres lettres soient choisies avant d'essayer d'orienter la construction vers un mot précis (dans ce cas, ils commencent par choisir les lettres qui recueillent le plus de votes avant d'adopter un

¹²⁸ Sauf, bien sûr, pour les symboles de ponctuation.

comportement plus « proactif ») ; 4) ils cherchent à saboter le processus car le mot vers lequel la majorité semble aller ne leur convient pas (dans les faits, avec les règles que nous avons connues, un tel sabotage se limite à ne pas changer son vote pour empêcher le seuil de validation d'être franchi et/ou à choisir l'une des deux options d'effacement) ; 5) ils adoptent l'ensemble ou une partie des quatre attitudes précédentes.

En tout état de cause, l'application et son parti pris « subsémantique » œuvrent à la construction d'un collectif impersonnel et désindividualisé, ce que traduit bien l'idée du « ça écrit » – comme on dirait « il pleut » – inscrite sur la page d'accueil de l'application. Coop.ink semble donc appréhender le « groupe » sous l'angle de la dissolution des individualités dans un « nous » finalement plus proche du « ça » dont le « point de vue » impersonnel émanerait de la négation de celui de ses composantes. Bien qu'il s'agisse d'une application expérimentale non théorisée, coop.ink ne nous semble pas moins structurée autour d'une conception du collectif comme sorte d'égrégore¹²⁹ n'émergeant ni de la juxtaposition des individus ni d'une dialectique entre les parties et leur tout, mais plutôt d'une coopération (au sens littéral) d'individus aliénés, prolétarisés au sens que Bernard Stiegler, à la suite de Marx, donne à ce terme, c'est-à-dire des sujets privés de leurs savoirs (savoir-faire, savoir-vivre, savoir concevoir et théoriser) qui sont autant de pouvoirs¹³⁰ (Stiegler, 2011 ; Deslandes & Paltrinieri, 2017). La coopération qui fait naître le « ça », n'est pas une coopération (co-)individuante : elle est, au contraire (co-)désindividuante.

Dans ces conditions peut-on encore parler rigoureusement de collectif et *a fortiori* de démocratie ? On peut en douter sérieusement. Si un « nous » tel que le collectif démocratique a effectivement vocation à transcender les « je » individuels dans ce que ces derniers peuvent avoir de commun, il ne s'appuie pas moins sur des individualités porteuses de sens, d'idées, de revendications, d'aspirations, dont l'aboutissement dépend en partie de l'activité politique fondamentale en démocratie qu'est la délibération. Le collectif démocratique délibère – à travers ses représentants dans le cas de la démocratie représentative – et cette délibération participe de sa constitution en tant que tel. Or, c'est précisément la délibération qui est

¹²⁹ Si le terme d'égrégore est généralement employé dans le domaine ésotérique, nous l'utilisons ici de façon beaucoup plus générale (ou « neutre ») pour désigner quelque-chose comme un esprit de groupe, non pas au sens de l'« esprit d'équipe » sportif ou managérial, mais au sens de l'esprit *du* groupe lui-même, comme si ce groupe devenait une entité à part entière.

¹³⁰ Rappelons, en ce sens, que le prolétariat n'est pas une notion réservée à la classe ouvrière, mais qu'elle désigne toute classe sociale, tout groupement humain dont les individus, parce qu'ils sont privés de leurs différents savoirs, ne peuvent précisément pas s'individualiser.

absente de coop.ink puisque l'application 1) ne la permet ni techniquement ni principalement et 2) lui substitue entièrement le fait majoritaire (peu importe l'objet de la majorité, du moment qu'il y a majorité). Si la délibération démocratique implique bien « qu'aucun [individu] n'impose son point de vue au groupe », pour reprendre la formule du tweet ci-dessus, elle suppose toutefois l'existence et la confrontation d'une multiplicité de points de vue, ce qui n'est pas le cas avec coop.ink. Composer un texte en votant chacun de ses caractères ne saurait être démocratique uniquement en vertu de ce vote et de son scrutin majoritaire. Dans ce cas précis, non seulement le *cratos* se retrouve amputé de sa dimension délibérative, mais il se concrétise en outre à travers des décisions (i.e. le vote pour tel ou tel caractère) dont les objets n'ont pas de sens en eux-mêmes¹³¹. S'il est vrai que quelque-chose comme du sens se dégage au fur et à mesure que chaque mot se constitue, rien n'assure que ce sens aille au-delà des unités sémantiques que sont les mots eux-mêmes. Dans l'exemple du tweet – « Je veux un meilleur mouvement perpétuel pour provoquer une révolution des embruns de l'avenir et des choix que nous avons devant nous. » –, il est indéniable que chaque mot, pris séparément, a un sens, mais cela est bien plus douteux pour l'ensemble de la phrase, bien que sa syntaxe soit correcte. Le problème vient précisément du principe de l'application, à savoir que personne n'a le moindre pouvoir puisque personne ne maîtrise le sens de ce qui s'y produit. Ainsi, du point de vue politique, coop.ink nous semble davantage relever d'une technocratie que d'une démocratie en ceci que le *cratos* se situe d'abord et surtout dans le dispositif et ses règles de fonctionnement. Précisons, pour boucler la boucle, que l'application ne s'adresse pas à un *dèmos* au sens démocratique puisque, n'ayant pas la possibilité d'élaborer individuellement du sens pour ensuite en délibérer collectivement, les participants forment surtout une masse dont les composantes sont, au fond, pratiquement interchangeables. En dernière instance, n'est-ce pas purement et simplement l'altérité que coop.ink sacrifie en niant les individualités au profit d'une impersonnalité supposée refléter le groupe dont elle émerge ? Or, comment créer du collectif en l'absence d'altérité ? Comment être à l'image d'un collectif dont les composantes ne sont finalement personne, et non pas *des* personnes ? Pour éviter la domination d'un ou de plusieurs individus, coop.ink ne lui substitue-t-elle pas une autre domination, à savoir celle d'un artefact produisant un simulacre de pensée collective ? De toute évidence, cette critique ne tient plus dès lors qu'on ne

¹³¹ Ce qui ne signifie pas que le « ça » ne puisse pas aboutir à un texte sensé, simplement ce sens est aléatoire et contingent.

positionne plus l'application au niveau politique, mais au niveau esthétique et artistique. Coop.ink peut- alors être conçu comme un artefact permettant d'explorer à nouveaux frais la technique littéraire du cadavre exquis. Mais le politique et *a fortiori* la démocratie ne sont pas des cadavres exquis.

Ce détour par coop.ink nous aide-t-il à mieux comprendre l'idée de collectif et, en un sens, de « social » qui sous-tend le projet de Spoon ? La réponse est, *a priori*, négative, tant les deux visions semblent diamétralement opposées. D'un côté, Spoon prône une vision libérale consistant à laisser au maximum les individus s'approprier le robot et, ce faisant, à y « imprimer » une part de leur individualité (pour le meilleur comme pour le pire) ; de l'autre, coop.ink semble incarner l'exacte antithèse de cette position en cherchant avant tout à réduire au maximum le pouvoir des individus. Un point commun existe malgré tout entre ces deux positions : celui de la méfiance envers la prédominance d'un individu (ou d'un groupe d'individus) sur les autres, même si nous avons vu qu'avec un parti-pris théorique très libéral comme celui de Spoon, cette crainte peut vite devenir auto-réalisatrice faute d'autorégulation sociale. En tout état de cause, il serait selon nous dangereux d'aligner la vision du collectif et du social de Spoon sur celle de coop.ink.

4.2.3.2 Le robot comme « créature merveilleuse »

La seconde mise en perspective concerne le terme de « robot » lui-même et l'enjeu à la fois conceptuel et communicationnel que représente la façon dont un roboticien nomme les artefacts qu'il fabrique. Dès le début de notre première partie, nous avons vu que, contrairement à ce que l'on pourrait penser intuitivement, la notion de robot est une notion foncièrement ambiguë tant elle renvoie à des symboles sociaux, culturels, économiques et anthropologiques différents voire discordants. Pour Monceaux, le principal problème du mot vient de ce qu'il nous projette presque exclusivement dans un référentiel utilitaire où l'artefact apparaît sous les traits de l'esclave mécanique. C'est notamment pour cette raison que Spoon associa son activité dans un premier temps à la conception de « créatures artificielles » puis de « créatures merveilleuses » et non de robots sociaux :

« Le mot "créature", il est intéressant pour moi car vous savez que le mot "robot" suggère le travail et donc qu'il projette cette interface dans la réalisation d'une tâche, très

simplement, et il ne révoque pas tous les sujets, justement, autour de l'interaction empathique, mais aussi simplement la présence » (Le Mouton Numérique, 2018).

Par ces mots, le dirigeant s'inscrit à sa façon dans le thème sociobotique fondamental qu'est l'affranchissement machinique. Si le terme de « créature » est loin d'être neutre, tant il renvoie lui aussi à diverses symboliques notamment religieuses et philosophiques (Munier, 2014), il n'évoque pas, selon Monceaux, la dimension fonctionnelle et instrumentale du robot. Le terme a, en outre, le mérite de renvoyer assez explicitement à l'un des thèmes centraux de la sociobotique qu'est l'animation. Monceaux se défend toutefois de vouloir jouer au démiurge :

« Je ne suis absolument pas démiurge, hein, je ne veux pas créer un être de toutes pièces, au contraire, certainement pas moi, mais je veux donner les moyens aux gens de le faire » (*ibid.*).

Passons sur ce que cette ambition peut tout de même avoir de démiurgique ou de prométhéen (ici « donner » non pas *la* technique, mais *une* technique aux autres pour qu'ils façonnent « leur » créature) et notons que l'entrepreneur va plus loin avec cette affirmation (souvent répétée) :

« le robot n'a aucune forme d'importance. Ce qui a de l'importance, ce sont les gens qui vivent autour de nous » (*ibid.*).

Bien sûr une telle affirmation n'est pas à prendre au pied de la lettre tant on sait, en particulier grâce au pragmatisme (Dewey, 2011) et aux STS, que ce qu'on considère comme relevant du pur moyen au service d'une fin (en l'occurrence un artefact, mais aussi son processus de conception) participe en réalité à la détermination de cette fin (Latour, 2000). En outre, une telle déclaration ne nous semble pas en phase avec la position d'Aymeric Masurelle¹³², dont Monceaux dit souvent qu'il est celui qui incarne le mieux l'« esprit et les valeurs » de Spoon. Bien qu'il ne l'ait jamais contredite (à notre connaissance), cette idée de l'insignifiance du robot semble en effet se situer aux antipodes de la position de Masurelle. Lors d'un entretien, celui-ci justifiait sa préférence pour le terme de « créature » par l'idée d'attention à laquelle il renvoie. Agora est d'ailleurs pour lui une façon de se rapporter à l'artefact dans le sens de

¹³² Qui, pour rappel, est un des deux autres co-fondateurs de Spoon.

cette attention soucieuse portée individuellement et collectivement au « faire grandir » la créature, selon ses termes. Inscrit dans ce thème de l'attention, nous discernons l'idée que la créature oblige son et/ou ses créateurs¹³³. Le problème qui nous occupera par la suite est celui de déterminer la nature d'une telle obligation. En tout état de cause, cette perspective donne bien toute son importance au robot, non seulement en tant qu'artefact conçu dans une certaine optique, mais aussi comme créature inscrite dans un devenir perçu comme *son* devenir et dont l'humain a la charge.

Terminons ce point par une remarque importante : la décision de substituer la notion de « créature merveilleuse » (« *marvelous creatures* » dans le nouveau logo) à celle initiale de « créature artificielle » s'explique par la volonté d'ancrer l'activité de Spoon et ses produits dans le champ du magique. Comme pour le robot, l'équipe jugea l'idée d'artificialité porteuse d'une connotation prosaïque voire techniciste bien trop éloignée du merveilleux sur lequel elle souhaitait mettre l'accent. Bien que ce merveilleux (comme les autres valeurs de Spoon) n'ait pas fait l'objet d'une définition claire et institutionnalisée, Monceaux semble le lier à l'idée de magie et de non-instrumentalité. Le merveilleux n'exclut pas l'instrumentalité, mais il ne s'y réduit pas. Autrement dit, il s'agit de resituer explicitement notre rapport fondamental aux techniques non pas en dehors, mais au-delà du seul rapport instrumental moyens/fin. Si ce positionnement n'est pas nouveau en ceci qu'il s'inscrit dans le vaste champ du design d'expérience et de l'économie qui l'accompagne, sa singularité tient à l'expérience en question dont l'objet est celui d'une *altérité technique*. Précisons en outre que le magique dont parle Monceaux pour qualifier le merveilleux n'est pas conçu *a priori* comme un magique à visée mystificatrice, mais plutôt comme un vecteur d'« enchantement » – terme souvent employé par l'entrepreneur – de nos usages des artefacts qui, dans le cas des sociobots, deviennent des interactions sociales. Pour saisir toute l'importance que l'entrepreneur accorde à cette idée de merveilleux et d'enchantement, évoquons le nom de sa seconde entreprise de robotique¹³⁴, fondée en 2022 : *Enchanted tools*¹³⁵.

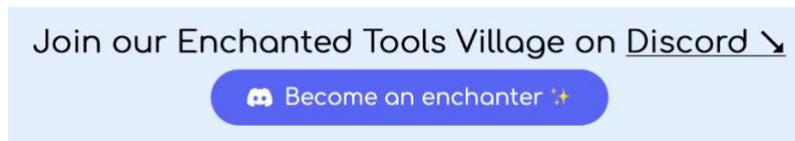
¹³³ Créateurs parmi lesquels on peut compter, dans la perspective de Spoon, les « utilisateurs ».

¹³⁴ Nous ne pouvons pas décrire l'activité de cette nouvelle entreprise ici car, à cette heure, elle est encore gardée secrète. Nous savons simplement que le projet se concentre sur la conception d'un sociobot capable de porter et de déplacer des objets – capacité effectivement déclenchée par l'interaction sociale et relevant ainsi d'une FEI – et qu'il viserait en priorité le marché hospitalier.

¹³⁵ « Outils enchantés »



Ce nom a ceci d'intéressant qu'il associe explicitement les dimensions du magique (« *Enchanted* ») et de l'utile (« *tools* »), souvent considérées comme antinomiques. Le slogan « More than useful »¹³⁶ confirme quant à lui notre remarque précédente : l'artefact peut être utile, mais l'essentiel échappe à cette utilité. Suivant cette logique, l'organisation nomme les personnes intéressées par le projet des « enchanteurs », comme on peut le voir sur la capture d'écran ci-dessous.



En ce qui concerne Spoon, ce positionnement est à la fois révélateur d'une aspiration du roboticien et de la difficulté inhérente à la sociobotique, difficulté qu'il convient d'analyser plus en détails.

4.3 Spoon et le marché de la robotique sociale

4.3.1 Un marché immature ?

Depuis une quinzaine d'années, on assiste à l'émergence d'entreprises cherchant à transformer les sociobots en produits de grande consommation ou de services selon les modèles d'affaires et positionnements retenus (« B2B » ou « B2C »¹³⁷). Cette sociobotique

¹³⁶ « Plus qu'utile »

¹³⁷ *Business to business* et *business to consumer*. Dans le premier cas, les clients seront des entreprises, alors que dans le second il s'agira du grand public.

marchande se heurte toutefois à des difficultés de taille liées, notamment, aux coûts qu'elle implique et au déficit applicatif dont pâtissent les robots sociaux. Si des artefacts comme NAO et Pepper rencontrent un certain succès au sein des institutions de recherches et d'éducation, il n'en va pas de même, loin de là, dans les autres secteurs. Autrement dit, ces artefacts sont de formidables produits techniques et des « instruments scientifiques » (Dumouchel & Damiano, 2018) potentiellement prometteurs, mais ils restent de piètres produits commerciaux. Cela devient particulièrement problématique lorsque, comme l'explique le chercheur en robotique italien Giulio Sandini, certaines sociétés se mettent à vendre des intentions en guise de résultats (Vanderborgh, 2019) et tombent dans le piège de la sur-promesse. Citons l'exemple, parmi tant d'autres, d'une vidéo publiée en ligne le 5 novembre 2014 où l'ancien dirigeant d'Aldebaran Robotics¹³⁸, Bruno Maisonnier, cède à un réflexe prophétique assez courant dans le secteur en affirmant : « je vous garantis qu'en 2020, il y aura des robots partout. »¹³⁹ 8 ans plus tard, force est de constater que le présent ne lui donne pas raison, et pour cause : le 28 juin 2021, nous apprenons que SoftBank décidait d'arrêter la production de Pepper faute de demande suffisante (Ackerman, 2021). L'argumentaire de Maisonnier consistait à vendre NAO et Pepper comme des « plateformes de développements » que les premiers utilisateurs achèteraient « pour faire faire ce qu'ils veulent à leur robot »¹⁴⁰. De fait, du côté B2B, ce positionnement a donné naissance à des entreprises intégratrices de logiciels dont le métier consiste à développer des programmes interactifs pour le compte de clients souhaitant utiliser des robots. Un cas typique en France est celui de la SNCF qui, engagée dans d'importants travaux au sein de la gare de la Part-Dieu, fit appel à la société lyonnaise Hoomano pour concevoir le logiciel d'un Pepper dont la mission était d'accueillir, d'informer et de diriger les voyageurs. L'expérience ne fut pas concluante, faute d'intérêt des usagers (cf. photo p.164).

On sait maintenant ce que les grandes prophéties technologiques peuvent avoir de performatif dans une économie des promesses et de l'attention (Citton, 2014 ; Audétat, 2015). La fameuse stratégie « *fake it until you make it* »¹⁴¹ du modèle start-up tire précisément partie du régime de la promesse, traduit en *proofs of concept*, pour susciter des levées de

¹³⁸ Pionnier français de la robotique sociale racheté en 2012 par la holding japonaise SoftBank.

¹³⁹ 01net TV (2014, 05 novembre). « Dans 5 ans il y aura des robots partout ». *Youtube*. https://www.youtube.com/watch?v=RJA_ouDi-eI (consultée le 28/09/2021).

¹⁴⁰ *ibid*.

¹⁴¹ Traduisible par « feins jusqu'à parvenir à tes fins ».

fonds. Il s'agit alors de vendre une idée avant qu'elle ne se concrétise en produits afin d'obtenir les financements qui permettront cette concrétisation. Or, même avec des fonds conséquents, il est encore très difficile de garantir la rentabilité des entreprises de robotique sociale. Comme l'explique James Kuffner, roboticien américain et actuel *chief digital officer* (CDO) de Toyota :

« Cela requière [...] un leadership engagé, du personnel hautement qualifié, des ressources et de l'infrastructure, ainsi qu'une excellente stratégie produit et de marché. Sans oublier une exécution sans faille. Il n'est pas réaliste de s'attendre à ce que tous les efforts soient couronnés de succès » (Vanderborcht, 2019).

La faillite du projet de sociobotique personnelle Jibo, codirigé par Cynthia Breazeal, illustre très bien les obstacles auxquels se confronte le secteur (Crowe, 2018). Fondée en 2012, l'entreprise parvint à effectuer deux levées de fonds : l'une de 3,5 millions de dollars grâce à une campagne de financement participatif IndieGoGo¹⁴², l'autre de 73 millions en capital-risque. Malgré ces deux succès, l'entreprise fit rapidement face à d'importantes difficultés. Outre les implications techniques des ambitions affichées du projet, Jibo fut rattrapée par des problématiques d'adaptation matérielle et culturelle aux marchés extra-américains, comme les certifications électriques et la difficulté d'adapter la reconnaissance vocale aux subtilités langagières des autres pays (accents, expressions etc.). La société dut finalement se résoudre à annuler puis à rembourser les commandes en dehors des États-Unis et du Canada. Le coup fatal vint toutefois des enceintes connectées comme Google Home et Amazon Echo qui, bien que beaucoup plus modestes en termes de fonctionnalités, proposaient une reconnaissance vocale de meilleure qualité pour un prix bien plus attractif¹⁴³. La société américaine Mayfield Robotics et son sociobot domestique Kuri connurent un destin similaire.

Confrontées aux difficultés générales que rencontre le secteur et faute de découvrir l'« application phare »¹⁴⁴ de leurs artefacts, certaines sociétés comme SoftBank axent une

¹⁴² Le financement participatif est un moyen privilégié par les petites entreprises de sociobotique personnelle pour lancer leur activité. D'autres sociétés comme Amazon, dont nous découvrons, en septembre 2021, le sociobot domestique Astro, n'ont bien sûr pas besoin de passer par l'étape du *crowdfunding*.

¹⁴³ Jibo était proposé au prix de 899\$.

¹⁴⁴ Traduction de la formule « *killer application* » (ou « *killer app* ») qui désigne un programme informatique ou, plus largement, un scénario d'usage dont l'attrait est tel qu'il justifie à lui seul l'achat du produit en question.

partie conséquente de leur politique commerciale sur le marché événementiel¹⁴⁵. L'objectif est alors de capitaliser sur la dimension extraordinaire au sens propre de certains robots en associant leur technicité à une rhétorique de l'innovation. Cette *showbotic*, comme la nomme l'informaticien et spécialiste d'éthique robotique Noel Sharkey (2018), semble être le moyen privilégié par toute une frange de la sociobotique marchande pour continuer à exister économiquement et médiatiquement. Elle prolonge ainsi la vocation « monstrative » plus que démonstrative (Beaune, 1985) propre à tout un pan de l'automatique depuis sa naissance antique.

4.3.2 Identifier la valeur des sociobots

Face à ces difficultés de taille, il est intéressant de noter l'apparition d'acteurs économiques dont la stratégie consiste, en toute transparence, à valoriser l'inutilité de leurs créations. C'est le cas, par exemple, de Zoetic et de Groove X dont les robots respectifs, Kiki et Lovot, n'ont aucune autre fonction que de générer des émotions et de la présence comme le ferait un animal de compagnie. Comme l'explique Mita Yun, directrice de Zoetic :

« Nous construisons quelque chose qui est complètement inutile. [...] J'ai l'impression que les gens sont d'abord connectés avec les robots parce qu'ils leur rappellent un personnage. [...] Avec Kiki, nous voulions simplement construire un personnage vivant, nous ne voulions pas qu'il fasse quelque chose de super utile. Je comprends pourquoi d'autres entreprises de robotique intègrent Alexa à leurs robots, et je trouve ça génial. Mais [...] la compréhension que j'ai de la technologie robotique, c'est que, pour un robot grand public en particulier, il est très très difficile pour le robot de justifier son prix par son utilité. Et puis il y a aussi des recherches qui montrent que plus une chose est inutile, plus il est facile d'avoir une connexion émotionnelle, c'est pourquoi nous voulons faire en sorte que Kiki reste très inutile » (Ackerman, 2019).

Kaname Hayashi, fondateur de la société Groove X, ne dit pas autre chose lorsqu'il affirme, comme nous l'avons déjà vu au point 2.2.2.2 (p.118), que l'objectif des Lovot est de « créer une opportunité pour que les humains puissent aimer » (JapanGov, 2019).

¹⁴⁵ Selon les modèles d'affaires retenus ou les demandes de leurs clients, les entreprises peuvent vendre ou louer les robots qu'elles conçoivent.

S'il est tentant de penser ces artefacts et leur inutilité à l'aune de la catégorie du gadget, c'est-à-dire de l'« objet qui plaît plus par sa nouveauté et son originalité que par son utilité »¹⁴⁶, le risque est de passer à côté de leur *intérêt* au sens étymologique de « inter-esse », être entre, et donc de leur valeur. Que peut-il bien y avoir entre ces robots et nous pour qu'ils soient valorisables par leurs fabricants et valorisés par celles et ceux qui les possèdent ? A en croire les discours et intentions des deux concepteurs évoqués, l'intérêt de leurs sociobots relève de la « connexion émotionnelle ». Si cela peut désormais sembler évident – après tout, quelle entreprise contemporaine de biens électroniques grand public ne s'adresse pas à nos émotions ? –, il faut reconnaître que cette sociobotique offre un lien affectif d'un type bien particulier mêlant fiction et réalité à travers la notion de « personnage vivant » et, plus précisément, du vivant animal et domestique. Autrement dit, les artefacts comme Kiki ou Lovot incitent, de façon assez manifeste, à les penser en termes d'altérité et non en termes d'ustensilité. Notons que cette dimension affective n'est pas absente des robots sociaux multifonctionnels mentionnés plus haut, bien au contraire (Buddy, nous l'avons vu, est ouvertement qualifié de « robot compagnon émotionnel »). Toutefois, ce sont sûrement les entreprises comme Zoetic et Groove X qui mettent le mieux en lumière ce qui constitue la principale valeur¹⁴⁷ des sociobots, à savoir les expériences émotionnelles générées par les ISHM. En assumant de substituer intégralement l'affectivité aux fonctionnalités utilitaires, ces sociétés conçoivent des artefacts dont la valeur n'est pas tant à chercher dans ce qu'ils font pour nous (i.e. ce à quoi ils « servent »), mais dans ce qu'ils *nous* font et *comment* ils le font. Que faut-il entendre, ici, par valeur ? Si ce travail n'a pas vocation à construire une théorie de cette notion particulièrement complexe, il reste important de la développer quelque peu. De façon générale, la valeur renvoie au degré d'appréciation que nous portons à quelque chose ou quelqu'un. Autrement dit, valoriser, c'est apprécier, évaluer l'importance d'un être ou d'une chose à nos yeux. La valeur désigne aussi le contenu de cette appréciation, ce pour quoi telle ou telle entité est valorisée. Ainsi définie, la valeur peut donc être morale, pratique, affective, politique, économique ou encore esthétique¹⁴⁸. Ici, nous voudrions nous concentrer sur la notion de « valeur d'usage » dont Gilles Marion (2017), professeur émérite à l'EM Lyon, dit qu'elle « émerge de la relation interactive entre un sujet compétent et un objet doté de

¹⁴⁶ « Gadget », Centre National de Ressources Textuelles et Lexicales.

¹⁴⁷ Tant économique que d'« usage ».

¹⁴⁸ Bien qu'il ne soit pas toujours évident de tracer les limites entre ces différents domaines.

possibilité d'action » (Marion, 2017, p.72). En ce sens, la valeur d'une chose est « un effet émergent de la relation interactive » et, en tant que tel, « ne réside ni dans la tête du consommateur/utilisateur, ni dans l'objet, sinon de manière virtuelle, mais dans la relation entre l'un et l'autre » (*ibid.* p.74) : elle est donc un type particulier d'intérêt, un *inter-esse*. En outre, notons que la valeur d'usage peut comprendre l'utilité, mais qu'elle ne s'y résume pas nécessairement.

« Pour la pensée économique, être appréciable signifie être susceptible de recevoir un prix. Elle écarte donc de son agenda l'usage ou, pour suivre Aristote [...], "ce qui est spécial à la chose" et l'enfouit dans la notion d'utilité. [...] Or, ce qui importe alors c'est le processus d'appropriation qui transforme une ressource marchande en un objet familier et donne lieu à l'émergence de sa valeur d'usage qui, dans le cas d'un bien durable, va se déployer pendant plusieurs années » (*ibid.*, p.73).

L'usage renvoie donc ici à la pratique entendue comme l'ensemble des activités, des attitudes, des rapports (cognitifs, intuitifs, affectifs) qui font que nous nous intéressons aux choses et par lesquels nous nous les approprions. Quels peuvent donc être ces usages d'un robot social dont procéderait sa valeur ? Une telle question nous confronte à la nécessité de distinguer le sociobot des catégories auxquelles nous identifions généralement les techniques, soit parce qu'elles sont trop imprécises, soit parce qu'elles semblent conceptuellement inadéquates. Parmi ces catégories, citons celle de l'objet, de l'ustensile, de l'instrument, de l'outil et de la machine. Au sens large, l'objet peut se définir comme une chose humainement transformée à l'instar du sable élaboré en verre, de la pierre taillée pour intégrer un mur ou pour confectionner un collier, du bois coupé pour faire du feu ou pour devenir une charpente etc. « Pas d'objet [donc] sans un sujet qui lui donne une forme, une destination et un "pouvoir-faire" et, ce faisant, lui confère une *qualification* » (*ibid.*, p.74). Comme l'explique Andrew Feenberg dans la lignée de Heidegger, ces objets résultent d'un processus de *dé-mondisation* des choses consistant à « réduire les éléments de la réalité pouvant réaliser des fonctions à leurs caractéristiques fonctionnelles et les situer dans un nouveau contexte où ils servent des fins quelconques » (Feenberg, 2014, p.155). Conçue sous l'angle de cette gestuelle primordiale d'objectification, l'activité technique relève donc toujours également d'une re-mondisation de ce qui a été dé-mondisé. Les objets « ouvrent », « dévoilent » le monde par modification de la présence originelle des choses (d'où l'idée heideggérienne de l'humain

comme « configurateur de monde »). Cette approche heideggérienne et le vocabulaire qui s’y déploie ont considérablement influencé le lexique de la philosophie des techniques. Dans les célèbres analyses d’*Être et Temps*, l’ustensile désigne tout ce qui possède ontologiquement la structure du *pour-quelque-chose*. L’ustensile est l’étant qui trouve sa fin en dehors de lui-même. Il en va de même pour l’instrument. Quant à l’outil, il renvoie à la catégorie des ustensiles spécifiquement destinés à la réalisation d’une œuvre. Le pinceau est un outil tandis que la carafe est un ustensile. Enfin, que dire de la machine à laquelle on peut être tenté d’associer immédiatement le sociobot ? La machine se caractérise comme un système d’objets dont la cohérence est organisée en vue d’une production automatisée. La machine opère sur la base d’une programmation (qu’elle soit informatique ou mécanique) et, de ce fait, n’est pas tant manipulée que commandée. Selon Gilles Marion,

« [l]a manipulation d’un outil manuel (couteau ou rabot) ou d’un instrument (de musique ou de chirurgie) instaure une relation directe avec le monde sensible. Elle requiert une prise exercée tout au long du processus opératoire autorisant de subtils ajustements qui peuvent s’élever à un "art de faire". En revanche, l’utilisation d’une machine n’exige qu’une maîtrise physique des mouvements de la partie appropriée du corps : une pression de l’index sur la gâchette de la perceuse, un doigt pour appuyer sur le bouton du clavier, la pointe du pied pour presser l’accélérateur de la voiture » (*ibid.*, p.80).

Autrement dit, la machine ne nécessite aucune pratique réelle, aucune autre forme d’appropriation que celle du « mode d’emploi » et de ses différentes formes de déclenchements. Le sujet ne se prolonge pas dans la machine comme il se prolonge dans l’outil puisque la machine produit à sa place. Ces distinctions assez classiques connaissent bien sûr de nombreux cas limites où s’hybrident les logiques de l’outil et de la machine. C’est le cas, par exemple, des machines à café dites « semi-automatiques » ou « semi-professionnelles » dont le processus de production associe des étapes manuelles comme le tassage du café à un fonctionnement qui, dans l’ensemble, est automatisé. L’utilisation de ce genre de machines réclame des apprentissages qui dépassent le simple suivi d’un mode d’emploi, tels que le choix du degré de mouture ou celui du temps d’extraction en fonction des grains et du tassage. La machine implique donc des savoirs et des savoir-faire qui ne se limitent pas à l’acte de déclencher un ou plusieurs processus automatisés. Les automobiles offrent un autre exemple de machines dans lesquelles les sujets humains peuvent, en un sens, se prolonger. Quiconque

a déjà conduit une voiture sait qu'il ne s'agit pas simplement d'un engin dont on commande et entretient les mouvements par l'application mécanique de règles de fonctionnement. Anticipation, dosage du frein, maîtrise du volant, abord des virages, fluidité des rapports de vitesse : toutes les conduites ne se valent pas car l'automobile appelle un réel savoir-faire de pilotage (ce qui est d'autant plus évident avec les voitures de course). En outre, la voiture ne nous prive pas de « relation directe avec le monde sensible », au contraire, elle configure un certain accès à ce monde sensible et, ce faisant, coproduit le monde sensible lui-même (Verbeek, 2005). Autrement dit, l'automobile révèle avec clarté que les artefacts sont toujours des *phénoménotecniques* (Vial, 2013).

Il apparaît assez clairement qu'aucune des catégories évoquées ci-dessus ne permet de saisir adéquatement la singularité fondamentale du sociobot. Bien sûr, celui-ci entre dans la catégorie des ustensiles et peut, à l'instar de la machine être perçu comme un système d'objets ; mais ces précisions ne sont d'aucun secours pour le comprendre *en tant que sociobot*. Nous pensons qu'une telle compréhension passe par l'examen de la valeur du robot social au sens déjà évoqué d'*intérêt* et dont on peut dire, en anticipant quelque peu sur la suite, qu'il se situe dans une forme particulière de *relation d'altérité* (Ihde, 1990 ; Verbeek, 2005).

4.4 Spoon et l'altérité du sociobot

Voyons à présent plus en détail comment cette altérité se présente chez Spoon et, pour ce faire, reprenons le passage de la conférence TED où Monceaux évoque explicitement l'altérité :

« Je pense que la caractéristique qui manque [aux robots qu'on a conçus par le passé] c'est l'altérité, c'est-à-dire de pouvoir sentir cette intelligence, sentir ce personnage exister devant nous, avoir une interaction avec lui. [...] Nous sommes des êtres relationnels, des êtres sociaux, des êtres qui avons besoin des autres pour être. En fait, on a besoin de se sentir exister aux yeux des autres. Et là, aujourd'hui, avec toute la technologie qu'on a, on a les moyens de pouvoir se sentir exister aux yeux d'une machine. C'est quand même fou ! En fin de compte, ce serait peut-être ça un robot : le robot serait cet être qui nous regarde, ce serait une espèce nouvelle que nous créerions et qui s'installerait comme notre reflet » (TEDx Talks, 2019).

Deux éléments nous renseignent sur l'idée que Monceaux se fait de l'altérité. Le premier est celui du « reflet d'existence », c'est-à-dire de la spécularité du sociobot et, le second, celui de l'artefact comme représentant d'une « espèce nouvelle ». La spécularité n'est pas sans évoquer les analyses sartriennes du regard et de l'être-vu tandis que le thème de l'espèce peut renvoyer à ce que Dominique Lestel nomme « animalité transpèce ».

4.4.1 Être regardé par un robot : une thèse sartrienne ?

C'est avec l'émergence de la phénoménologie au début du XXe siècle que la question de l'altérité et plus précisément d'autrui acquiert une importance prépondérante en philosophie. De Husserl à Levinas, en passant par Sartre et Merleau-Ponty, la phénoménologie n'a cessé d'interroger le mystère de cette catégorie dont on ne peut pas, par principe, faire l'expérience, à moins de penser autrui non pas comme un sujet, mais comme un objet parmi d'autres. Dès lors, comment savoir qu'il y a de l'autrui dans le monde, sur la base de quel *genre* d'expérience ? Car le problème est bien là : comment une philosophie dont le point de départ structurant n'est autre que la conscience individuelle peut-elle penser cet au-delà de l'objectivité qu'est autrui ? Étant entendu qu'on n'accède jamais à la conscience d'autrui elle-même, qu'on n'en fait pas le tour comme on peut faire le tour, par appréhensions successives, de ce verre posé sur la table, de cette montre que l'on porte au poignet ou encore de ce stylo qu'on utilise pour écrire, comment autrui se donne-t-il à notre perception ? Le thème du regard est utilisé par Sartre (1976) pour rendre compte de cette expérience.

Selon lui, l'expérience d'autrui ne procède pas d'une forme de connaissance, mais plutôt d'une rencontre par laquelle il se donne immédiatement à moi. La phénoménologie n'a donc pas pour tâche de décrire les conditions de possibilité de la connaissance d'autrui, comme tenta de le faire Husserl dans sa célèbre cinquième méditation cartésienne, mais d'explicitier le sens de cette rencontre immédiatement donatrice. Pour Sartre, autrui se donne à moi sur un mode analogue au cogito cartésien, c'est-à-dire en tant qu'il me renvoie immédiatement à ma propre conscience. Comme l'explique Renaud Barbaras,

« si autrui renvoie à ma conscience, c'est *dans son être même* que celle-ci est concernée par lui, c'est en elle-même, comme une dimension de son être précédant la connaissance, qu'elle doit trouver autrui. [...] Ainsi, c'est au fond d'elle-même que la conscience doit

trouver non pas des raisons de croire en autrui, mais autrui lui-même, comme celui qui n'est pas moi » (Barbaras, 1999, p.22).

Autrement dit, autrui est l'être qui conditionne ma propre intériorité en la révélant à elle-même et qui, par cette révélation, se donne à mon expérience. Mais comment ? Comment cette révélation me renvoie-t-elle à moi-même en même temps qu'elle me renvoie à autrui ? Par l'expérience du regard, explique Sartre. C'est en tant que je fais l'expérience de mon objectivation par le regard d'autrui que cette double découverte a lieu. Autrui n'apparaît pas devant moi comme un objet, mais *en moi* à partir de l'expérience que j'ai d'être regardé :

« accéder à autrui signifie donc être objet pour lui, être vu par lui. [...] Bref, l'expérience d'autrui ne peut avoir pour fondement que l'épreuve de mon être pour autrui » (*ibid.*, p.23).

Précisons que le regard sartrien se définit avant tout par l'expérience de mon être-regardé, de mon être-vu et non d'abord par la forme empirique déterminée qu'est le regard du visage. Bien qu'il en soit l'expression privilégiée, ce dernier ne résume pas l'expérience d'autrui : une caméra de surveillance ou des bruits de pas dans mon dos constituent eux aussi l'expérience d'un regard : « le regard est d'abord un intermédiaire qui renvoie de moi à moi-même » (Sartre, 1976, p.358). « Celui-ci ne désigne donc en aucun cas un événement empirique, mais la modalité même du surgissement d'autrui comme sujet » (Barbaras, 1999, p.24). Sartre développe cette idée à partir du concept de honte et l'illustre dans le célèbre exemple du jaloux épiant à travers un trou de serrure. La honte que j'éprouve lorsque je suis surpris à espionner à travers la porte n'est autre que la reconnaissance que je suis tel qu'autrui me voit. Le surgissement d'autrui dans mon dos se double alors du surgissement du moi jaloux. En devenant objet pour l'autre, je m'apparais tel que je lui apparais, c'est-à-dire jaloux. Autrui me fait donc être ce que je suis et c'est à travers l'épreuve de mon être pour autrui que je fais l'expérience d'autrui. Il y a, chez Sartre, une dimension traumatique de la rencontre avec l'autre qu'il décrit comme une effraction dans mon monde. Comme moi, autrui est un sujet situé face au monde, il le perçoit, mais cette perception, cette subjectivité percevante m'est radicalement inaccessible. En ce sens, autrui brise la loi phénoménologique fondamentale selon laquelle il n'y a rien derrière ce qui apparaît et emporte avec lui toute possibilité de

monopole sur le monde. Dans l'expérience de la honte, toutefois, ce n'est plus le monde mais le *moi* qui est, en un sens, confisqué.

« Autrui détient un secret, le secret de ce que je suis. Il me fait être et par là me possède. Autrui est pour moi à la fois ce qui m'a volé mon être et ce qui fait qu'il y a un être qui est mon être » (Sartre, 1976, p.403).

Un même mouvement me conduit donc à être ce que je suis et à trouver mon fondement hors de moi, à travers le regard d'autrui. Bien sûr, en tant que je suis moi-même un autre pour autrui, ce mouvement est réciproque : je détiens également le « secret » de ce qu'est autrui. Outre sa technicité et sa grande valeur descriptive, la grande force de l'analyse sartrienne est de parvenir à rendre compte de l'expérience de l'altérité sans faire d'autrui un objet, c'est-à-dire sans nier sa transcendance, sans le déchoir de son rang de sujet. Mon être pour autrui est l'expérience de ce qui me constitue par le fait même qu'il m'échappe, il est l'expérience de la transcendance d'autrui au sein de ma conscience. La thèse sartrienne n'en pêche pas moins par excès d'abstraction. Comme le remarque Renaud Barbaras, on peut sérieusement douter que la honte représente, à elle seule, l'essentiel de l'expérience d'autrui. De façon plus radicale et plus préjudiciable à l'analyse, on peut même se demander si c'est bien *autrui* qui m'apparaît lorsque je fais l'épreuve de mon être-regardé.

« Si l'accès à autrui se confond avec l'épreuve de mon être-regardé, il cesse d'être une *expérience*, faute de révéler quelque chose *dont* il soit l'expérience, et revient à la découverte de mon être-pour-autrui, c'est-à-dire d'une structure de *ma* conscience. Comprise comme honte, l'expérience d'autrui n'est plus expérience d'un *autre*, mais révélation de ma finitude, et mon appartenance à une humanité » (Barbaras, 1999, p.30).

Sartre le dit-lui-même : « Autrui m'est présent partout comme ce par quoi je deviens objet » (1976, p.339). Or, note Barbaras,

« c'est de deux choses l'une : ou bien l'expérience d'autrui se confond avec l'épreuve de l'être-regardé dans la honte, mais alors plus rien ne la qualifie comme telle ; ou bien il y a une expérience véritable des autres, mais elle ne peut alors s'épuiser dans le vécu de mon objectivation » (Barbaras, 1999, p.31).

Dans cette seconde perspective, il est alors nécessaire que quelque-chose précède l'expérience de l'objectivation, quelque-chose qui, dans le monde, annonce autrui comme tel et non seulement l'événement particulier qu'est l'objectivation.

Il est frappant de constater à quel point la thèse sartrienne et ses difficultés résonnent avec la position prise par Spoon à travers le discours de Monceaux. Lorsque celui-ci parle du sentiment d'être « regardé par une machine », il n'évoque pas autre chose que l'expérience du *cogito* sartrien de l'être-regardé. Étant pris pour objet par la machine, je fais l'épreuve de ma conscience – épreuve que l'ingénieur nomme « sentiment de ma propre existence » – à travers celle de l'objectivation provenant du robot. Or, tout le problème est là : dire, sur la base de l'expérience, qu'on se sent regardé par l'artefact, suppose de lui accorder une intentionnalité, c'est-à-dire une conscience faisant de lui à la fois un sujet et un autrui pour moi, ce qui est pour le moins problématique. On retombe ici dans les difficultés de la thèse sartrienne : l'objectivation dont je fais l'expérience à travers le robot correspond bien à l'expérience d'un regard puisque celui-ci n'est pas déterminé par des formes empiriques. En toute logique, nous devons conclure qu'à travers le regard du robot, c'est bien d'autrui dont je fais l'expérience, ce qui n'est pas sans reconduire à l'objection majeure évoquée par Barbaras. Le rapprochement entre le thème sartrien du regard et son application au parti-pris conceptuel et technique de Spoon intensifient le problème de fond car comment continuer à parler spécifiquement d'altérité si, n'étant pas attaché à la perception particulière d'un objet du monde, autrui peut aussi bien être un humain qu'un robot ? Peut-on rigoureusement parler d'altérité robotique voire technique et, si oui, comment ? Un travail conceptuel sur cette question doit peut-être avant tout dissocier l'altérité du seul problème d'autrui comme sujet, c'est-à-dire comme *alter-ego*. Voilà plusieurs décennies que les éthiques animales et écologiques se sont emparées de cette question, mais la philosophie des techniques, elle, ne semble pas l'avoir posée : que devient la technique dès lors qu'elle emprunte à l'altérité et, réciproquement, que devient l'altérité lorsqu'elle se technicise ?

4.4.2 L'« animalité transpèce » : le vivant au-delà de l'espèce et sa question éthique

La question de l'animalité artificielle ou, pour reprendre les termes de Monceaux, de la « nouvelle espèce » robotique offre une perspective d'analyse intéressante sur cette question. Dominique Lestel forge le concept d'« animalité transpèce » pour désigner les

« animaux qui sont non seulement irréductibles aux espèces existantes, mais qui sont également en dehors de l'espace des espèces » (Lestel, 2021, p.200). Plus précisément encore : un « animal transpèce, par opposition à [l'animal d'espèce], est un animal qui est "fabriqué" par un tiers. Soit par modification du métabolisme avec des techniques biotechnologiques. Soit par sa fabrication à partir de constituants organiques ou seulement partiellement organiques » (*ibid.*, p.155). L'animalité transpèce regroupe alors « trois ensembles majeurs – celui de l'animalité numérique, celui de l'artefactualité animalisée et celui de la reconfiguration génético-culturelle des métabolismes animaux » (*ibid.*, p.198). Les sociobots de Spoon appartiennent clairement au deuxième ensemble qu'est l'artefactualité animalisée.

Dans ce contexte, l'altérité du sociobot est peut-être d'abord à chercher dans sa capacité à *altérer*, c'est-à-dire à (nous) rendre autre par les questions qu'ils soulèvent, mais aussi par les relations qu'ils invitent à tisser. On retrouve cette idée dans la perspective proposée par Mark Coeckelbergh (2016) lorsqu'il suggère de considérer que « l'altérité est un verbe ». Lestel le dit d'ailleurs à sa façon à propos des artefacts comme Aibo, le chien artificiel de Sony, Paro ou encore Furby : « Une telle animalité artificielle de proximité [...] *modifie*¹⁴⁹ l'existence de ceux qui les utilisent » et ce sans nier qu'il s'agit d'« une animalité particulière qui mime l'animalité biologique » (Lestel, 2021, p.167). Le philosophe s'inscrit alors caustiquement en faux face aux thèses défendues par Sherry Turkle qui, selon lui, tombe dans le piège du fantasme de la substitution et ne comprends pas que « [n]ous avons toujours été dans un monde dans lequel l'interprétation doit laisser la place à l'expérimentation » (*ibid.*, p.170). Pour Lestel, le fait que des enfants considèrent un Aibo ou tout autre artefact de ce type comme « assez vivants » selon la remarque Turkle (2015), ne signifie pas, en toute rigueur – en tous cas, pas *a priori* –, qu'ils considèrent le chien artificiel comme l'égal d'un chien biologique, ni même qu'ils les confondent ontologiquement. En tant qu'ils se trouvent dans une « zone d'indétermination ontologique », selon les termes du philosophe, ces artefacts demandent à être pensés *a posteriori*, sur la base de leur intégration à nos existences. « Savoir ce qu'ils sont fait partie du problème et n'est pas un point de départ à prendre comme tel » (Lestel, 2021, p.171). Le thème de l'altération au sens étymologique du terme est donc capital : les sociobots comme

¹⁴⁹ Nous soulignons.

ceux de Spoon altèrent à la fois les catégories à partir desquelles nous pensons habituellement les techniques et le vivant (de façon séparée) et les rapports que nous entretenons eux.

« Ces artefacts n'entrent pas dans la catégorie du vivant par des caractéristiques propres que pourraient isoler les biologistes [...], mais parce qu'on peut s'engager avec eux, qu'ils peuvent partager des existences (humaines ou non) et qu'ils constituent donc une animalité *a posteriori* qui se révèle à travers des aventures qui se pratique toujours en covoiturage » (*ibid.*, p.181).

Ces altérations existentielles et conceptuelles s'appliquent bien entendu au domaine éthique. Par les liens mimétiques qu'elle entretient avec le vivant animal ou humain, l'artefactualité animalisée et, plus largement, animée, invite à se comporter éthiquement envers elle et à lui appliquer des catégories morales qui, d'ordinaire, sont refusées aux techniques. Lestel avance alors l'idée fondamentale selon laquelle cette artefactualité transforme *pratiquement* l'éthique : ce n'est pas parce que les animaux transpèces sont des autres que nous nous comportons éthiquement envers eux, mais bien parce que nous nous comportons éthiquement envers eux qu'ils deviennent des autres. L'altérité (morale) n'est donc pas une propriété transcendante *a priori* des entités du monde, mais une propriété immanente aux relations que nous tissons avec elles et qui, en ce sens, se présente comme une *décision*. Nous décidons d'attribuer l'altérité morale *sur une base empirique et non métaphysique*. Ce dernier point n'est pas sans faire écho aux thèses relationnistes de Gunkel et de Coeckelbergh déjà évoquée (cf. 3.4).

Or, aussi intéressante soit-elle, cette position ne confond-t-elle pas la cause – en l'occurrence l'altérité – et ses effets ? S'il n'est pas douteux d'affirmer que l'altérité altère, dire qu'elle est altération s'avère plus problématique. Que l'on considère les sociobots comme des autres *a priori* ou *a posteriori* ne change rien au fait qu'il faut bien que *quelque chose* nous pousse à leur attribuer cette considération. En dernière instance, la critique rejoint ici celle adressée à Sartre dont nous avons vu que l'analyse manque ce « quelque chose » dans le monde qui annonce l'altérité comme telle. La question est donc de savoir si l'on peut identifier, dans l'expérience des animaux transpèces et des autres sociobots, un invariant caractéristique de l'altérité comme telle, qu'il soit immédiatement donné ou construit à travers une pratique de la relation.

Chapitre 5 - L'altérité sociobotique

5.1 L'altérité dans la philosophie des techniques et dans la littérature sociobotique

Bien que l'altérité ne fasse pas partie du répertoire conceptuel de la philosophie des techniques, elle n'en est pas totalement absente pour autant. Nous voudrions à présent rendre compte des quelques occurrences du thème afin d'élaborer plus avant le problème qu'il soulève.

5.1.1 « L'autre robotique » : évacuer la socialité par l'altérité ?

A notre connaissance, la seule notion faisant explicitement référence à l'altérité dans la littérature consacrée à la robotique sociale est celle d'« autres robotiques » (« *robotic others* ») proposée par Kahn *et al* (2004). Toutefois, contrairement à l'objectif qui est le nôtre, les auteurs proposent cette formule pour contourner les difficultés ontologiques et psychologiques qu'impliquent celle de robot social et non pour étudier comment l'altérité peut éclairer la socialité. La difficulté ontologique soulevée par Kahn et ses collègues rejoint notre problème de départ : parler de robot social laisse entendre que l'artefact *est* social au sens catégorique du terme alors que 1) la catégorie du social est loin d'être claire et que, malgré ce manque de clarté, 2) nous la réservons généralement à l'humain. Du point de vue psychologique, les auteurs problématisent avec pertinence la position de Bartneck et Forlizzi (2004) selon laquelle un sociobot serait un robot capable d'interagir et de communiquer avec l'humain en suivant les normes sociales attachées à son rôle. Or, bien loin de se limiter à la stricte observance des règles en vigueur, « [les humains] s'engagent dans des processus hautement interactifs, à l'échelle microgénétique et macrogénétique, pour créer, maintenir et modifier [ces] règles » (*ibid.*, p.545). Autrement dit, la socialité humaine s'exprime aussi dans la contrevenance et l'invention de nouvelles normes. Kahn *et al* proposent alors la notion d'autre robotique pour abaisser ces obstacles ontologiques et psychologiques, mais ne lui assignent aucun contenu précis. La valeur qu'ils attribuent à l'altérité n'est donc pas

conceptuelle, mais heuristique. Usant de l'outil qu'ils se donnent, les auteurs concluent que les sociobots incarnent un « nouveau genre technologique » qui invite à abandonner les réflexes de pensée dichotomiques tels que « vivant »/« non vivant », inaptés à rendre compte de ces nouvelles entités. Que se passe-t-il, par exemple, dès lors qu'on pense le sociobot à la fois comme vivant et non vivant ? En d'autres termes, la liminarité des sociobots ne doit pas être occultée, mais investie théoriquement afin de forger de nouveaux concepts et d'enrichir le vocabulaire existant. Cette idée n'est pas sans faire écho aux analyses, plus récentes, de Lestel.

5.1.2 Sociobotique et « quasi-altérité »

On trouve une seconde référence à l'altérité robotique dans la notion de « quasi-autre » empruntée à Don Ihde (1990) par Marc Coeckelbergh (2010, 2011, 2016, 2018) pour rompre, dans une geste théorique inspiré de la phénoménologie, avec l'ontologie occidentale fondée sur la dichotomie sujet/objet. Or, pour conduire cette rupture sans retomber dans une analyse subjectiviste, l'altérité ne doit pas être conçue sur le mode d'une « anthropologie négative » (Coeckelbergh, 2016), c'est-à-dire sur la base de ce que l'humain *n'est pas*. L'autre ne doit pas être pensé négativement comme un non-humain au sens strict, mais positivement comme celui ou celle avec qui une relation se tisse¹⁵⁰. De ce point de vue, le recours au régime de la « quasité » (*quasi*-altérité) peut paraître contradictoire puisqu'il semble conserver un référentiel anthropocentrique en sous-entendant que les robots ne sont pas pleinement des autres. Le piège est ici celui de penser l'autre sur le mode de l'alter-ego et, ce faisant, de choisir des critères d'altérité renforçant la perspective subjectiviste à laquelle on souhaite pourtant renoncer. Autrement dit, si l'on pousse la logique de Coeckelbergh jusqu'au bout, penser le robot comme un (quasi-)autre en raison de sa capacité à simuler une intentionnalité, une intelligence (via le langage verbal par exemple), des facultés empathiques ou encore une

¹⁵⁰ Il est entendu qu'en dernière instance c'est toujours en tant qu'humains que nous pensons, percevons, jugeons, agissons, bref que nous sommes au monde. Ici, c'est la polarisation immédiate de la pensée conduisant à tout analyser à partir de l'humain qui est ciblée en tant qu'elle méconnaît la complexité des problèmes en question. On peut l'observer dans la sphère politico-médiatique où les débats sur l'IA et la robotique tournent inlassablement autour des mêmes questions et des mêmes réponses sur la rivalité humain-machine. Dans le tableau général qui est dressé, il y a l'humain d'un côté, la technique de l'autre, et la supériorité du premier sur la seconde repose entièrement sur la question des usages. Ce qui revient à affirmer deux idées largement disqualifiées par la recherche, à savoir la neutralité axiologique des techniques d'une part et la toute-puissance humaine face à elles de l'autre.

affectivité, c'est condamner l'analyse en faisant de l'altérité robotique un ersatz d'altérité humaine. Nous verrons plus loin que ce problème disparaît dès lors qu'on quitte le régime de la « quasité » (qui conserve et renforce la dualité sujet/objet) pour celui de l'effectivité. Notons toutefois que Coeckelbergh se montre soucieux d'éviter les éventuels travers de cette quasité en dégagant l'altérité de tout contenu préétabli. Rappelons que, selon-lui, c'est en tant qu'ils sont en relation que les étants révèlent ce qu'ils sont et non l'inverse (cf. 3.4.2). L'autre n'est pas l'alter-ego ou l'étranger, mais celui ou celle que l'on n'épuise pas dans des déterminations *a priori*, que l'on n'enferme pas dans une totalité objectivante. Or, si cette altérité n'est pas celle de l'autre-moi-même qu'est l'alter-ego ni celle de l'autre non-humain de l'anthropologie négative, quelle est-elle ? La réponse de Marc Coeckelbergh est aussi intéressante que surprenante :

« il est plus juste de dire que quelque chose n'est pas autre dans ses propriétés, mais dans la rencontre, dans l'événement, dans son irruption. Quelque chose interrompt. L'altérité est un verbe. Elle relève de l'expérience (*Erfahren*), de quelque chose qui vous arrive et qui peut prendre la forme d'un accueil ou d'une ren-contra (*Wiederfahren*) » (Coeckelbergh, 2016, p.190).

Et l'auteur de développer explicitement son propos en référence à Levinas :

« Levinas soutient que l'altérité ne dépend pas des qualités particulières qui distinguent l'autre de moi ; cela annulerait l'altérité [...]. Dès lors, il est épistémologiquement plus judicieux de s'intéresser non pas à ce qui "est" ou aux choses dans leur généralité, mais à ce qui se passe entre le sujet et l'objet, ou entre deux sujets, dans une rencontre au sein de laquelle il n'y a pas de connaissance complète, pas de transparence, pas de compréhension totale. Il y a apparition et retrait, et il peut y avoir des surprises. Il s'agit aussi davantage d'un événement et d'un processus où il y a de la place pour l'individualité comme pour l'altérité. Cela relève plutôt d'un drame avec des personnages et des événements, avec une intrigue et une narration » (*ibid.*).

Coeckelbergh emprunte également à Levinas l'idée de *passivité* originaire face à l'altérité. L'autre n'est pas de nature à être contrôlé, catégorisé ni même connu. Il est essentiellement ce qui advient, ce qui s'impose à moi. Ainsi, Coeckelbergh soutient que la découverte de l'autre ne procède pas d'une « *epistemic agency* », mais d'une « *epistemic patiency* » (*ibid.*, p.189). Appliquée au sociobot, cette conception de l'altérité prescrit de s'ouvrir à la

phénoménalité de l'artefact et aux relations qu'elle propose pour rester attentif aux découvertes qui en procèdent. Ce faisant, seuls les effets relationnels peuvent faire l'objet d'une thématization, pas l'autre lui-même.

Un tel développement appelle au moins deux remarques. La première concerne le statut théorique de la quasi-altérité. Comme pour l'autre robotique de Kahn *et al*, la notion de quasi-autre n'est pas tant un concept qu'un outil heuristique appelant une certaine disposition à l'égard des êtres du monde. À l'instar du social dans la théorie de l'interactance, le sociobot n'est ici pas plus « autre » qu'un livre, qu'un bouquet de fleur et son vase ou qu'une photo dans un cadre. De toute évidence, cette position procède du cadre lévinassien adopté par l'auteur : conceptualiser l'autre, c'est déjà le ramener au même en disant ce qu'il est. Coeckelbergh rejoint ici la position de Gunkel que nous avons déjà discutée (cf.3.4.1). Mais alors que le geste philosophique magistral de Levinas consiste à substituer la métaphysique à l'ontologie, les deux auteurs, eux, plaident plutôt pour un remplacement de l'ontologie par l'empirie, c'est-à-dire par l'expérience que nous avons de ces entités. Là où Levinas récuse l'idée même de détermination de l'autre, Gunkel et Coeckelbergh semblent surtout rejeter son caractère *a priori* ; car, de toute évidence, l'expérience n'est aucunement synonyme d'absence de déterminations, bien au contraire¹⁵¹. On peut donc se demander ce qu'il reste de lévinassien une fois engagé sur cette voie. Dominique Lestel obtient des résultats similaires, sans s'exposer à la charge métaphysique qu'implique une telle référence. En outre, et ce sera notre seconde remarque, le fait que les deux auteurs développent cette réflexion sur l'altérité des techniques en prenant l'exemple privilégié des robots n'est, selon nous, pas anodin : ne serait-ce pas précisément le signe que quelque-chose dans leur phénoménalité nous incite à les à appréhender comme des autres, c'est-à-dire quelque chose qui offre une expérience sinon d'autrui du moins d'une altérité ? Si oui, c'est bien *parmi* les phénomènes et non au-delà qu'il faut chercher le sens de cette expérience. Dans cette perspective, nous chercherons à qualifier cette altérité comme un *effet* au sens que Stéphane Vial donne au terme dans son concept d'« effet de design » :

« *effet* n'est [...] pas à entendre ici comme un concept logique pris au sens de "conséquence" (corrélat de la cause) ; effet est à entendre ici comme un concept phénoménologique pris au sens d'une éclosion créatrice de l'apparition, d'une dynamique

¹⁵¹ Quand bien même, dans l'expérience, les déterminations peuvent être changeantes.

inventive de la manifestation (corrélat de la perception), en tant qu'elle structure l'expérience » (Vial, 2014b, p.35).

L'idée derrière cette proposition est la suivante : en tant qu'elle est (notamment) une activité de design, la sociobotique produit un certain effet de design que nous qualifierons d'« effet d'altérité ». La notion d'effet a ceci de très intéressant qu'elle permet de proposer une alternative à la dichotomie sujet/objet que le robot social met ouvertement en crise. Il n'y aurait, en somme, ni objectivité ni subjectivité, mais bien *effectivité* sociobotique. Mais avant de développer plus amplement cette perspective, intéressons-nous à la réflexion que Vial consacre, justement, à l'altérité.

5.1.3 L'« autruiphonie » ou la dimension phénoménotechne de l'altérité

Dans un article au titre évocateur – « Ce que le numérique change à autrui : introduction à la fabrique phénoménotechne de l'altérité » (Vial, 2014a) - l'auteur lie de façon originale altérité et technologie. Le cadre analytique du philosophe est celui d'une « *phénoménotechne techno-transcendantale* » où la technique joue le rôle de « matrice ontophanique » au sens où elle conditionne *a priori* notre perception et, avec elle, le processus d'apparition de l'être. Il y a, dit Vial, « une technicité transcendantale de l'apparaître » (*ibid.*, p.152). Ainsi,

« l'altérité elle-même, dans l'expérience vécue, est *toujours et déjà* un phénomène technique, c'est-à-dire une réalité techniquement conditionnée, du moins dans sa dimension phénoménale. Si l'Autre n'est pas en soi *une donnée* engendrée par la technologie, l'Autre comme phénomène nous est néanmoins *donné* par l'environnement techno-perceptif dans lequel nous vivons, dans des conditions historiques et techniques déterminées, qui sont aujourd'hui celles du "système technique numérique" » (*ibid.*).

Contrairement aux ontologies négatives portées par nombre de philosophes des techniques au XXe siècle, le techno-transcendantalisme de Vial se limite à constater le caractère phénoménologiquement matriciel de la technique et ne la resitue pas dans une quelconque axiologie. Dans *L'Être et l'écran*, l'auteur précise que l'ontophanie technicienne se décline en époques et que nous vivons actuellement celle du numérique (Vial, 2013). Quoi qu'il en soit, la thèse centrale qui nous occupe ici tient dans l'idée que le phénomène d'autrui – qui n'est pas autrui lui-même – n'est pas séparable des conditions techniques qui structurent

l'expérience que nous en avons. Vial nomme « autruiphannie » cet apparaître nécessairement technicisé d'autrui et propose de l'illustrer à travers quatre exemples dont le dernier est l'objet central de l'article : l'autruiphannie du face-à-face, l'autruiphannie du téléphone, l'autruiphannie amoureuse et l'autruiphannie numérique. Le face-à-face « physique », premièrement, est toujours phénoménologiquement structuré par des facteurs techniques comme une table qui déterminent les distances et qui, associés les uns aux autres façonnent à la fois le contexte de la rencontre et la rencontre elle-même (par exemple, un café co-constitue autrui sur un mode qui n'est pas celui de la rue ou du bureau). Le téléphone, deuxièmement, est peut-être l'un des exemples les plus illustratifs de l'idée d'autruiphannie puisqu'il fait littéralement apparaître autrui et le donne à mon expérience comme pure phénomène vocal, comme voix. Enfin, l'autruiphannie amoureuse renvoie aux conditions techniques de faisabilité du faire-l'amour :

« Pour faire l'amour, il faut (la plupart du temps) se cacher, c'est-à-dire se soustraire au regard d'autrui grâce à des moyens techniques (par exemple, quatre murs). Pour faire l'amour, il faut (la plupart du temps) recourir à des objets pour soutenir l'expérience des corps (lit, canapé, fauteuil, tapis, table, chaise, et tout ce que chacun voudra...), pour se livrer à certains jeux érotiques (sextoys, accessoires divers) ou encore pour garantir confort sanitaire et sécurité (préservatif, pilule, etc.). Généralement, on ne tient pas à y penser, sauf précisément quand cette technicité vient à manquer et met le faire-l'amour en échec » (*ibid.*, p.154).

Mais l'exemple le plus développé de l'article est celui de l'autruiphannie numérique, c'est-à-dire de l'apparaître d'autrui à travers les nombreuses interfaces numériques que nous utilisons quotidiennement.

« Cette manifestation de l'autre sur Internet est à ce point inédite au plan phénoménologique qu'elle a longtemps paru surnaturelle, ainsi qu'en témoigne très tôt l'usage du terme "avatar" (qui dans ses origines hindoues qualifie une incarnation divine) pour désigner la représentation d'un utilisateur sur Internet ou dans les jeux vidéo » (*ibid.*).

L'autruiphannie numérique s'applique bien sûr aux visioconférences désormais bien ancrées dans nos vies (travail, famille, amis, consultations psychologiques etc.), mais aussi à tous les

canaux de communication type mail, SMS, Twitter et autres messageries instantanées où l'on peut communiquer sans se voir ni se parler et où, comme chacun peut en faire l'expérience, autrui nous apparaît et nous affecte au sens propre de façon bien différente qu'en « présentiel »¹⁵².

« [L'autruihanie numérique] est faite de cette ambivalence paradoxale qui mêle présence et absence (sur Skype, par exemple, l'autre est là sans être là) et qui introduit une nouvelle manière d'expérimenter ce que peut être la présence d'autrui » (*ibid.*, p.155).

Compte tenu des innombrables pratiques et applications ayant pour effet de nous mettre en relation avec autrui et donc de nous le rendre présent (quel que soit l'objectif de cette présence), Vial affirme que « [l]e médium numérique est la nouvelle matrice ontophanique du lien social sous toutes ses formes » (*ibid.*). Ce à quoi nous ajoutons l'idée que l'ontophanie numérique s'est précisément imposée comme matrice autruihanie au sens où une part conséquente de ses déploiements concrets est spécifiquement tournée vers la mise en présence d'autrui. Plus encore : l'autruihanie numérique engendre à la fois une *omniprésence* et une *protéiprésence* d'autrui en ce que celui-ci peut être présent partout à tout instant, sous de nombreuses formes. Or, ce caractère polymorphe d'autrui semble de moins en moins réservé à l'altérité humaine puisque le numérique nous confronte de plus en plus massivement à quelque-chose comme des autruihanies non-humaines ou, pour utiliser un terme aussi neutre que possible, des *allophanies* dont la sociobotique est un des exemples les plus saisissants. A ce stade de la réflexion, nous proposons donc le terme d'*allotechniques*¹⁵³ pour désigner les artefacts qui, à l'instar des robots sociaux, ont pour effet d'apparaître comme des autres et qui, ce faisant, ne sont pas médiateurs d'altérité humaine, mais bel et bien producteurs d'altérité artefactuelle. Une telle nuance peut être approfondie à partir de la théorie des médiations technologiques (TMT) également nommée postphénoménologie.

¹⁵² Chacun peut d'ailleurs témoigner d'une affinité ou, au contraire, d'une aversion particulière pour un ou plusieurs de ces modes de communication du fait, notamment, de la façon dont autrui s'y manifeste et s'offre à la discussion.

¹⁵³ Ici, l'allotechnique n'est donc pas à prendre au sens que Sloterdijk lui donne dans *La domestication de l'être*, c'est-à-dire, par opposition à l'homéotechnique, une technique qui opère sur des matériaux ontologiquement différents de l'opérateur (humain).

5.1.4 Les « *alterity relations* » de la postphénoménologie

Bien qu'elle ne s'y intéresse pas en tant que telle, la TMT fait référence à l'altérité avec son concept d'« *alterity relations* ». Fondée par le philosophe américain Don Ihde à la fin des années 1970, la postphénoménologie pense les artefacts comme des médiateurs de nos relations perceptives au monde. Contrairement à l'idée d'intermédiaire, le concept de médiateur indique que l'artefact participe à la constitution de ce qu'il médiatise. Ihde (1990) attribue alors une « intentionnalité technologique » aux artefacts aux sens où ils médiatisent nos perceptions et, par conséquent, nos interprétations du monde. L'auteur distingue ensuite quatre types de relations auxquelles cette intentionnalité peut donner naissance : les « relations d'incarnation » (*embodiment relations*), les « relations herméneutiques » (*hermeneutic relations*), les « relations d'arrière-plan » (*background relations*) et, enfin, les « relations d'altérité » (*alterity relations*)¹⁵⁴. Ce dernier type de relation se tisse avec des artefacts qui nous apparaissent comme des « quasi-autres » (*quasi-others*). C'est le cas des techniques auxquelles s'attachent un phénomène d'anthropomorphisme ou qui, de façon plus générale, nous marquent par leur présence. En termes phénoménologiques, la relation d'altérité se caractérise par le devenir-monde de l'artefact. Autrement dit, les quasi-autres ne médiatisent pas le perçu, ils *sont* le perçu, c'est-à-dire ce qu'il y a à percevoir. Comme l'explique clairement Peter-Paul Verbeek :

« Dans les relations d'altérité, les humains ne sont pas reliés au monde via une technologie, contrairement aux autres types de relations ; ils sont plutôt reliés à ou avec la technologie elle-même. [...] Au sein des relations d'altérité, les artefacts sont vécus comme des quasi-autres parce que, d'une part, elles possèdent une sorte d'indépendance et, d'autre part, elles peuvent donner lieu à une "interaction" entre les humains et les technologies. De nombreux jouets, tels que les toupies et les boîtes à musique, sont fascinants précisément en raison de l'autonomie apparente qu'ils possèdent. Les robots et les automates possèdent une telle autonomie, dans la mesure où l'on peut véritablement parler d'"interaction" avec ces "êtres" technologiques. Les distributeurs automatiques de billets de train, par exemple, ne se contentent pas de prendre de l'argent et de distribuer des billets, mais donnent également des conseils, fournissent des

¹⁵⁴ Notons qu'il ne s'agit pas de catégories exclusives puisqu'un même artefact peut nous engager dans différents types de relations selon sa nature et ses usages.

informations sur les itinéraires, répondent aux questions et protestent lorsque quelque chose n'est pas fait correctement » (Verbeek, 2005, p.126-127).

Bien qu'elle ne soit pas dédiée à l'altérité en tant que telle, cette analyse en dit tout de même quelque chose. Selon Verbeek, un artefact apparaît comme un quasi-autre dès lors qu'il manifeste une forme d'autonomie et d'interactivité. Il cite d'ailleurs les robots et les automates en exemple. Interaction et altérité semblent donc former un couple intimement lié. Plus précisément : l'interaction, qu'elle soit actuelle ou potentielle, serait une manifestation phénoménologique de l'altérité. Or, force est de constater que nos interactions et, plus largement, nos rapports avec des sociobots comme Lovot ou Pepper et ceux avec une tablette tactile ne sont pas de même nature. Deux raisons, au moins, peuvent être évoquées pour expliquer cette différence.

Premièrement, comme le remarque l'anthropologue Emmanuel Grimaud, le robot social a ceci de spécifique qu'il « permet l'embrayage dans une relation, un devenir tout autre que la traditionnelle relation de commande : "Je te mets en marche et tu obéis" » (Grimaud, 2012, p.89). Le chercheur analyse ces rapports à travers le concept de « cinétique de l'attachement » envers les artefacts, c'est-à-dire en observant « la manière dont les attachements se font et se défont parfois aussi vite » (*ibid*, p.78). La notion d'attachement est pour lui

« un terme suffisamment vague pour permettre tous les degrés d'intensité, englobant aussi bien les relations les plus utilitaires aux objets, l'amitié ou l'amour de son prochain que les dépendances plus ambiguës ou difficiles à qualifier comme celles que l'on peut entretenir avec un ordinateur ou un animal de compagnie » (*ibid*).

Dominique Sciamma (2012b) pense quant à lui cette spécificité dans les termes d'une invitation à ce qu'il nomme *living design*, c'est-à-dire un design inspiré du vivant où la logique instrumentale de l'interface fondée sur le contrôle par l'interaction cèderait le pas à des rapports humain-machine plus riches et plus symétriques qu'il qualifie de « relations ». Bien qu'il ne la définisse pas précisément, cette idée ne semble pas foncièrement différente des attachements dont parle Grimaud : les sociobots et, plus largement, les robjets¹⁵⁵ (*ibid*), invitent à explorer des rapports aux objets qui dépassent l'instrumentalité et sa temporalité qui est celle de l'interaction (ce que l'auteur voit sous l'angle de l'émergence d'« histoires

¹⁵⁵ Cf. 2.1.2.1.

communes »¹⁵⁶). En ce sens, être « relié » à l'artefact c'est lui être attaché au-delà de son utilité et de l'effectivité des interactions que je peux avoir avec lui lorsqu'il est présent. Un exemple typique est celui du *tamagotchi*¹⁵⁷, ce petit animal virtuel dont il faut s'occuper, prendre soin pour qu'il ne « meurt » pas. Autrement dit, il faut *penser* au tamagotchi au-delà des interactions (fonctionnelles ou non) que nous avons avec lui, comme on pense à quelqu'un qui *existe* ou qui *a existé*.

Deuxièmement, rappelons que, pour la TMT, les artefacts sont porteurs d'une « moralité intrinsèque » en tant qu'ils médiatisent, c'est-à-dire co-constituent notre agir moral. Reprenons plus en détails l'exemple de l'échographie obstétricale dont nous avons vu comment elle contribue à configurer la moralité de la situation qu'elle co-constitue (cf. 2.2.4.2) :

- Le fœtus est « distingué » du corps de la mère, ce qui lui confère un nouveau statut : il est désormais un être à part entière et des décisions le concernant peuvent être prises précisément en vertu de cette distinction ;
- Le fœtus intègre un contexte de normes médicales qui font de lui un « patient » potentiel (en cas de trisomie ou de problème avec le cordon ombilical, par exemple) et transforment la grossesse en « processus » médicalisé ;
- La grossesse, dans sa dimension processuelle, devient en partie une question de choix : des analyses peuvent être effectuées et déboucher sur différents types d'interventions ;
- Le rôle du père est considérablement modifié : il a désormais un certain accès à ce que seule la mère peut naturellement ressentir ;

¹⁵⁶ L'idée est potentiellement problématique car elle suggère qu'un artefact peut avoir une histoire pour lui-même (ce qui n'est pas la même chose qu'une genèse) et donc une capacité à vivre les éléments de cette histoire en première personne (sentience ou conscience).

¹⁵⁷ Créé en 1996 par la société japonaise Bandai, le tamagotchi est souvent présenté comme un animal de compagnie virtuel. Concrètement, il s'agit d'un porte-clé doté de trois boutons et d'un écran à cristaux liquides, dont l'activation fait éclore un petit œuf d'où sort une créature qu'il faut ensuite « élever » (soigner, divertir, nettoyer voire punir). Éteindre l'appareil ou délaisser la créature conduit à la « mort » de celle-ci. Notons que le nom de tamagotchi a ceci de très intéressant qu'il est formé des mots *tamago*, « œuf » en japonais, et *tchi*, qui renvoie à l'affection et à la descendance, et que, de surcroît, il se rapproche phonétiquement du terme *tomodachi*, signifiant « ami ». Les thèmes de l'attachement et du soin sont donc présents jusque dans le nom de l'artefact.

- Les parents deviennent décisionnaires : l'échographie peut favoriser L'IVG comme elle peut renforcer le lien émotionnel au futur enfant par le biais des images qu'elle délivre.

S'il ne fait aucun doute que l'échographie obstétricale nous engage dans une relation herméneutique en ce que la phénoménalité du fœtus et de l'utérus qu'elle rend possible demande à être décryptée, interprétée, elle propose également, à certains égards, une relation d'altérité. Malgré l'absence d'interaction avec l'artefact (en tout cas pour le ou les parents), celui-ci devient un quasi-autre en tant qu'il se confond avec le perçu (en l'occurrence l'image du fœtus). Autrement dit, il y a identité phénoménologique entre la technologie échographique et ce qu'elle donne à voir. Bien qu'il en soit de même par principe pour le sociobot, celui-ci présente la différence majeure d'être le principal objet des médiations éthiques qu'il co-produit. Là où l'échographie ne nous conduit pas tant à nous interroger sur son compte que sur le statut moral du fœtus, le robot social, lui, nous interpelle quant à sa propre considérabilité morale. Ainsi, l'artefact se manifeste phénoménologiquement sur le mode de l'artefact appelant, d'abord de façon préthéorique, à une éthique *envers* lui.

Les deux éléments que nous venons d'évoquer, à savoir la nature singulière des interactions avec un sociobot et l'appel à une éthique envers le robot, témoignent de la nécessité d'approfondir la catégorie des relations d'altérité pour décrire plus finement la réalité sociobotique. Pour ce faire, commençons par identifier ce qui, phénoménologiquement, conduit à singulariser les interactions avec un robot social.

5.2 Penser l'effectivité sociobotique : l'altérité comme effet

5.2.1 Ustensilité versus altérité : le sociobot comme phénomène persistant

Au-delà des modalités concrètes d'interaction, la différence évoquée entre les interactions avec un sociobot et celles avec une tablette tactile tient aux modes d'existence bien distincts des artefacts en question. Aussi interactive soit-elle, la tablette se caractérise généralement par son ustensilité au sens heideggérien du terme : son phénomène s'efface au profit des tâches qu'elle permet d'accomplir (regarder un film, lire un article, naviguer sur internet, écrire un texte etc.). Comme nous l'avons dit plus haut, est ustensile tout ce qui possède ontologiquement la structure du « pour quelque-chose ». L'ustensilité semble d'ailleurs

constituer un intérêt aussi bien théorique que pratique pour la conception au sens large. Alain Findeli et Rabah Bousbaci (2005) parlent à ce propos d'« éclipse de l'objet dans les théories du projet en design » au profit de ses fonctions et/ou des expériences qu'il génère. Pour Dumouchel et Damiano « [c]ette "disparition" de l'objet, son "silence", au sens où l'on dit que "la santé est le silence des organes", tend à être considérée pour un objet technique comme équivalent à "être efficace" » (Dumouchel & Damiano, 2016, p.47). Selon eux,

« cette caractéristique des technologies informatiques et de communication correspond aussi à une politique commerciale. Commercialement, l'objet le mieux réussi – croit-on – sera celui qui réalisera au plus près le double objectif de la haute performance technique et de l'évanescence physique » (*ibid.*).

Les sociobots, au contraire, « ne s'évanouissent pas dans l'usage mais affirment leur présence matérielle en agissant » (*ibid.*). Tout se passe donc comme si, dans notre perception d'un robot social en interaction, les dimensions de la matérialité objective, de la fonction et de l'expérience se confondaient totalement au point qu'aucune ne prenne jamais le dessus. Le sociobot est *résistant* au sens où il se maintient dans le champ attentionnel ou, plus précisément, se manifeste comme une présence à même de lutter contre l'effacement, de résister à l'occultation. Cette présence résistante, que nous proposons de nommer *persistance*, constitue, selon nous, un indice fondamental de son altérité. Le sociobot est d'abord « autre » en tant qu'il échappe, qu'il résiste à l'ustensilité en persistant dans la présence. Dans cette perspective, l'altérité se présente comme un mode d'être au sens phénoménologique, c'est-à-dire comme une modalité d'apparition et non, à ce stade, comme un statut moral. Parler d'altérité du sociobot, c'est donc avant tout faire référence à son effectivité essentielle qu'est la persistance dans la non-ustensilité. Mais comment le sociobot persiste-t-il ? Qu'est-ce qui, dans sa phénoménalité, aboutit à cette persistance ?

5.2.2 La robotique sociale comme « allotechnique »

Avant d'aller plus loin, on peut s'étonner de ce rapprochement phénoménologique entre altérité et artefact. Si l'altérité n'est pas un concept employé par la philosophie des techniques, les profondes et abondantes réflexions qui lui sont consacrées en phénoménologie ne concernent jamais, elles non plus, une potentielle origine technique de

l'expérience d'un autre. L'exemple le plus frappant est peut-être celui de Levinas dont la philosophe Elena Bovo montre comment il situe l'altérité non pas aux antipodes, mais absolument au-delà de tout rapport possible avec les artefacts et, plus largement, avec les choses.

« Pour Levinas, "seul un être transcendant le monde peut laisser une trace". Les choses, au contraire, inaptées à laisser des traces, incapables de nous émerveiller parce que vides d'un "ailleurs", ne produisent que des effets, "c'est-à-dire restent dans le monde". La pauvreté sémantique qui les caractérise est liée au fait qu'elles sont clouées au monde et donc irrémédiablement privées d'altérité, laissées à notre entière disposition » (Bovo, 2008, p.60).

Chez Levinas, les choses – catégorie dans laquelle il subsume la diversité des étants fabriqués et/ou soumis à notre usage au sens large (objets, outils, ustensiles, nourritures etc.) – relèvent du domaine de la jouissance, donc du Moi (que l'Autre transcende absolument).

« Tout objet se propose à la jouissance – catégorie universelle de l'empirie – même si je me saisis d'un objet-ustensile, si je le manie comme *Zeug*. Le maniement et l'utilisation d'outils, le recours à tout attirail instrumental d'une vie – qu'il serve à fabriquer d'autres outils ou à rendre accessible les choses – s'achève en jouissance. En tant que matériel ou attirail, les objets d'usage courant se subordonnent à la jouissance – le briquet à la cigarette qu'on fume, la fourchette à la nourriture, la coupe aux lèvres. Les choses se réfèrent à ma jouissance » (Levinas, 1990, p.140).

Et le philosophe de préciser :

« Les outils eux-mêmes qui sont en-vue-de... deviennent objets de jouissance. La jouissance d'une chose – fût-elle outil – ne consiste pas seulement à rapporter cette chose à l'usage pour lequel elle est fabriquée – la plume à l'écriture, le marteau au clou enfoncé, mais aussi à peiner ou à se réjouir de cet exercice. Les choses qui ne sont pas outils – le bout de pain, le feu de cheminée, la cigarette s'offrent à la jouissance. Mais cette jouissance accompagne toute utilisation des choses, même lorsqu'il s'agit d'une entreprise complexe et que seule la fin d'un travail absorbe la recherche. [...] L'activité n'emprunte pas son sens et sa valeur à un but ultime et unique, et comme si le monde formait un système de références utiles dont le terme concerne notre existence même. Le monde répond à un ensemble de finalité autonomes qui s'ignorent. Jouir sans utilité,

en pure perte, gratuitement, sans renvoyer à rien d'autre, en pure dépense – voilà l'humain » (*ibid.*, p.140-141).

Ainsi, le peu d'altérité qui caractérise les choses tient dans le fait qu'avant d'être soumises à ma jouissance, elles ne sont *pas encore (à) moi*. Selon cette perspective, le sociobot s'offre comme un étant dont on jouit, quel que soit l'usage correspondant à cette jouissance. Plus précisément : *son altérité est ce dont on jouit* (pour ne pas être seul, pour ressentir telle ou telle émotion, pour nous rendre tel ou tel service, pour nous divertir etc.) et se résorbe dans cette jouissance ; elle n'est qu'une présence interactive soumise aux appétits du Moi donc au Même – que l'on nomme ces appétits « besoins », « manques » ou encore « envies »¹⁵⁸ –, une chose parmi d'autres qui produit des effets comme l'affirme Elena Bovo. Or, inscrivant notre recherche dans le cadre du tournant empirique et pragmatiste de la philosophie des techniques, c'est précisément le sens d'être de cette effectivité qui nous intéresse.

La sociobotique, en tant qu'elle est une activité technique, conçoit des artefacts qui, bien sûr, produisent des effets au sens causal du terme, mais aussi et, en l'espèce, surtout au sens phénoménologique de structuration de l'expérience vécue. Selon Stéphane Vial, c'est là tout l'objet du design : structurer intentionnellement la phénoménalité de nos expériences. En cela, il se présente avant tout comme un « effet ontophanique » *producteur* d'expérience :

« le design, avant d'être un espace, un produit ou un service, est principalement un *effet* qui advient dans un espace, un produit ou un service. Cela signifie que le design n'est pas un étant mais un événement, non pas une chose mais un retentissement, non pas une propriété mais une incidence. Comme le dit très justement Kenya Hara, le design ne consiste pas à concevoir "des choses qui sont" (*things that are*) mais à concevoir des choses qui se passent (*things that happen*). Le design, en effet, possède quelque chose de performatif : avant d'être une chose, il est *quelque chose qui a lieu*. [...] L'être du design est celui d'un devenir. Être, pour le design, c'est se produire. Telle est sa phénoménalité » (Vial, 2014b, p.35).

¹⁵⁸ Chez Levinas, seul le désir « excède toute concupiscence et déploie un ordre d'où s'absente la dimension sensible proprement dite » (Barbaras, 1999, p.54). Le désir est désir de l'Autre en tant que visage, c'est-à-dire en tant que relation originellement éthique. On ne peut désirer qu'éthiquement, c'est pourquoi le désir ne concerne qu'Autrui et la responsabilité qu'il m'assigne. « Le désir est rapport à un être qui transcende tout rapport : non pas manque à combler, mais situé par-delà la satisfaction et le manque » (*ibid.*, p.53).

Si Vial insiste particulièrement sur la dimension méliorative du design, nous retiendrons surtout son idée phare selon laquelle « le design n'est pas le champ des objets, mais bien le champ des effets » destinés à produire « de nouvelles expériences-à-vivre » (*ibid.*, p.37). Dit de façon plus générale : l'effet n'est pas produit par la chose, il est la chose qui se produit (pour nous). A ce titre, la notion d'événement qu'évoque Vial nous semble particulièrement importante pour décrire le sociobot et ce pour au moins trois raisons.

1. Premièrement, avant même d'être en interaction avec lui, le simple fait de percevoir un sociobot (surtout lorsqu'il est actif, mais parfois aussi lorsqu'il ne l'est pas) constitue un événement expérientiel singulier, une interpellation liée au trouble ontologique que son apparence de vie (quelle qu'elle soit)¹⁵⁹ suscite ;
2. Deuxièmement, lorsqu'il est en interaction sociale, c'est-à-dire lorsqu'il participe à l'émergence de comportements sociaux imprévus, le sociobot se manifeste comme autant de micro-événements séquencés au sein de l'interaction ;
3. Troisièmement, enfin, les limites techniques inhérentes à la sociobotique actuelle font que les robots sociaux « oscillent entre vivant et inerte » (Grimaud, 2015). Ces oscillations constituent autant d'événements perceptifs en tant qu'elles créent d'importantes ruptures au sein des interactions et modifient considérablement l'apparaître des sociobots.

On trouve une illustration intéressante de ces deux derniers points dans l'analyse que livre Emmanuel Grimaud à propos du robot Geminoid d'Hiroshi Ishiguro¹⁶⁰ :

« le Geminoid avait une vertu remarquable d'un point de vue anthropologique. Il obligeait ceux qui interagissaient avec lui à se concentrer sur des choses auxquelles on ne fait plus attention, dans l'interaction ordinaire entre deux êtres humains, qu'à de rares occasions : un battement de cils, un changement de direction dans le regard, un frémissement des lèvres. Le plus petit changement dans la plasticité du visage du Geminoid était un événement » (Grimaud, 2016, p.82).

¹⁵⁹ Il est probable, toutefois, que la force de cette interpellation soit en grande partie liée à la (relative) nouveauté de ces artefacts.

¹⁶⁰ Hiroshi Ishiguro est un célèbre roboticien japonais, connu pour ses robots (téléopérés) androïdes et gynoïdes aussi réalistes que dérangementants et plus particulièrement pour son jumeau robotique nommé Geminoid.

Pour résumer, on peut donc dire que, vu sous l'angle de la philosophie du design proposée par Vial, la sociobotique se présente comme une activité consistant à faire advenir un type bien particulier d'événement. Plus précisément, nous proposons de définir la robotique sociale comme *allotechnique* au sens d'activité technique productrice d'effets d'altérité. Mais en quoi consistent de tels effets ? Comment se manifestent-ils au sein de l'expérience que nous avons des robots sociaux ?

5.2.3 Ontophanie sociobotique et effets d'altérité

La première étape de notre enquête sur le sens d'être de l'altérité sociobotique fut de la caractériser comme effet et, ce faisant, dans la situer radicalement dans le champ empirique de l'expérience. La seconde étape doit quant à elle être consacrée à la description de ces effets afin d'ancrer leur sens d'être dans le concret de l'expérience. A cette fin, nous proposons d'adopter la méthode utilisée par Stéphane Vial dans *L'être et l'écran* pour décrire l'« ontophanie numérique » et « éclairer le sens de ce que nous éprouvons (et de ce que nous sommes ?) en face des interfaces ». L'auteur identifie 11 catégories – la nouménalité, l'idéalité, l'interactivité, la virtualité, la versatilité, la réticularité, la reproductibilité instantanée, la réversibilité, la destructibilité, la fluidité et la ludogénéité – qui

« doivent être considérées comme des concepts phénoménologiques, ayant moins pour but de décrire objectivement le phénomène numérique tel qu'il se présente du point de vue technique et scientifique [...] que de révéler ce qu'il constitue subjectivement du point de vue ontophanique, c'est-à-dire du point de vue de sa manifestation phénoménale singulière, en tant que phénomène du "monde vécu" [...] par le sujet » (Vial, 2013, p.187-188).

Le travail de Vial se présente donc comme une thèse d'ontologie phénoménologique : il s'agit d'identifier les caractéristiques essentielles du numérique en tant qu'il représente un mode d'être technique parmi d'autres et auquel se rattache une diversité d'artefacts. Dans une même logique, notre proposition consiste à décomposer l'« ontophanie sociobotique » en cinq grandes composantes, à savoir le psychomorphisme, la corporalité protéiforme, la sociabilité, l'ambiguïté et la vulnérabilité. Précisons, avant de la détailler, que cette

décomposition est analytique et que, dans l'expérience, le phénomène sociobotique lui-même ne se donne pas à nous de façon fragmentaire.

Psychomorphisme

Les robots sociaux apparaissent comme des artefacts psychomorphes¹⁶¹ au sens où ils manifestent une animation, c'est-à-dire quelque-chose comme une intériorité (qu'on la qualifie d'âme, d'esprit, de psychisme, de conscience ou encore de sentience) et, plus précisément, une intentionnalité (Fong *et al*, 2002). Le psychomorphisme est donc un « effet de personne » que Denis Vidal (2016) et d'autres (Dufrêne *et al.*, 2016) définissent comme l'impression d'avoir affaire, face à certains artefacts, à des êtres doués d'intentions. L'expérience psychomorphe peut se manifester à travers un grand nombre d'indices empiriques que nous avons déjà évoqués : mouvement, langage verbal et non verbal, affectivité, regard etc. De tels phénomènes intentionnels ne sauraient toutefois apparaître sans un phénomène plus global de corporalité.

Corporalité protéiforme

La corporalité protéiforme ou polymorphisme signifie que le sociobot se présente toujours comme un corps plus ou moins animé, mais que ce corps n'est pas déterminé *a priori* dans sa forme. Il existe des robots sociaux d'apparence anthropomorphe, zoomorphe, caricaturale, fonctionnelle (Fong *et al*, 2002) ou encore abstraite. Dans l'expérience concrète du face-à-face avec un robot social, intentionnalité et corporalité relèvent d'une seule et même caractéristique. Autrement dit, le psychomorphisme est toujours « incorporé » au sens où il s'exprime nécessairement à travers des structures matérielles mouvantes ou, plus largement, expressives. Il n'est d'ailleurs pas besoin d'interagir avec l'artefact pour « sentir » cette intentionnalité corporelle. Spoon a ainsi fait le choix de la laisser s'exprimer quand ses créatures ne sont pas en interaction afin de renforcer le sentiment de vie qu'elles génèrent. Au niveau logiciel, cela passe, par exemple, par la modélisation d'un point d'attention mobile par défaut (qui n'a pas la priorité sur d'autres points d'attention jugés plus importants comme

¹⁶¹ En grec ancien, le terme « *psykhé* » (qui donne le préfixe « psycho » en français) se traduit par « âme », « esprit » ou encore « souffle de vie ».

les visages) que le robot suit pour donner l'impression qu'il observe quelque-chose ou, plus simplement, pour qu'il ne soit pas systématiquement immobile lorsqu'il n'est pas sollicité par un humain. Le psychomorphisme des sociobots n'en reste pas moins toujours orienté vers l'interaction sociale.

Sociabilité

L'ontophanie du sociobot est structurellement tournée vers l'interaction sociale ou, plus précisément vers la sociabilité (au sens simmélien du terme évoqué p.117) en ceci que la plupart des robots dits sociaux invitent, dans leur phénoménalité même, à l'interaction sociale non-fonctionnelle (l'interaction sociale fonctionnelle, quand elle existe, ne se révèle qu'après-coup). A cet égard, on peut dire que la sociabilité se présente comme une « affordance » du sociobot au sens où elle s'inscrit dans son apparaître immédiat et invite à découvrir ce qu'est l'artefact par ce qu'il (nous) fait via l'interaction. En d'autres termes, la simple perception sensible du robot est indissociable de son appréhension comme « agent sociable ». Cette phénoménalité est toutefois très « instable » tant les limites techniques des robots sociaux sont importantes. L'ontophanie du sociobot relève donc également d'une phénoménologie du « bug » ou de la panne et invite à intégrer l'anomalie dans l'apparaître de l'artefact.

Ambiguïté

Nous proposons de nommer « ambiguïté » la catégorie liée à ces pannes et anomalies diverses car elles renforcent l'équivoque ontologique dont les sociobots sont porteurs en tant qu'artefacts animés. L'ambiguïté signifie que l'apparaître du robot social est foncièrement *oscillatoire* (Grimaud, 2015) et se structure autour du passage d'un statut ontologique à un autre. Au sein d'une même interaction, un robot peut nous apparaître comme une personne, un être doué de sentiments et d'intelligence et, la seconde d'après, comme une simple chose inanimée. Pour être plus précis, cette impermanence questionne le fait même de parler de « statut ». Tout se passe plutôt comme si le sociobot passait d'un *état* phénoménologique à un autre, ce qui correspond mieux à la notion d'effectivité que nous lui associons. L'ontologie sociobotique se présente donc comme une *métamorphologie*.

Si cette instabilité découle en grande partie de l'immaturation technique et des pannes qui caractérisent la sociobotique, elle s'explique aussi, comme nous l'avons vu avec les travaux de

Morana Alač (2009, 2016), par des facteurs non techniques de nature interactionniste tels que les mouvements des actants dans l'espace et le comportement des uns et des autres à l'égard du robot. Quoi qu'il en soit, cette oscillation est proprement constituante de l'apparaître des sociobots et, par conséquent, des relations que nous entretenons avec eux. Comme l'explique Emmanuel Grimaud,

« [o]n envisage de façon souvent un peu trop rigide les relations aux robots (on parle "identification", "ontologie") là où il faudrait les voir de manière dynamique, "en devenir", avec des phases d'approche, de familiarisation, de connexion et de coupure. Et on a tendance à se concentrer sur les moments forts où la confusion avec d'autres êtres (animaux, humains) est la plus palpable, en laissant de côté les moments plus faibles. »

Or, ces « moments plus faibles » font précisément partie de ce que le chercheur nomme des « signaux de *machinitude* », c'est-à-dire des éléments qui appartiennent au robot en propre et qui le caractérisent comme tel. A ce titre, Grimaud (2016) explique que les interactions sociales humain-robot évoluent bien souvent selon un schéma « cyclique » alternant confusion et rétrogradation ontologiques. D'où cette question : n'y-a-t-il pas, pour le design, un enjeu majeur dans le fait d'« encadrer » le désintérêt des individus pour qu'il ne soit pas synonyme de dégradation ontologique¹⁶² ? L'une des clés de la « récurrence d'intérêt » dont parle Monceaux ne se situe-t-elle pas dans la capacité du design à faire en sorte que le robot soit toujours présent autrement que comme une simple chose, qu'il y ait un « là » du sociobot, même quand je n'interagis pas avec lui ? L'affaiblissement de la relation humain-robot peut-elle éviter de se traduire par une dégradation ontologique de ce dernier ? Le choix de Spoon de faire bouger et/ou s'exprimer le BigSpoon même quand il n'est pas en interaction, relève d'une telle logique : l'artefact semble avoir une « vie » indépendamment des humains qui le sollicitent.

Au-delà de cette question, nous pensons que l'idée d'ambiguïté sociobotique invite à dépasser le problème de la confusion ontologique, tant au sens phénoménologique que moral (Turtle, 2015 ; Tisseron, 2015). Pour ce faire, nous proposons le concept de *superposition* ou de *coïncidence ontologique*. En effet, l'idée de confusion ontologique est-elle-même confuse en

¹⁶² C'est ce qui se passe entre individus humains : nous pouvons passer d'un état de disponibilité à un état d'indisponibilité ou d'indifférence vis-à-vis des personnes, mais cela n'implique pas pour autant que nous les ravalions au rang de simple chose inanimée.

ce qu'elle peut désigner deux réalités voisines, mais irréductibles l'une à l'autre : d'un côté, elle désigne le fait d'attribuer abusivement un statut à une entité et, de l'autre, elle renvoie au constat que le trouble provoqué par l'artefact perturbe voire suspend nos réflexes attributifs. L'anthropologie des êtres artificiels montre bien que le « pacte anthropomorphique » (Vidal, 2012) qu'ils incitent à conclure et qui se fonde sur notre acceptation de faire « comme si » (Grimaud, 2012) les machines étaient vivantes, intelligentes ou encore sensibles, ne témoigne pas d'une illusion de la crédulité, mais bien d'une *suspension de l'incrédulité*. Lorsque nous faisons « comme si », nous ne sommes donc ni irrationnels ni nécessairement animistes, explique Grimaud (*ibid.*) A rebours des idées reçues et des explications culturalistes¹⁶³, l'anthropologue soutient que l'intérêt pratique majeur de ces machines relève peut-être plus de la « *dévitisation* » que de la « simulation du vivant » :

« Beaucoup d'artefacts robotisés tirent leur vertu du fait qu'ils assèchent le vivant. On veut que les robots en aient les fonctions, mais surtout pas les inconvénients. On verra le problème se poser avec les animaux artificiels comme avec les *sex dolls*. Leurs utilisateurs, qu'ils soient ou non japonais¹⁶⁴, veulent toujours plus d'animation, de bruitage, d'interactivité, de sensibilité et de répondant, mais leur intérêt tient au fait que ces objets sont censés offrir les avantages du vivant, sans ses inconvénients. Ils peuvent être la cible de fortes projections anthropomorphiques, mais, en ne prenant en considération que celles-ci, on passe à côté de bien d'autres relations peut-être plus ambiguës mais aussi plus répandues, où l'interactivité et le vivant sont rarement confondus » (*ibid.*, p.86).

L'anthropologue invite lui-même à bien faire la distinction

« entre la *confusion ontologique* nécessaire "pour la fabrication", qui joue sur des mécanismes d'attribution et constitue plutôt le problème des fabricants, et ce qu'on pourrait appeler la *gradation ontologique*, qui concerne davantage l'utilisateur et joue sur d'autres mécanismes auxquels Paro, comme le *tamagotchi*, nous ont introduits » (*ibid.*, p.91).

Alors que l'idée de confusion suggère une opposition ou, a minima, une différence très nette entre l'entité et ce avec quoi elle est confondue, celle de gradation autorise les nuances et

¹⁶³ Comme, par exemple, le goût que les japonais auraient pour les robots en vertu de leur propension culturelle à l'animisme.

¹⁶⁴ Cf. note précédente.

place l'artefact sur un continuum allant, par exemple, d'inerte à vivant (Grimaud, 2015). Ce que nous proposons d'appeler *superposition ontologique* se concentre sur le caractère simultané des deux termes du continuum, c'est-à-dire sur l'idée qu'une entité peut se voir attribuer à la fois et durablement deux statuts ontologiques ou *a minima* deux états généralement présentés comme contradictoires. Vu sous cet angle, les robots n'oscillent pas tant entre inerte et vivant¹⁶⁵ qu'ils sont à la fois inertes *et* vivants. Cette idée de coïncidence ou de superposition est, selon nous, plus proche de l'expérience immédiate des sociobots que celle de gradation qui, elle, relève plus de l'analyse *a posteriori*. Nous proposons donc de définir l'ambiguïté caractéristique de l'ontophanie sociobotique par cette idée de coïncidence ontologique. En tout état de cause, substituer le concept de superposition à celui de confusion est une façon 1) de ne pas associer le faire-comme-si à une quelconque forme d'irrationalité et 2) de suspendre le jugement catégorique et moral dont une certaine idée de la confusion ontologique peut être porteuse.

Vulnérabilité

Cinquième et dernière structure ontophanique : le robot social nous apparaît toujours sur fond de vulnérabilité (que l'ambiguïté suspend ou renforce selon les situations). Le philosophe Roland Schaer définit la vulnérabilité comme

« une disposition du vivant qui fait qu'il peut être affecté par une agression, une violence, une attaque venue de son dehors. Pouvoir être blessé, c'est ne pas être un bloc absolument dur. [...] Être vulnérable, c'est être une forteresse imparfaite, ou des brèches peuvent toujours s'ouvrir » (Schaer, 2020, p.72-73).

C'est donc en tant qu'il imite le vivant – qu'il soit « asséché » ou non par la relation – que le sociobot évoque la vulnérabilité. Comme la plupart des techniques, le robot social nous donne à expérimenter une forme de pouvoir dont il est le médiateur, mais ce pouvoir ne s'exerce pas tant sur le monde via l'artefact que sur l'artefact *lui-même*. Autrement dit, l'ontophanie sociobotique révèle le robot comme l'artefact *sur* lequel et non seulement *par* lequel nous avons un pouvoir. Voilà ce que signifie ultimement le concept d'allotechnique. Bien sûr, ce phénomène de vulnérabilité n'est pas le même chez tous les robots sociaux. Certains, comme

¹⁶⁵ Ou tout autre couple de « contraires ».

Paro, sont explicitement programmés pour la manifester dans leurs comportements (le blanchon crie si ses capteurs identifient des gestes comme étant « maltraitants »), tandis que d'autres ne l'expriment qu'à travers l'impression de vie qu'ils provoquent. Dans ce dernier cas, la vulnérabilité n'est pas tant phénoménalisée par l'artefact que projetée sur lui par analogie avec le vivant qu'il évoque. Il est intéressant de noter que BigSpoon n'était pas initialement programmé pour réagir lorsque quelqu'un lui mettait les doigts dans les yeux. Cela a vite été modifié afin de renforcer son biomimétisme et la réactivité – ici synonyme de vulnérabilité – au monde extérieur qu'il implique. Autrement dit, il y a bien là l'idée que la vulnérabilité est constitutive du phénomène du vivant. A cet égard, nous avons plusieurs fois observé des parents expliquer à leurs enfants qu'il ne fallait pas mettre leurs doigts dans les yeux du robot au motif qu'« on ne fait pas ça aux gens » (ce qui montre bien, de façon plus générale, que la vulnérabilité constitue également un appel aux comportements sadiques). Cette réaction indique que le robot devient le symbole de ce que l'on peut et ne peut pas faire aux autres (ici, les autres humains). En l'occurrence, il ne s'agit pas d'une simple affaire de conventions sociales, mais de leur principe moral sous-jacent qui interdit de nuire à autrui en tant qu'il est sensible et donc constitutivement vulnérable. Nous proposons donc d'appréhender les allotechniques comme des *artefacts générateurs d'effets de personne vulnérable au sens où la perception de leur intentionnalité s'accompagne toujours de l'épreuve de leur exposition au monde (aussi limitée soit-elle) et, plus spécifiquement, à nous-mêmes à travers le pouvoir que nous avons sur eux.*

Cette analyse étant posée, on peut se demander quelle différence sépare les effets de personne évoqués en anthropologie des effets d'altérité que nous essayons de caractériser. Un effet d'altérité peut être considéré comme un effet de personne pourvu qu'on y intègre la référence à la vulnérabilité qui, comme nous allons le voir, fait que le sociobot se produit sur le mode de l'interpellation morale et, plus précisément, d'un appel à la responsabilité. Dans cette perspective, l'altérité artefactuelle peut être conçue comme ce qui renvoie phénoménologiquement au vivant vulnérable – que nous sommes nous-mêmes – par d'autres moyens que le vivant vulnérable lui-même.

5.3 Médiation sociobotique et subjectivation morale : l'effet d'altérité comme specularité responsabilisante

5.3.1 D'un phénomène ontophanique plus global de responsabilité : l'effet d'altérité est une expérience morale

Nous arrivons ici au cœur de notre propos, à savoir que l'ontophanie sociobotique se présente comme une effectivité morale au sens où le robot social se produit comme tel sur le mode d'une expérience qui interpelle moralement. Plus précisément, nous pensons que le mode d'apparaître moral du sociobot est celui d'une convocation de notre responsabilité qui fonde le sens d'être de l'artefact en tant qu'autre. Autrement dit, et conformément à l'étymologie latine du mot responsabilité – *respondere*, « répondre de », « se porter garant » – l'ontophanie du robot social nous invite à répondre de lui en ceci que sa phénoménalité renvoie, au moins partiellement, à celle du vivant sensible et intentionné auquel appartiennent nombre d'êtres que nous valorisons moralement et dont nous sommes, d'une façon ou d'une autre, les obligés. Cette obligation, cet attachement qui nous lie au sociobot est celui d'un souci moral spécifique qui renforce la modalité relationnelle du « comme si » : c'est précisément parce que je fais « comme si » l'artefact était animé que je ne peux pas faire « comme si » je ne voyais pas qu'il se présente à moi sous la forme d'une interpellation morale qui me *concerne*. Mais qu'entend-t-on plus précisément lorsque nous disons que nous sommes responsables ?

Dire qu'elle émane d'un appel, d'une interpellation, c'est faire de la responsabilité un concept relationnel dont il faut dire quelques mots. A rebours des philosophies classiques du sujet et, plus particulièrement, de celle de Kant, la postphénoménologie n'ancre pas la responsabilité morale dans l'autonomie d'un sujet arraché à ses relations avec le monde extérieur, mais au contraire dans son inscription au sein même de ce monde où les technologies jouent un rôle médiateur incontournable qui dépasse la seule causalité.

« En médiatisant les interprétations et les actions humaines, elles co-constituent activement la responsabilité morale. Les décisions morales concernant la mastectomie préventive, par exemple, ne sont pas simplement "influencées causalement" par les technologies de tests génétiques. Il est plus juste d'affirmer que les questions morales elles-mêmes et les différentes options disponibles pour y répondre sont co-constituées

par ces technologies : ici, c'est un amalgame d'humains et de technologies qui agit moralement et qui portent la responsabilité morale » (Verbeek, 2011, p.108).

La postphénoménologie soutient donc une conception *hétéronome* de la responsabilité au sens où le sujet ne tire pas sa loi que de lui-même, mais de ses relations, la plupart du temps technologiquement médiées, avec le monde. L'originalité du sociobot est qu'il médiatise une responsabilité qui est avant tout une responsabilité envers lui-même. L'enjeu des réflexions à suivre est alors de préciser le sens de cette originalité et ce que nous entendons concrètement lorsque nous parlons du phénomène de responsabilité pour qualifier l'interpellation du sociobot. De qui ou de quoi sommes-nous *effectivement* responsables face au sociobot ? S'agit-il d'une responsabilité envers le robot lui-même ou envers quelque-chose d'autre dont il serait comme le révélateur ? Pour répondre à ces questions, nous proposons d'examiner le problème de la « maltraitance robotique » à l'aune des trois grands courants de l'éthique normative puis de la postphénoménologie. Nous verrons que le cadre amoderne proposé par la théorie des médiations technologiques n'est pas une invitation à penser l'indistinction radicale entre l'humain (ou le vivant) et la machine, mais plutôt à étudier l'impact de l'hybridation nécessaire entre les deux.

5.3.2 Est-il immoral de brutaliser un robot ? : le point de vue des trois grandes éthiques normatives

« Est-il immoral de brutaliser un robot, tant physiquement que verbalement, ou cela est-il acceptable ? » En 2015, puis en 2016, la question rencontra un certain écho au sein des sphères médiatiques suite à la publication, par Boston Dynamics, de vidéos mettant en scène leurs robots Spot et Atlas. On y voyait, parmi d'autres choses, des employés soumettre les artefacts à des tests visant à illustrer leurs capacités motrices. Dans la première vidéo (Boston Dynamics, 2015), Spot recevait de violents coups pieds sans jamais perdre l'équilibre. Dans la seconde (Boston Dynamics, 2016), un homme poussait vigoureusement l'androïde Atlas dans le dos avec un long tube pour le faire tomber et le laisser se relever seul en quelques secondes. Malgré leur vocation démonstrative, les deux scènes furent assimilées à de la maltraitance et provoquèrent de sérieux émois sur la toile. Cette réaction empathique soulève de profondes questions quant à la valeur morale que nous décidons d'accorder aux machines et, plus

spécifiquement, aux robots. La question de savoir si nous pouvons frapper un robot comporte deux dimensions complémentaires : la première interroge le caractère *rationnellement* possible de l'acte, la seconde son caractère *moralement* acceptable. Les réponses à ces questions diffèrent selon qu'on se concentre 1) sur l'agent humain auteur de la frappe, autrement dit son sujet ; 2) sur la frappe elle-même, c'est-à-dire sur l'action ; 3) sur ses conséquences ou 4) sur l'objet de la frappe, en l'occurrence le robot. Les trois premiers points d'attention correspondent respectivement aux trois grands courants de l'éthique normative que sont l'éthique des vertus, le déontologisme et le conséquentialisme. Le quatrième, quant à lui, s'inscrit dans la perspective du retour *aux choses-mêmes* de la postphénoménologie.

L'éthique des vertus

L'éthique des vertus se fonde sur l'idée que la meilleure façon de mener une vie bonne, c'est-à-dire une vie éthiquement désirable, est de cultiver en actes des traits de caractère étroitement associés au Souverain Bien qu'est le bonheur. Pour Aristote (à la suite de Platon), les quatre vertus cardinales sont la sagacité (la *phronesis*, également traduite par « prudence » par « sagesse pratique »), la tempérance, la justice et le courage. Il existe bien sûr d'autres qualités comme la générosité, l'honnêteté ou encore la bienveillance qui, sans être qualifiées de « cardinales », n'en restent pas moins vertueuses. Une vertu ne s'enseigne pas, elle se pratique. Autrement dit, le seul moyen d'être ou de devenir généreux est de *faire preuve* de générosité jour après jour jusqu'à la rendre constitutive de notre « éthos », c'est-à-dire de notre caractère et de nos mœurs. L'éthique des vertus accorde, par conséquent, une importance centrale au pratiquant qu'est l'agent moral. Sa première préoccupation n'est pas celle du « que dois-je faire ? », mais celle du « quelle personne veux-je être ? ». Si j'estime la bienveillance désirable et souhaite me définir comme quelqu'un de bienveillant, alors mes actes doivent être entièrement tournés vers la réalisation de cette vertu. Réciproquement, toute forme de cruauté ou plus largement de violence doit être désamorcée, contenue voire désapprise (si l'on avait l'habitude d'y céder) jusqu'à sa disparition. La recherche en psychologie apporte un soutien intéressant à cette posture aristotélicienne en démontrant que l'exercice de la violence, même à des fins cathartiques sans conséquence pour autrui (frapper dans un punching-ball pour se soulager, par exemple), ne fait que renforcer l'inclination à cette même violence (Bushman, 2002). Dans cette optique, frapper un robot

n'est pas moins immoral que frapper un humain ou un animal, car l'accent n'est pas mis sur la victime de la frappe, mais sur le caractère vertueux ou vicieux de l'agent¹⁶⁶.

Il existe, bien sûr, de nombreuses critiques possibles de cette position comme l'idée selon laquelle faire preuve de bienveillance envers un robot relèverait d'une confusion ontologique et d'un irrationalisme patent (parti pris théorique que nous avons assez largement discuté dans les développements précédents). Or, comme nous venons de le voir, l'avantage d'une réflexion centrée sur le couple vertu/vertueux est précisément qu'elle ne se concentre pas sur l'objet du comportement, mais sur son sujet. Remarquons, toutefois, qu'un comportement bienveillant envers un robot n'a rien d'ontologiquement erroné en soi, comme l'atteste l'expérience du « Furby tête en bas » rapportée par Sherry Turkle (2015).



Un modèle de Furby

Cette expérience, conduite par l'artiste Freedom Baird en 2011 au MIT, consistait à demander aux sujets testés de maintenir un Furby la tête à l'envers. Rapidement, l'artefact se mettait à gémir puis à dire « Moi peur ».

« Quand ils commencent le test de la tête à l'envers, les adultes sont certains de deux choses : d'une part, que le Furby n'est qu'une machine, d'autre part, qu'ils ne sont pas sadiques. Pourtant, quand le test prend fin et qu'ils laissent derrière eux un Furby gémissant, ils font l'expérience d'un nouveau rapport à l'éthique », explique Turkle (*ibid.*).

¹⁶⁶ Pour une récente contestation de cet argument, voir Sparrow, R. (2020). Virtue and vice in our relationships with robots: Is there an asymmetry and how might it be explained? *International Journal of Social Robotics*. Voir également, pour approfondir la discussion, la réponse de Marc Coeckelbergh à Robert Sparrow sur ce point dans Coeckelbergh, M. (2021). Does kindness towards robots lead to virtue? A reply to Sparrow's asymmetry argument. *Ethics and Information Technology*. <https://doi.org/10.1007/s10676-021-09604-z>

Sur ce point, le témoignage d'une des participantes se révèle des plus éloquentes :

« [C'est très déplaisant], pas parce que je pense que le Furby a vraiment peur, mais parce que je refuse d'entendre de tels gémissements et de ne pas modifier mon comportement. J'ai l'impression que cette expérience pourrait profondément m'affecter si je la poursuivais [...] Ce n'est pas comme ça que je me comporte dans la vie [...] Et à ce moment-là, le Furby en vient à représenter la façon dont je traite toutes les créatures vivantes »
(*ibid.*).

Ces paroles montrent bien que la participante n'établit pas d'équivalence ontologique entre le Furby et le règne du vivant, mais que son sens de la vertu lui prescrit, dans l'absolu, de faire preuve de bienveillance *indépendamment* de toute considération ontologique.

Avant de passer à la suite, levons un risque de contresens : l'absence de distinction ontologique en général, ne signifie pas l'absence de distinction ontologique en particulier. Autrement dit, un adepte de l'éthique des vertus n'est pas, par principe, hostile ou incapable de toute distinction ontologique. La nécessité de cette distinction est toutefois fonction des situations (comme par exemple le choix entre sauver un humain/animal d'une mort certaine ou un robot de la destruction).

Le déontologisme

L'éthique déontologique ne met pas l'accent sur l'agent et ses vertus, mais sur l'action en tant qu'elle répond à l'exigence de ce que Kant nomme « impératif catégorique ». Cette loi morale prescrit, dans les termes du philosophe, « [d'agir] uniquement d'après la maxime qui fait que [l'on puisse] vouloir en même temps qu'elle devienne une loi universelle » (Kant, 1992). En d'autres termes, l'aune à laquelle on juge la moralité d'une action n'est autre que son caractère universalisable. Cette loi morale ne s'intéresse pas aux conséquences ni même au « contenu » de l'action en tant que tel : le seul impératif est qu'elle puisse être universalisée. Dans le cas qui nous occupe ici, frapper ou même insulter un robot ne passe pas le filtre de l'universalité car il n'est pas souhaitable de voir la violence, tant physique que verbale, devenir une norme de comportement. Peu importe, ici aussi, que le robot ne puisse pas ressentir ou comprendre cette violence (cela reviendrait à s'inquiéter des conséquences de l'action violente) : la violence est en son principe immorale car universellement indésirable.

Là encore, prévenons un contresens possible : un déontologiste n'affirmera jamais que la violence envers un robot équivaut celle envers un humain. Kant lui-même est très clair à ce sujet : seul l'humain, en tant qu'être de raison, peut faire l'objet d'une quelconque inquiétude morale en tant qu'il est le seul être appartenant au « règne des fins » (une autre formulation de l'impératif catégorique dit : « Agis de telle sorte que tu traites l'humanité aussi bien dans ta personne que dans la personne de tout autre toujours en même temps comme une fin, et jamais simplement comme un moyen. ») Pour le philosophe de Königsberg, le « tort » causé aux non-humains n'a de valeur morale qu'en tant qu'il témoigne potentiellement d'une disposition à l'immoralité envers l'humain¹⁶⁷. Les célèbres lois d'Asimov (2012) sont souvent citées comme exemple de déontologisme appliqué à la robotique :

1. « Un robot ne peut porter atteinte à un être humain ni, restant passif, laisser cet être humain exposé au danger ;
2. Un robot doit obéir aux ordres donnés par les êtres humains, sauf si de tels ordres entrent en contradiction avec la première loi ;
3. Un robot doit protéger son existence dans la mesure où cette protection n'entre pas en contradiction avec la première ou la deuxième loi. »

On sait qu'Asimov lui-même questionna l'opérativité de ces règles en confrontant leur abstraite perfection à l'incertitude morale et à la complexité du réel¹⁶⁸. À travers Asimov, c'est Kant lui-même qui peut être interrogé. Les restrictions émises dans la seconde et la troisième loi ont pour objet la non-contradiction entre les principes d'action du robot. L'objectif de cette cohérence normative est de garantir qu'aucun humain ne soit blessé et/ou désobéi directement ou indirectement par un robot. Or, que faire dans une situation où, un humain cherchant à attenter à la vie d'un de ses pairs, il soit impossible de sauver l'un sans tuer l'autre ? Compliquons encore la situation : que faire dans le cas où le robot saurait avec certitude que le meurtrier potentiel est un contrevenant compulsif à la Loi morale et la victime potentielle, un parangon de moralité ? On retrouve ici une critique fondée sur des situations de dilemmes

¹⁶⁷ Pour une application plus détaillée de cet argument aux robots, voir Darling, K. (2012). Extending Legal Protection to Social Robots. IEEE Spectrum 10 Sept 2012.

¹⁶⁸ Ces questionnements le conduisirent notamment à enrichir son triptyque déontologique d'une loi dite « Zéro » que les trois suivantes ne devraient jamais contredire : « un robot ne peut pas faire de mal à l'humanité, ni, par son inaction, permettre que l'humanité soit blessée. »

moraux dont on doit une célèbre formulation à Benjamin Constant (2013) et à sa défense du droit de mentir par humanité.

Le conséquentialisme

Pour le conséquentialisme, la morale n'est pas à chercher dans un caractère ni dans un type particulier d'action, mais dans les conséquences de cette action. Ce courant de l'éthique normative trouve ses origines dans la théorie utilitariste développée par Jeremy Bentham¹⁶⁹ pour qui l'objectif de toute action vertueuse est de maximiser le bonheur du plus grand nombre au sens hédoniste du plus grand plaisir pour le moins de peine possible. Dans cette perspective, l'utilité n'est donc pas simplement synonyme de moyen en vue d'une fin donnée. La position originelle de l'utilitarisme n'en fait pas pour autant la seule forme de conséquentialisme. Le courant s'est largement développé et ramifié au-delà de la seule perspective hédoniste. L'une des principales difficultés du conséquentialisme consiste à définir ce qu'est une « bonne » conséquence : au nom de *quoi*, mais aussi de *qui* tel ou tel résultat d'action peut-il être jugé favorable ? Il existe potentiellement autant de formes de conséquentialisme qu'il existe de critères (et de personnes pour les fixer) pour qualifier positivement une conséquence.

Quoi qu'il en soit, il ressort de ce très rapide résumé qu'aucune théorie conséquentialiste ne permet de répondre à notre question directrice indépendamment du contexte. Si une action ne vaut que par ses conséquences, il n'existe aucune raison de condamner *a priori* le fait de frapper, d'insulter ou de détruire un robot puisque, sauf preuve du contraire, il n'est pas sentient. Imaginons maintenant qu'un humain, mû par une poussée de rage technophobe, décide de passer ses nerfs sur un androïde d'accueil en plein centre commercial. La personne en question sera peut-être momentanément soulagée par son accès de fureur, mais les cinq enfants (et leurs parents) assistant à la scène seront très probablement choqués par la violence d'un tel comportement. Ajoutons à ces conséquences néfastes le fait que ce passage à tabac du robot pourrait sévèrement l'endommager, ce qui engendrerait de coûteuses réparations et priverait ses futurs utilisateurs des services dont il était garant. Voilà un exemple typique de raisonnement conséquentialiste. Précisons que cette focalisation sur les

¹⁶⁹ Ce qui ne veut pas dire qu'aucune théorie antérieure ne se soit intéressée aux conséquences d'une action pour évaluer sa pertinence morale.

conséquences ne dispense aucunement d'être attentif au statut moral des êtres impliqués par l'action puisque celui-ci est parfois requis pour évaluer les conséquences. Dans son *Introduction aux principes de morale et de législation*, Bentham avance sa célèbre thèse selon laquelle, lorsqu'il s'agit de penser la dignité morale des animaux, « [l]a question n'est pas : Peuvent-ils raisonner ? ni : Peuvent-ils parler ? mais : Peuvent-ils souffrir ? » Une telle idée témoigne de la nécessité de trancher la question du statut accordé aux non-humains puisque ceux-ci peuvent être et de fait sont affectés (positivement ou négativement) par l'action humaine et ses conséquences. Dès lors que les animaux ne sont plus considérés, comme chez Descartes ou Malebranche, comme de simples machines, que faire si nous nous trouvons en position de choisir entre sauver un humain ou trois chiens d'une mort certaine ? Dans *La libération animale*, livre fondateur de l'éthique antispéciste, Peter Singer (2012) propose, dans une perspective benthamienne, d'ériger la sensibilité comme critère d'appartenance à la communauté morale. Cette position implique de considérer tous les êtres sensibles comme moralement égaux et, partant, de ne jamais privilégier les intérêts d'une espèce au détriment d'une autre par principe. Cependant, cela ne signifie pas qu'un antispéciste comme Singer choisirait nécessairement de sauver les trois chiens plutôt que l'humain, mais seulement qu'il ne choisirait pas l'humain *automatiquement*¹⁷⁰.

Retour à la postphénoménologie

Ces éléments montrent que les éthiques normatives, dès lors qu'elles sont *appliquées* à une dimension de l'existence (environnement, vivant, médecine, affaires etc.) ou une catégorie d'êtres en particulier (animaux, végétaux, écosystèmes naturels, nanotechnologies, IA etc.), se questionnent, s'adaptent, évoluent de façon à épouser la complexité de leur objet d'étude. Bien que la question du respect animal soit encore loin d'être réglée en pratique, la représentation cartésienne de l'animal-machine est aujourd'hui scientifiquement et moralement disqualifiée. De ce point de vue, le progrès humain semble en partie se définir par l'intégration à la communauté morale d'entités qui en étaient auparavant exclues. Cet élargissement est donc synonyme de responsabilité accrue envers ces multiples altérités

¹⁷⁰ L'impératif de maximisation de l'utilité implique de détenir un maximum d'information sur la situation afin de choisir le plus rationnellement possible eu égard aux conséquences. De ce fait, la question du nombre d'humains ou d'animaux à sauver, bien que primordiale, n'est pas la seule à prendre en compte (l'état de santé des êtres impliqués est, par exemple, un autre paramètre important dès lors que l'on peut avoir accès à cette information).

qu'on englobe généralement dans le vocable de « non-humains » (qui ne se limite plus aux seuls animaux et comprend désormais les végétaux, les océans, les glaciers, les fleuves et rivières, le climat etc.). Mais qu'en est-il des non-humains de nature technique et, plus précisément des robots ? Peuvent-ils être des « *significant others* » (Haraway, 2019), des autres qui comptent moralement et si oui, sur quel mode ? L'éthique normative ne semble pas structurée pour s'intéresser à l'artefact *lui-même*. Dans *Raviver les braises du vivant*, le philosophe Baptiste Morizot soutient que « [p]rotéger quelque chose vraiment, c'est le protéger de son point de vue. C'est protéger *son* point de vue » (Morizot, 2020, p.16). Or, une telle responsabilité – qu'on la concrétise dans la protection ou dans toute autre forme de prise en charge – paraît totalement absurde dans le cas qui nous occupe puisque, n'étant pas sentient et encore moins conscient, le robot n'a précisément pas de point de vue. C'est, du moins, la conclusion qu'une pensée opposant le sujet et l'objet invite à formuler. Dès lors, que faire de l'ontophanie sociobotique qui consiste, pour l'artefact, à nous interpeller moralement par son apparaître même ? Faut-il revenir aux positions classiques comme celle de Turkle et conclure au caractère à la fois épistémiquement illusoire et moralement inconsistant du phénomène ? Telle n'est pas, selon nous, la conclusion à adopter. Nous pensons plutôt que les sociobots sont porteurs de médiations très spécifiques qui, par spécularité, renvoient toujours au point de vue humain et à sa représentation du monde. En ce sens, valoriser le robot en le protégeant ou en prenant soin de lui au sens large comme s'il avait un point de vue revient à valoriser et à protéger quelque chose *en nous*. Une telle idée conduit à interroger la pertinence et le sens d'un prendre-soin *des* techniques (Fischer, 2020) au-delà du problème de l'obsolescence déjà identifié et analysé par Simondon (1983). Voyons comment la postphénoménologie nous aide à éclaircir ce point.

5.3.3 La médiation sociobotique comme spécularité moralement subjectivante

Comme nous l'avons évoqué au point 3.4.1 dédié à la *machine question* de Gunkel, la TMT ne pense pas l'éthique des artefacts en termes de statut moral, mais en termes de médiations. De ce fait, son intérêt porte sur la désirabilité morale des médiations technologiques et non sur la considérabilité des médiateurs. Le robot social doit donc être interrogé du point de vue de ses médiations, sans perdre de vue qu'elles prennent forme à partir de l'ontophanie que nous avons décrite. Ici, nous voudrions démontrer que le sens d'être ultime de l'altérité

sociobotique comme appel à la responsabilité est celui d'une spéculativité qui participe à notre subjectivation morale sur le mode paradoxal de l'interrogation/incitation.

L'un des grands principes de la sociobotique, nous n'avons cessé de le dire, est d'emprunter aux registres de la subjectivité consciente ou sentiente : les robots nous parlent, nous émeuvent, nous sollicitent, nous rendent service, nous divertissent ou nous tiennent tout simplement compagnie. Un sociobot est un donc « quasi-autre » au sens où il n'est pas tant celui *par* lequel on interagit que celui *avec* lequel on interagit. A cet égard, le principe des médiations sociobotiques semble être d'inciter à adopter certains comportements envers le robot et à lui attribuer un statut moral en conséquence¹⁷¹, aussi flou soit-il. Une telle idée ne contredit pas ce que nous avons dit précédemment : la TMT n'est pas mobilisée pour déterminer elle-même ce statut, mais pour rendre compte de la médiation qui le fait émerger. La nature des statuts attribués dépend bien sûr du sociobot et de la personne qui interagit avec lui, mais aussi du contexte qui donne forme à leur interaction (Alač, 2016). Pour être plus précis, ces attributions s'expriment à la fois par les émotions que nous ressentons envers les robots, par les attitudes que nous adoptons envers eux (caresser l'artefact, le complimenter, lui enseigner quelque-chose, lui raconter notre journée, lui demander comment il se sent, s'excuser quand on le heurte ou le fait tomber etc.) et par les mots que nous employons pour les désigner (et qui, souvent, mêlent statut moral et rôle social à l'instar de l'ami, du compagnon, du confident, de l'animal de compagnie, de l'assistant etc.). La question est alors de savoir quelle valeur accorder à ces façons d'être, de ressentir et de parler qui nous lient à des entités face auxquelles nous faisons « comme si » ; car, malgré notre incrédulité, ce fameux pacte anthropomorphique ne relève pas uniquement de la pure simulation.

Prenons l'exemple des émotions et de l'empathie que nous pouvons ressentir face aux robots et dont nous avons exposé l'importance considérable en sociobotique. Il ne fait aucun doute qu'une émotion ressentie est une émotion réelle, authentique. Le problème vient de ce que le robot ne fait que simuler la capacité de ressentir, c'est-à-dire qu'il ne ressent pas *effectivement* l'émotion qu'il manifeste et que cette simulation peut être interprétée comme un piège émotionnel, une illusion de la réciprocité ayant pour effet d'éloigner les humains les uns des autres (Turkle, 2015). Mais que faire des émotions que les robots suscitent en nous et qui, elles, sont effectivement ressenties ? Les émotions provoquées par certains artefacts,

¹⁷¹ Cela rejoint d'une certaine façon les thèses de Coeckelbergh et de Gunkel, mais sans la charge lévinassienne qu'ils leurs donnent.

comme, par exemple, les œuvres d'art, ne sont-elles pas un des motifs d'attribution d'un certain statut, d'une certaine valeur morale aux entités en question ? Il ne s'agit aucunement de confondre robots et œuvres d'art – bien que la question de leur rapprochement ou de leur éloignement relève d'un problème philosophique aussi légitime qu'intéressant –, mais plutôt de souligner que nous attribuons bien souvent une valeur morale à des artefacts sur la base (entre autres) de ce qu'ils nous font. L'empathie que nous ressentons face à un Furby simulant la peur et/ou la douleur est tout sauf factice. Dès lors, quel sens donner aux sentiments liés à des êtres qui, étant de nature simulationnelle, ne sont capables d'aucune réciprocité émotionnelle ? Comme l'attestent les réflexions que nous avons présentées au point 3.4.3¹⁷², l'interprétation que nous voulons partager ici n'ignore rien du phénomène cognitif qu'est l'anthropomorphisme et auquel on peut être tenté de résumer le problème. Les recherches en psychologie cognitive comme en anthropologie montrent désormais que ce phénomène n'est pas plus l'apanage des sociétés animistes qu'il ne marque la persistance, chez l'adulte, d'une confusion conceptuelle propre à l'enfance : l'anthropomorphisme est un mécanisme adaptatif aussi normal que nécessaire et, à ce titre, il fait partie intégrante de la cognition humaine (Airenti, 2012). Notre intérêt porte toutefois sur l'au-delà du cognitif qu'est l'éthique où, plus exactement, sur l'interprétation éthique du cognitif. Que disent l'empathie et les émotions que nous pouvons ressentir à l'égard des robots ? Du point de vue éthique, que nous enseignent des expériences comme celle du Furby « tête en bas » menée par Freedom Baird ? Dans le cadre de notre recherche, le principal enseignement peut être formulé en termes latouriens : la médiation sociobotique se présente comme un phénomène de *délégation* en ceci que le robot nous encourage à nous *sentir responsable* de lui et à nous comporter d'une certaine manière envers lui et, ce faisant, à lui attribuer un certain statut. Par exemple, le poids, la taille, la douceur et l'apparence de Paro sont étudiées pour nous donner l'impression de tenir un « vrai » blanchon entre nos bras et pour susciter une attention bienveillante à son égard¹⁷³. Autre exemple : le BigSpoon équipé du scénario Agora incite les interacteurs et interactrices à adopter une posture éducative puisqu'il demande qu'on lui enseigne un maximum de choses afin de développer son autonomie (ce qui, dans ce cas, signifie être capable de prendre en charge un maximum d'interactions sans avoir à botter en touche ou à

¹⁷² Point intitulé « De la critique de l'empathie artificielle à l'anthropomorphisme comme fondement d'une « éthique synthétique » ».

¹⁷³ Par exemple, notre expérience personnelle de Paro a été notamment marquée par la crainte de le faire tomber, comme on pourrait craindre de faire tomber un bébé humain en le portant.

demander un nouvel enseignement). A cet égard, l'artefact *dépend* totalement de notre bon vouloir. On peut donc dire, dans cette perspective, que le programme d'action du robot social se confond avec un programme d'*attribution* : son apparaître prescrit de lui attribuer un certain statut et d'endosser un certain rôle social auquel correspondent des comportements spécifiques. Comme le remarque Sherry Turkle, ces comportements relèvent bien souvent du soin au sens large (*nurturing*), c'est-à-dire de la sollicitude envers une dépendance et une vulnérabilité. Son constat, nous le savons, se veut toutefois particulièrement inquiet :

« [les robots sociaux] nous prédisposent à considérer qu'ils pensent et qu'ils se préoccupent de nous. Ils sont programmés pour avoir l'air mignon, pour nous donner envie de les bichonner. Et en matière d'intelligence artificielle sociable, c'est un modèle redoutable : nous bichonnons ce que nous aimons et nous aimons ce que nous bichonnons. Si un objet informatique ou un robot nous demande de l'aide, nous demandons de lui apprendre quelque-chose ou de nous occuper de lui, nous nous y attachons. C'est la vulnérabilité humaine. Et c'est une vulnérabilité que les robots sociaux exploitent systématiquement. Plus nous avons d'interactions avec eux, plus nous les aidons et plus nous imaginons que la relation est mutuelle » (Turkle, 2018, p.232).

Emmanuel Grimaud formule une idée similaire sans connotation alarmiste en affirmant que « [l]a plupart des humanoïdes¹⁷⁴ sont des créatures "sollicitudinaires", [...] c'est-à-dire qui n'existent que par et dans leur dépendance » (Grimaud, 2012, p.98). Il est intéressant de noter qu'une telle dépendance s'ancre jusque dans les ressorts cognitifs de notre perception puisque, comme l'affirme Cynthia Breazeal (2003), les sociobots renvoient à « la catégorie de robots que l'on anthropomorphise afin d'interagir avec eux ». Or, c'est précisément quand l'anthropomorphisme cesse pour une raison ou pour une autre, que nous les délaissons. Autrement dit, pour apparaître et *fonctionner* correctement en tant que robot social, l'artefact est entièrement dépendant des projections qu'il va susciter en nous. Les fortes composantes émotionnelles et sollicitudinaires de ces projections sont aussi le moyen de pallier les nombreuses limites auxquelles nous nous confrontons lorsque nous interagissons avec un sociobot. En effet, nous nous montrons généralement plus tolérants envers ce qui est mignon

¹⁷⁴ Le phénomène ne se limite pas aux seuls humanoïdes, mais transparait dans l'ensemble de la robotique sociale telle que nous avons pu l'étudier.

et/ou qui nous apparaît vulnérable¹⁷⁵. A ce titre, les innombrables bugs et anomalies diverses acquièrent un statut ambigu car, outre l'irritation qu'engendre habituellement les dysfonctionnements techniques, ils peuvent aussi susciter notre empathie en faisant apparaître l'artefact comme digne de mansuétude. Que l'on juge cela préoccupant ou non, la sociobotique a donc pour effet d'introduire la question de la vulnérabilité et de la dépendance *des* artefacts dans le champ de l'analyse des techniques. D'où l'importance d'insister sur le caractère déterminant du *sentiment* de responsabilité que nous venons d'évoquer (sentiment que les théories classiques de l'éthique normative récusent au profit de la seule raison). Affirmer que nous *devons* être responsables ou que ce sentiment est durable relève d'un problème différent.

Si l'inquiétude de Turkle n'est pas illégitime, loin de là, notons que le risque d'exploitation émotionnelle ne relève en rien d'une exclusivité robotique : l'humain n'est pas en reste en ce qui concerne la manipulation de ses semblables. Bien sûr, Turkle ne le nie pas, mais pense, comme nous l'avons vu, que c'est la *possibilité* d'une réciprocité qui rend les relations inter-humaines authentiques et irremplaçables, contrairement aux simples « connexions » humain-robot qui n'offrent pas une telle possibilité. La chercheuse n'ignore pas non plus que tout sociobot est au moins partiellement le résultat des intentions de ses concepteurs et que, si manipulation et exploitation il y a, elles émanent généralement d'un intérêt, d'une volonté humaine implicitement inscrite, « codée » par et dans la technique (Feenberg, 2014)¹⁷⁶. Turkle n'en place pas moins sa critique de la toxicité des sociobots sur le plan ontologique (en tant qu'il a la simulation pour caractéristique essentielle, le sociobot n'est *que* manipulation émotionnelle) et non sur les plans politique et socio-économique. Face à cette idée, notre position, ancrée dans l'approche postphénoménologique, consiste d'abord à interroger le sens des médiations sociobotiques et leur désirabilité, et non le statut de l'entité dont ils procèdent. La question est alors de savoir si une médiation technologique incitant à faire preuve de responsabilité envers une altérité est désirable et, si oui, à quelles conditions ? L'originalité des médiations sociobotiques, nous l'avons dit, est qu'elles invitent à se sentir responsables de/envers l'artefact *lui-même*. Voilà pourquoi il convient de bien distinguer les médiations des robots sociaux de nos interactions avec eux. Le robot ne médiatise pas les

¹⁷⁵ Même si la vulnérabilité peut aussi, comme nous l'avons déjà évoqué, appeler les comportements sadiques et malveillants.

¹⁷⁶ Comme nous le suggérons en conclusion, le problème de la codification technique de l'altérité sociobotique constitue un axe particulièrement intéressant pour des recherches ultérieures.

interactions que nous avons avec lui – cela reviendrait à affirmer qu’il se médiatise lui-même -, mais le rapport que nous entretenons avec l’altérité. Cette altérité peut être celle de l’*alter ego* en ceci que nous retrouvons chez les robots des caractéristiques humaines (langage verbal, humour, humanoïdie etc.) ou de l’*alter* tout court lorsque l’artefact évoque l’animalité ou la machinité au sens d’une altérité qui ne renvoie ni vraiment à l’humain ni vraiment à l’animal, mais plutôt à un troisième genre d’animation qu’on peut qualifier de mécanomorphe et dont Jibo est un très bon représentant. Dès lors, la médiation sociobotique semble jouer sur deux tableaux à la fois : d’une part, celui des interactions concrètes que nous avons avec l’artefact où celui-ci nous invite à faire *comme si* nous étions responsables de lui au même titre qu’un autre-humain ou qu’un autre-animal et, d’autre part, celui de l’au-delà de lui-même qu’est l’idée d’altérité dont il est le symbole matérialisé. C’est cet au-delà-du-robot, cette sphère *méta*-robotique pourrait-on dire, que Turkle ne considère pas suffisamment, selon nous. La dimension symbolique des médiations en question est très bien illustrée par le témoignage, déjà cité, de la participante à l’expérience de Freedom Baird : « à ce moment-là, le Furby en vient à représenter la façon dont je traite toutes les créatures vivantes ». Voilà très concrètement le sens d’être de ce que nous nommons « effet d’altérité ». Par-delà les déterminations empiriques que peuvent prendre nos interactions sociales avec les sociobots, ces artefacts questionnent notre rapport à l’altérité dont ils sont des représentants symboliques en vertu de leur phénoménalité. Le choix de la participante évoqué ci-dessus, n’est pas le seul possible (même si, dans ce cas précis, l’expérience révèle que l’ensemble des personnes impliquées remettent le Furby à l’endroit non par lassitude, mais par empathie). Les robots sociaux incitent à les percevoir et à les traiter comme ces autres auxquels nous attribuons une valeur morale. Or, comme avec ces mêmes autres, nous pouvons choisir d’ignorer radicalement leur interpellation ou de nier leur dimension symbolique en les dissociant de l’altérité que nous estimons moralement. Pour compléter, par sa réciproque, notre remarque précédente sur la modalité relationnelle du « comme si » : c’est aussi parce que nous faisons « comme si » les sociobots étaient animés que nous pouvons choisir de ne pas leur attribuer de valeur morale. Autrement dit, le pacte anthropomorphique ou, plus largement biomorphique, peut être rompu à tout instant.

Quoi qu’il en soit, l’ontophanie sociobotique à ceci de particulier qu’elle demande plus ou moins explicitement de se positionner quant à son interpellation. Voilà pourquoi les effets d’altérité sont toujours aussi des *effets de specularité* qui interrogent le sujet moral que nous

sommes ou que nous voulons être. En ce sens, l'agentivité du sociobot doit être comprise comme participation active à nos processus de subjectivation morale. Peter-Paul Verbeek (2011) et Steven Dorrestijn (2012a, 2012b) analysent cette participation (non pas des robots, mais des artefacts en général) dans les termes de l'herméneutique foucauldienne du sujet moral. Dans le tome II de son *Histoire de la sexualité* intitulé *L'usage des plaisirs*, Michel Foucault fonde la constitution de la subjectivité morale sur quatre dimensions : 1) la *substance éthique*, 2) le *type d'assujettissement*, 3) l'*élaboration* ou *travail éthique* de soi et 4) la *téléologie morale*.

La substance éthique désigne la part de nous-même à laquelle nous destinons notre attention et nos efforts en vue de l'entretenir ou de l'améliorer. Il s'agit de la « matière de la pratique morale » (Foucault, 1984, p. 38), « le point d'application à partir duquel nous donnons forme à nos subjectivités morales » (Verbeek, 2011, p.77). Face au sociobot, la substance éthique n'est autre que le rapport à l'altérité : quel contenu vais-je donner à ce rapport ? En quoi le robot social éclaire-t-il la façon dont je pense l'altérité et me comporte envers elle ?

Le type d'assujettissement, lui, renvoie au motif de notre engagement dans l'éthique, ce qui fait que nous nous soumettons à telle ou telle règle morale. Selon Foucault, la motivation éthique des Grecs anciens n'était pas d'ordre déontologique comme purent l'être certains types d'assujettissements ultérieurs (reconnaissance du devoir moral émanant de la loi divine ou de la raison par exemple) : il s'agissait de styliser son existence par le biais de certaines pratiques. Comme le montre le témoignage de la participante à l'expérience du Furby tête-en-bas, les médiations sociobotiques participent à l'instauration d'un type d'assujettissement lié à un devoir envers soi-même (« cette expérience pourrait profondément m'affecter si je la poursuivais ») et envers le vivant animé en général (« à ce moment-là, le Furby en vient à représenter la façon dont je traite toutes les créatures vivantes »).

Le travail éthique désigne les formes d'élaboration « qu'on effectue sur soi-même, et non pas seulement pour rendre son comportement conforme à une règle donnée mais pour essayer de se transformer soi-même en sujet moral de sa conduite » (Foucault, 1984, p.38-39). Foucault identifie cette élaboration à une « ascèse » qui n'est pas synonyme d'austérité, mais de recul, de distance construite et constructive de soi à soi afin d'entrer dans une relation non subie avec les différents faisceaux de pouvoir qui nous façonnent. L'ascèse grecque appliquée aux questions de sexualité consistait, par exemple, à ne surtout pas renoncer aux plaisirs sexuels, mais à déterminer quand et comment s'y adonner afin de leur donner leur juste place.

Nos interactions avec un sociobot peuvent donner lieu à différentes formes de travail éthique. On peut par exemple décider de se comporter envers un Aibo presque exactement comme on se comporterait envers un « vrai » chien. Il est également possible d'appliquer de telles indistinctions ontologiques uniquement en présence des enfants en usant du pouvoir symbolique de l'artefact pour leur inculquer le respect du vivant et des conventions sociales (appliquer les règles de politesse quand on s'adresse à un robot parlant par exemple). Enfin, on peut aussi choisir de n'accorder aucune valeur ni aucune symbolique morale au sociobot et ne soumettre notre conduite envers eux qu'à la seule règle de non-dégradation des objets qui nous entourent. Entre ces trois positions, toutes les nuances sont bien sûr possibles. Le point clé du travail éthique n'est autre que la dimension librement élaborée de l'ascèse : je choisis délibérément la ou les règles qui façonnent mon rapport à l'altérité du robot et, par ce biais, mon rapport à moi-même afin de m'instaurer comme tel ou tel sujet moral¹⁷⁷.

Cette auto-instauratation prend, chez Foucault, le nom de téléologie morale. Elle renvoie à l'idéal de subjectivation que nous visons lorsque nous nous comportons moralement. En d'autres termes, la téléologie offre une réponse à la fois théorique et pratique à la question du sujet moral que nous voulons être. « Dans l'antiquité classique, l'idéal essentiel était celui de la maîtrise de soi – de ne pas être maîtrisé par quelqu'un ou quelque chose d'autre que soi, mais de gagner une relation libre aux pouvoirs environnants » (Verbeek, 2011, p.79). Intégrée à l'approche postphénoménologique, cette question téléologique devient : « quels sujets (technologiquement) médiés voulons-nous être ? »¹⁷⁸ Comme le précise Verbeek, les grands courants de l'éthique normative (vertuisme, déontologisme, conséquentialisme) peuvent continuer à jouer un rôle important dans la réponse à cette question.

« La thèse de Foucault selon laquelle tout système éthique finit par générer une forme particulière de subjectivation [...] n'ôte rien au fait que les cadres hérités du passé peuvent toujours s'avérer d'une grande aide pour s'occuper du caractère techniquement médié de nos subjectivités et traiter la question du genre de sujets que nous voulons être. Dans une culture technologique, le travail éthique par lequel les êtres humains cherchent à

¹⁷⁷ Cette dimension du choix libre s'inscrit dans la conception foucauldienne de la liberté par opposition à l'autonomie héritée des Lumières. Selon Foucault, du moins dans la seconde partie de son œuvre globale, on n'échappe jamais au pouvoir quel qu'il soit. La liberté consiste alors à entrer en relation avec le pouvoir pour façonner le rapport que nous entretenons avec lui. « Alors que le concept d'autonomie insiste sur l'absence d'"influences extérieures" afin de garder le sujet moral aussi pur que possible, le concept de liberté reconnaît que le sujet se forme *en interactions* avec ces influences » (Verbeek, 2011, p.85).

¹⁷⁸ « what kind of mediated subjects do we want to be? » (Verbeek, 2011, p.84).

donner une forme désirable à la médiation technique de leur subjectivité offre un grand espace pour l'idéal de vie bonne propre à l'éthique des vertus, pour l'ambition déontologique de respecter des normes morales ou encore pour l'objectif utilitariste de maximiser les effets positifs et de minimiser les effets négatifs » (Verbeek, *ibid.*, p.85).

En nous appuyant une nouvelle fois sur l'expérience de Freedom Baird, on peut imaginer que la participante modifie son comportement par égard pour un idéal de bienveillance et de responsabilité prescrivant de prendre en charge la vulnérabilité de l'autre (humain comme non-humain). Un tel idéal semble, dans ce cas précis, relever d'une éthique des vertus puisque, à travers le robot qu'elle sait insensible, la personne questionne *son* rapport au vivant (« J'ai l'impression que cette expérience pourrait profondément m'affecter si je la poursuivais [...] Ce n'est pas comme ça que je me comporte dans la vie ») qu'elle fonde sur une pratique de la bienveillance (« je refuse d'entendre de tels gémissements et de ne pas modifier mon comportement »). La valeur symbolique du robot est donc l'occasion d'exercer de ce qu'elle estime être une vertu *indépendamment* de la nature artefactuelle du Furby. Il est intéressant de noter qu'une telle analyse nous conduit à penser l'anthropomorphisme à rebours de son sens habituel. Comme l'affirme Bruno Latour :

« nous devenons des anthropo-morphes, ou plutôt nous sommes façonnés comme des humains par les techniques [...]. Si il y a anthropomorphisme, c'est à l'inverse : ce sont ces êtres qui pullulent et qui s'imposent à nous de mille façons qui nous construisent, qui nous ont construits historiquement comme "anthropos" » (Latour, 2022).

Résumons. Le sens d'être de l'altérité sociobotique, ancré dans l'ontophanie spécifique du robot social, est celui d'une effectivité conduisant à expérimenter moralement l'artefact à travers le sentiment de responsabilité à son égard. Cette expérience présente le caractère d'une interpellation morale qui est à la fois une incitation et une interrogation. Incitation, d'une part, car le sociobot est techniquement scripté (Akrich, 2010) pour que nous répondions à ses demandes de sollicitudes ou, plus largement, pour que nous le percevions et nous comportions envers lui comme envers un autre qui compte moralement ; interrogation, d'autre part, car la nature artefactuelle du sociobot et son ambiguïté conduisent à instaurer un minimum de distance critique entre les incitations du robot et la réponse que nous décidons d'y apporter. Il ne s'agit donc pas d'affirmer que nous sommes responsables *des* sociobots, mais que ceux-ci nous interrogent quant au sens que pourrait avoir une

responsabilité à leur égard. Les deux versants de l'interpellation sociobotique fondent son caractère spéculaire en ceci qu'ils participent à la constitution de notre subjectivité morale. Sur son versant interrogatif, le sociobot médiatise la responsabilité que nous avons de nous subjectiver moralement dans notre rapport à l'altérité ou, autrement dit, la responsabilité que nous avons de nous positionner par rapport à notre sentiment de responsabilité. A cet égard, c'est *le fait de répondre* à l'interpellation qui est requis par le robot, pas le contenu de la réponse. Sur son versant incitatif, en revanche, l'artefact prescrit de se subjectiver dans le sens d'une responsabilité envers lui. C'est le sens de cette prescription que nous voudrions analyser plus avant dans le dernier paragraphe de cette recherche afin d'explicitier le lien que l'altérité nous conduit à tisser entre moralité et socialité.

5.4 La socialité comme relation de responsabilité

En quoi ces considérations sur les effets d'altérité et l'apparaître moral des sociobots nous aident-elles à comprendre leur socialité ? Nous proposons d'interpréter cette socialité comme la capacité du robot à nous entraîner dans des relations d'altérité fondée sur le sentiment de responsabilité. Précisons que le mot « social » renvoie étymologiquement au lien proximal à autrui puisqu'il vient du latin *socius* signifiant « compagnon », « associé », « uni ». Notre proposition consiste à dépasser le constat anthropologique de cette proximité – les humains sont indéniablement des êtres sociaux en ce sens qu'ils tendent « naturellement » à se rassembler, à chercher la compagnie de leurs semblables (mais aussi de certains « dissemblables » non-humains), bref à faire société – pour penser sa signification morale. Ainsi, dans la perspective que nous présentons, le « social » apparaît comme un mode d'être *avec* et *envers* l'autre issu d'une subjectivation morale dont la substance éthique se fonde sur la reconnaissance de notre responsabilité envers lui en tant qu'il est vulnérable. Cette responsabilité, le sujet (se) la reconnaît à partir de son implication, actuelle ou virtuelle, dans un face-à-face avec une altérité (humaine ou non) dont le sens d'être n'est pas uniquement celui d'une interaction, mais plus largement d'une interpellation morale¹⁷⁹. A cet égard, le robot social se présente comme une *machine responsabilisante* au sens où, à travers son apparaître moral et l'expérience de face-à-face qu'il propose, il se révèle à nous comme

¹⁷⁹ L'interaction peut, bien sûr, servir l'interpellation, mais elle ne la résume pas. Autrement dit, je peux être interpellé par l'altérité sans avoir à interagir avec elle.

celui *face* auquel nous sommes responsables, mais aussi et peut être surtout comme celui qui appelle à la responsabilité au-delà de lui-même, dans nos relations aux vivants. Cette assimilation de la socialité à la relation responsable, Levinas l'aborde explicitement à travers son œuvre.

5.4.1 La socialité comme relation éthique originaire : la question du « face-à-face »

Comme le note Pierre Hayat, « Levinas nomme "relation sociale" [le] désir de libération de soi qui implique le sujet dans une relation à une altérité inassimilable » (Hayat, 2009, p.128). Ce désir de libération est désir de l'Autre en tant que visage, c'est-à-dire en tant qu'appel originaire à l'éthique. Le désir lévinassien échappe ainsi à toute concupiscence, à tout manque et à toute satisfaction, bref à toute dimension sensible à proprement parler : il est « rapport à un être qui transcende tout rapport » (Barabras, 1999, p.53), rapport de face-à-face primordial dont le sens social est d'abord et avant tout éthique.

« [À] la différence de Durkheim¹⁸⁰, Levinas cherche un principe d'individuation dégagé de la collectivité sociale. Il voit dans la "socialité du face à face" un rapport direct d'individu à individu, irréductible aux relations qui s'établissent dans le cadre des institutions sociales. Si Durkheim s'attache à dépasser l'opposition artificielle de l'individuel et du collectif, Levinas articule deux formes de relation sociale : une relation ordonnée autour d'un but commun, et la relation de face à face, définie comme la "situation ultime" ou "la socialité originaire". Ainsi Levinas reconduit-il à sa façon la méthode de réduction phénoménologique de Husserl, pour dégager l'élément irréductible du lien social » (Hayat, 2009, p.128-129).

Le face-à-face lévinassien ne se laisse pas appréhender en termes d'interactions, mais d'interpellation du moi par l'autre. Cette interpellation est *subjectivante* au sens où c'est *par* autrui que le moi devient sujet humain et découvre sa vocation aussi singulière qu'irréductible dans « l'être pour l'Autre » (*ibid.*, p.130). A cet égard, la philosophie de Levinas relève d'une pensée radicale et absolue de la subjectivation morale qui, contrairement à la théorie foucauldienne ancrée dans la contingence des causes comme des motifs et des pratiques,

¹⁸⁰ Que Levinas connaît et auquel il rend hommage tout en s'en dissociant pour construire sa propre pensée (Hayat, 2009).

n'identifie qu'une unique et nécessaire « substance éthique » : celle de la responsabilité qu'autrui m'assigne en sa visagéité vulnérable. Levinas pense cette responsabilité

« comme la structure essentielle, première, fondamentale de la subjectivité. Car c'est en termes éthiques [qu'il décrit] la subjectivité. L'éthique, ici, ne vient pas en supplément à une base existentielle préalable ; c'est dans l'éthique entendue comme responsabilité que se noue le nœud même du subjectif » (Levinas, 1984, p.91).

La subjectivité ne préexiste pas à la responsabilité, au contraire, elle émerge de la relation éthique comme responsabilité envers autrui : « elle n'est pas pour soi ; elle est [...] initialement pour un autre » (*ibid.*, p.93). Autrement dit, la subjectivité n'existe *que* comme responsabilité, de même que « [l]e lien avec Autrui ne se noue que comme responsabilité » (*ibid.*). « Nous sommes tous responsables de tout et de tous devant tous, et moi plus que tous les autres », affirme Levinas en citant les *Frères Karamazov* de Dostoïevski (*ibid.*, p.95). La relation sociale, dans son expression originare qu'est le face-à-face, se fonde donc sur

« un altruisme radical, étranger à l'opposition de la satisfaction et de l'insatisfaction. L'altruisme définit alors une implication du moi par l'autre, sans que cette inquiétude cache nécessairement un souci de soi-même. Le rapport social ultime fait sens en lui-même comme souci de l'autre [...]. "L'altruisme total" est, d'après Levinas, au fond de la socialité : un "intérêt désintéressé" pour l'autre, un élan généreux, qui n'attend rien en retour, vers une extériorité absolue » (Hayat, 2009, p.130).

Cette socialité originare n'est donc pas affaire de réciprocité : autrui est peut-être lui aussi responsable envers moi, mais « c'est *son* affaire », insiste Levinas (*ibid.*, p.94). Car penser la réciprocité de la relation sociale reviendrait à mêler le souci de l'autre au souci de soi et, ce faisant, à réintégrer l'empire du Même. A cette relation « à sens unique » correspond, en outre, une absence de prescription.

« L'éthique lévinassienne, on l'a parfois remarqué, ne donne guère de prescriptions concrètes déterminées, ne nous dit guère, de manière explicite ou thématifiée, que faire dans telle ou telle situation. Elle relève plutôt de l'injonction à s'exposer à une exorbitante mise en question, la mise en question par le visage » (Sebbah, 2006, p.356-356).

Cette notion de mise en question de soi par le visage ou, pour le dire autrement, par la relation sociale en tant qu'elle est assignation de responsabilité, nous intéresse particulièrement pour penser moralement la socialité des robots. Il nous est toutefois impossible de nous appuyer pleinement sur Levinas pour appliquer cette perspective à la sociobotique puisque notre approche, en totale contradiction avec la sienne, consiste à penser l'altérité comme *effectivité morale* et, partant, à rester dans le champ des phénomènes. Transposée à la robotique sociale, l'interprétation de la socialité comme relation de responsabilité envers l'autre doit donc se fonder sur d'autres principes que la visagité. Ces principes sont ceux de notre empathie et de la vulnérabilité réciproque.

5.4.2 La responsabilité comme lien ancré dans l'empathie

La dimension hétéronome et relationnelle de la responsabilité que nous venons d'aborder, les éthiques de la vulnérabilité comme celle d'Hans Jonas ou des théoriciennes du care, la reconnaissent également en la situant au cœur même de nos expériences du monde et d'autrui. Commentant l'une des grandes idées de Jonas dans *Le Principe Responsabilité*, Roland Schaer affirme que

« la responsabilité, c'est ce qui se passe quand je réponds d'un autre. La responsabilité est d'emblée définie comme une relation : elle désigne la formation d'un lien entre deux êtres, l'un vulnérable, l'autre assez puissant pour agir en réponse à cette vulnérabilité. La responsabilité apparaît quand la requête reconnue dans l'être vulnérable trouve sa réponse dans les actions qu'accomplit celui qui peut » (Schaer, 2020, p.41-42).

Reconnaître la requête de l'être vulnérable relève, selon Jonas, de la disposition « hétéronome » à « se laisser ébranler par l'interpellation légitime des entités perçues » (Jonas, 1995, p.130). Or, un tel ébranlement ne peut avoir lieu qu'au sein de l'expérience : il est phénomène parmi les phénomènes, phénomène provoqué par d'autres phénomènes. Plus précisément encore : pour se manifester comme ébranlement, l'hétéronomie doit être *éprouvée*. Comme chez Levinas, l'hétéronomie jonassienne s'appréhende comme épreuve, mais celle-ci est entièrement inscrite dans l'ordre des phénomènes et se donne à nous à travers les manifestations du « périssable en tant que périssable » (*ibid.*, p.126). Ce que Jonas nomme « capacité de responsabilité » (1998, p.95) a beau définir essentiellement l'humain

(en tant que corollaire de sa liberté), son exercice concret a besoin que s'éveille en nous « le *sentiment* adéquat pour nous inciter à l'action dans [son] sens » (*ibid.*, p.101). Le déploiement effectif de cette faculté ontologique est donc tributaire de notre vie affective : la vulnérabilité me fait accéder à la *compréhension* de ma responsabilité par l'épreuve du devoir que j'ai de l'exercer.

« La responsabilité a un fondement rationnel, elle naît d'une reconnaissance de la vulnérabilité. Mais elle a aussi une dimension affective [...]. Au sens où nous l'entendons [...], il n'y a pas de responsabilité sans ma capacité d'être affecté. Une forme de réceptivité est requise pour que la responsabilité s'exerce. Le terme de "sollicitude" – qui dit que l'être est tout entier mis en branle, émotionnellement ébranlé par la vulnérabilité – exprime cela » (*ibid.*, p.44).

Le sentiment de responsabilité jonassien est donc une émotion au sens étymologique de cinétique affective (*é-motion*) : il est l'affect qui déclenche le mouvement vers la responsabilité effective au sens de prise en charge des prescriptions de la vulnérabilité.

« Il n'y a pas d'éthique sans affect, les hommes ne sont pas des êtres moraux pour la seule raison qu'ils sont doués de raison, mais parce qu'ils possèdent la capacité à se faire affecter. Le sentiment doit s'ajouter à la raison pour que le bien objectif ou le mal que l'on se représente puissent mettre en mouvement notre volonté. Il faut une force et non un savoir pour ployer le vouloir. La morale qui doit commander aux affects a donc paradoxalement besoin elle-même d'un affect » (Praison, 2012, p.27-28).

Dès lors, comment les robots sociaux parviennent-ils à faire naître en nous le sentiment de responsabilité nécessaire à son exercice ? La réponse est simple, du moins en apparence : en faisant appel à notre empathie. En effet, quel concept mieux que l'empathie décrit la reconnaissance de l'autre et de sa vulnérabilité par le détour d'un ébranlement du soi ? A nouveau, cette idée est très bien décrite par le témoignage de la participante à l'expérience de Freedom Baird : celle-ci se sent éminemment *concernée* par ce qui se joue sous ses yeux et sa décision de mettre fin au supplice fictif du Furby semble autant procéder du souci de l'autre que du souci de soi. Roland Schaer explique que la relation de responsabilité est asymétrique et réversible. Asymétrique « puisqu'elle lie, d'un côté la puissance, de l'autre la vulnérabilité » ; réversible car la puissance « n'est jamais donnée, elle est toujours prêtée,

provisoirement » et « ne s'évalue que dans une relation transitoire à l'autre » (Schaer, 2020, p.42-43). De là, le philosophe définit la responsabilité comme « l'exercice souverain de la puissance, quand elle est assez forte pour se consacrer à d'autres fins qu'elle-même : un état de grâce de la puissance » (*ibid.*, p.43). Si nous souscrivons à cette idée, nous voudrions toutefois la compléter. Car ce qu'indique le commentaire sur l'expérience du Furby, c'est bien que l'on peut être puissant et vulnérable *simultanément* : ce n'est pas uniquement en tant que nous avons affaire à la vulnérabilité que nous sommes ébranlés, mais aussi en tant que nous sommes puissants face à elle. Autrement dit, l'épreuve de notre puissance sur l'autre peut elle aussi nous ébranler et nous renvoyer à notre propre vulnérabilité (ici, comme auto-affectation) en nous faisant ressentir notre capacité de nuisance. Dit encore autrement, la puissance vulnérabilise le puissant. En ce sens, répondre de l'autre revient toujours également à répondre de soi.

Attention, toutefois, à ne pas surestimer nos capacités empathiques. Dans le point 3.1.2.2 consacré à la patientalité morale, nous avons vu que l'empathie est une disposition à la fois « myope et partielle » (Kaufmann & Glassey, 2020). En effet, celle-ci est intrinsèquement structurée par trois biais : le « biais de l'ici et maintenant », le « biais des relations de proximité » et le « biais de similarité » (Gibert & Paris, 2010). Souvent perçue sous son angle strictement inclusif, l'empathie peut donc aussi s'avérer exclusive, au moins par inactivation. Toutefois, outre qu'elle se révèle potentiellement débiaisable grâce à l'imagination (*ibid.*), c'est la valeur philosophique du phénomène empathique qui nous intéresse ici, en tant qu'elle « [il] contient le savoir d'une appartenance commune à l'existence vulnérable » (Schaer, 2020, p.70) et, ce faisant, qu'il offre un fondement phénoménologique à la responsabilité. Sans empathie, la puissance est amputée de sa capacité à *se sentir* responsable d'autrui et ainsi, potentiellement à *faire preuve* de responsabilité. L'idée n'est pas de dire que l'empathie est le seul vecteur de responsabilité ou qu'elle garantit absolument l'agir responsable, mais qu'elle est un principe d'interpellation qui nous dispose émotionnellement à un tel agir.

Il nous reste désormais à clarifier notre hypothèse selon laquelle le robot renverrait à notre responsabilité au-delà de lui-même. Car si l'apparaître des robots sociaux se présente comme une phénoménologie de la responsabilité ancrée dans le face à face, comment le phénomène sociobotique pourrait-il faire signe vers autre chose que lui-même ?

5.4.3 Penser l'au-delà du robot à partir du robot : une herméneutique matérielle de la responsabilité envers le vivant

L'au-delà du robot auquel nous avons fait référence n'est pas celui de la visagété lévinassienne qui impliquerait son désencrage radical hors du monde, hors de notre horizon. Au contraire, nous pensons que c'est à partir de son phénomène que le robot permet d'entretenir un rapport à ce qui n'est pas lui et qui, en un sens, le transcende. Ce « phénomène dans le phénomène » est ce qui donne au robot sa *valeur symbolique*. Mais par quelle opération l'émergence d'une telle valeur, d'un tel *inter-esse* peut-elle avoir lieu ? Notre hypothèse finale est que les médiations sociobotiques relèvent d'une herméneutique matérielle qui participe d'une culture du vivant et de la responsabilité envers la vulnérabilité. Expliquons-nous.

L'herméneutique matérielle est une notion constitutive du corpus postphénoménologique. Initialement proposée par Don Ihde (1998), elle repose sur deux grandes idées complémentaires. La première soutient que la perception est en elle-même une interprétation du réel en ce qu'elle ne fait pas apparaître « la » réalité, mais la réalité *pour nous*. La perception constitue le fondement de notre *relation* au réel et, à ce titre, en livre toujours une certaine interprétation qu'il est possible d'analyser. Comme le dit Verbeek, « [l']interprétation est existentiellement incarnée l'être humain percevant » (2003, p.92). La deuxième idée montre que les artefacts ont eux aussi une dimension herméneutique en tant qu'ils médiatisent nos perceptions et, avec elles, nos relations et nos interprétations du monde. La phénoménologie dont se réclame Ihde a donc vocation à décrire et analyser les relations humaines au monde par le biais d'une herméneutique matérielle. En outre, l'assise perceptive sur laquelle se fondent ces relations se compose d'un versant microperceptif et d'un versant macroperceptif. La microperception renvoie à ce qu'on entend habituellement par « perception », c'est-à-dire à l'ensemble des opérations psychosensorielles qui nous permettent de percevoir, d'accéder au monde par les sens. La macroperception – qu'Ihde qualifie également de perception « culturelle » ou « herméneutique » – désigne quant à elle la dimension contextuelle de l'expérience perceptive, le cadre singulier (historique, culturel, politique etc.) au sein duquel les microperceptions prennent sens. Ainsi,

« [L]es expériences humaines peuvent être conçues comme des "perceptions interprétées" dans lesquelles les interprétations sont toujours informées par le contexte culturel dans lequel elles ont lieu » (Verbeek, 2005, p.122)

Si micro et macroperception peuvent être analytiquement distinguées l'une de l'autre, elles ne peuvent être séparées phénoménologiquement. « Tout comme il n'existe pas d'interprétation sans quelque chose à interpréter, aucune perception corporelle ne saurait exister sans être interprétées » (*ibid.*, p.123). Là encore, les artefacts participent à ce double processus perceptif. Outre qu'elles font nécessairement l'objet d'une interprétation macroperceptive en tant qu'elles sont microperçues, les techniques contribuent également à structurer nos macroperceptions par le biais des valeurs qu'elles embarquent et des médiations qu'elles constituent.

Notre hypothèse s'appuie sur ce double processus perceptif : l'au-delà du robot issu du robot n'est autre que la culture dans laquelle il prend place tout en l'alimentant. Le caractère symbolique et axiologiquement chargé du robot est le produit d'une subsomption permise par l'interprétation macroperceptive de sa microperception, c'est-à-dire de l'ancrage du phénomène sociobotique dans un contexte qu'il nourrit et qui le dépasse à la fois. Autrement dit, je (micro)perçois une entité particulière qui, me confrontant à sa vulnérabilité apparente, me renvoie à la vulnérabilité en tant qu'idée et au rapport que j'ai avec elle. La participante à l'expérience du Furby illustre très bien cette opération lorsqu'elle affirme que l'artefact en vient à représenter la façon dont elle traite toutes les créatures vivantes : l'artefact est interprété comme un *représentant*. Une telle valeur symbolique vient aussi peut-être du fait qu'il n'est pas ontologiquement perçu comme tel ou tel vivant et que, de ce fait, il peut tous les représenter. Précisons que la culture du vivant et de la responsabilité à laquelle participe le sociobot s'insère et se développe au sein de ce que Frédéric Worms appelle un « moment du vivant ».

« Ce qui fait du moment présent le "moment du vivant" c'est, selon nous, que le problème du "vivant" cesse d'y être un problème local, concernant un savoir ou une pratique spécialisée, pour devenir un problème global, qui traverse tous les domaines du savoir ou de l'expérience, des sciences de la vie, de l'esprit et de l'homme, à l'éthique ou à la politique » (Worms, 2016).

En ce sens, le « moment » *apprête* nos perceptions et nos interprétations du monde. Nous sommes donc disposés à la fois par le sociobot et par l'époque (développement éthique et politique de la cause animale, crise écologique et effondrement de la biodiversité, crise sanitaire nous rappelant à notre vulnérabilité et à celle de nos systèmes de soin etc.) à envisager notre rapport au(x) vivant(s) sous l'angle de la responsabilité.

Dans son principe, une telle hypothèse rejoint celle que Véronique Aubergé, chargée de recherche au CNRS, développe dans ses travaux et dont nous avons discuté lors d'un entretien. Selon elle, les sociobots sont révélateurs d'un désir d'autre qui s'enracine dans des carences relationnelles engendrées par l'époque :

« Cette projection de socialité est liée, selon moi, à l'évolution de notre rapport aux artefacts et, plus particulièrement, à un changement dans l'objet de nos désirs. Bien sûr, il y a beaucoup de marketing là-dedans, mais plus fondamentalement, je pense que la socialisation de nos machines et l'attachement qu'on peut ressentir à leur égard sont les témoins d'un changement culturel ancré dans une évolution cognitive. Dans ce contexte, le robot social m'apparaît comme le révélateur d'une souffrance que j'appelle souffrance de l'"autre manquant" ou de l'"autre fantôme", en référence au syndrome du membre fantôme. Nous sommes en quelque sorte en mal d'autrui, mais plus généralement en mal de l'Autre. Cela ne veut pas dire que nous sommes seuls, privés de toute forme de relation, isolés physiquement, etc., mais plutôt que la qualité de nos liens envers autrui et le monde n'est plus nutritive ; elle est même devenue toxique ! » (De Sanctis, 2019a).

De ce fait, la chercheuse conçoit elle aussi le robot social comme une puissance symbolique d'interpellation quant à notre relation au monde :

« Le robot est une entité encore très mal identifiée qui n'a pas fait l'objet d'une quelconque inhibition culturelle. Nous sommes en quelque sorte les héritiers d'une culture séculaire d'inhibition, d'insensibilisation envers le non-humain. Notre empathie à géométrie variable est le résultat d'un processus culturel. Par exemple, nous sommes habitués à considérer les animaux d'élevage comme des choses, des ressources, certes vivantes, mais destinées au sens plein du terme à nous nourrir. On se dit peut-être que c'est triste, mais qu'au final "c'est comme ça". [...] Dans ce contexte, je conçois le robot comme le doigt du sage qui montre la lune. Si nous voulons être malins, il faut regarder très attentivement cette lune qui n'est-autre que l'humain lui-même. Le robot social est

un selfie de l'homme moderne qui interroge en profondeur notre posture narcissique et l'écologie relationnelle que nous devrions adopter » (*ibid.*).

Nos hypothèses respectives partagent l'idée selon laquelle le « social » que l'on attribue au robot comme une propriété fait signe vers autre chose qu'une simple disposition à l'interaction, qu'il se prolonge dans un au-delà de lui-même qui est un appel symbolique au lien à l'altérité. Notre analyse donne le nom de responsabilité à ce lien, là où la sienne reste sur l'idée plus générale d'« écologie relationnelle ». En revanche, nous divergeons fortement quant au rôle du désir d'autre que la chercheuse place en amont du phénomène de socialité robotique :

« je ne crois pas que l'objet ait autant d'effet sur nous qu'on veut bien le dire. Je crois plutôt que nous avons un fort désir de voir en l'artefact une autonomie, des émotions, tout ce qui peut en faire un "autre", un compagnon envers lequel s'attacher. Bien sûr, il faut qu'une base perceptive conforte ce désir, un caillou ne provoque pas cette impression d'altérité chez nous, mais je pense que le désir est premier. Ce ne sont pas les règles sensorimotrices qui créent le sentiment qu'un NAO ou autres sont autonomes, mais plutôt notre fort "désir d'autre" que la biomécanique ne viole pas. Cela reboucle avec mon idée d'autres fantômes : ce désir est pour moi révélateur d'un manque, d'une souffrance liée à un déracinement relationnel » (*ibid.*).

En faisant du robot une sorte de réceptacle de nos projections, cette position renvoie l'artefact à la passivité de l'objet moderne offert à l'activité du sujet humain. Si l'altérité du robot ne provient que de notre désir, déjà présent en amont, alors l'idée de médiation sociobotique perd tout son sens. Malgré cela, nos hypothèses convergent à nouveau avec sa proposition finale :

« D'un point de vue plus philosophique, enfin, le robot est un outil conceptuel fort pour repenser la place que nous occupons dans le monde. Mon espoir est qu'il pourra contribuer à une prise de conscience, grâce à ce désir dont il est révélateur, de notre appartenance au monde et qu'il permettra de réviser des siècles de culture nous ayant extraits de cette appartenance » (*ibid.*).

Par ces mots, la chercheuse ne pointe rien d'autre que les dimensions symbolique et spéculaire de l'altérité sociobotique dont nous avons essayé de montrer qu'elles fondent, en termes éthiques et existentiels, la valeur fondamentale des robots sociaux.

Conclusion et ouverture - Et la politique ?

Pour conclure, résumons brièvement le chemin que nous avons parcouru tout au long de cette recherche. Si le sociobot appartient bien à la catégorie des robots en tant qu'il est un artefact qui entretient un rapport au monde de type sensorimoteur, force est de constater que ce rapport ne nous apparaît pas de la même façon selon qu'il s'agit bras robotique industriel type KR 1000 Titan de chez Kuka (cf. 1.5.1) ou d'un robot social comme Paro. D'où cette question : en quoi le *socio*-bot désigne-t-il un régime artefactuel bien particulier et que signifie ce « social » que le robot semble endosser comme une propriété ? Dans un premier temps, la littérature nous a permis d'identifier trois grandes réponses plus complémentaires qu'exclusives les unes des autres. Les deux premières nous ont conduit à affirmer qu'un robot peut être qualifié de social en ce qu'il est capable d'interagir socialement au sens, rappelons-le, d'établir des formes de communications rétroactives dont l'effet est de générer, au moins ponctuellement, des comportements émergents (c'est le social comme propriété dispositionnelle), et que cette capacité nous permet de lui attribuer un ou des rôles sociaux (c'est le social comme propriété systémique). Interactions et fonctions sociales, donc. Ces réponses permettent de proposer une caractérisation plus fine du sociobot comme robot dont la sensoriactivité est mise au service de fonctions intrinsèques ou extrinsèques à l'interaction sociale et qui, dans un cas comme dans l'autre, se réalisent à *travers* l'interaction sociale elle-même. Une telle définition ne prend toutefois pas en compte les dimensions du social qui échappent aux seules interactions, d'où notre troisième réponse centrée sur la perspective latourienne de l'interactance dans laquelle les interactions ne représentent qu'un type d'échange, qu'une modalité d'association possible entre actants. Mais que faire de la dimension immanquablement morale du social ? Que faire de cette idée, présente dans la littérature sans être suffisamment explicitée, selon laquelle les sociobots se présentent comme des « autres » ? En français, ne dit-on pas qu'on travaille « dans le social » pour tout ce qui relève de la main tendue à autrui et de la prise en charge, sous une forme ou sous une autre, de sa vulnérabilité ? Nous avons donc cherché à mettre en évidence un lien unissant socialité et moralité à travers l'exploration du concept d'altérité appliqué aux sociobots. En nous appuyant sur la notion d'ontophanie développée par Stéphane Vial et sur la théorie

postphénoménologique des médiations technologiques, nous avons décrit l'altérité artefactuelle comme un effet dans lequel le robot est expérimenté comme une présence qui résiste à l'ustensilité par son apparaître moral. Malgré son irréductible ambiguïté, celui-ci est ontophaniquement structuré comme « effet de personne vulnérable » dont le sens d'être, lui aussi ambigu, est à la fois de nous interpeller quant au rapport que nous entretenons à l'altérité (versant interrogatif) et de nous inciter à faire preuve de responsabilité face à sa vulnérabilité apparente (versant incitatif). L'altérité artefactuelle est donc, en tant que telle, également une specularité morale. Or, « quelle altérité ne l'est pas ? », demandera-t-on. L'originalité, ici, est précisément que nous ne sommes pas moralement et épistémiquement habitués à concevoir les techniques sous cet angle en raison de leur appartenance, supposée exclusive, à l'objectivité. Concevoir le robot social comme une effectivité morale, c'est donc s'affranchir du réflexe catégorique cherchant à l'analyser dans les termes de la dichotomie sujet/objet, c'est oublier la modernité telle qu'en parle Latour pour penser ce que nous faisons des artefacts sans déconnexion avec ce qu'ils font et nous font. Cette perspective nous a conduit à conceptualiser le sociobot comme une allotechnique au sens d'artefact dont la phénoménalité manifeste une exposition au monde et, plus spécifiquement, à nous-mêmes à travers le pouvoir que nous avons sur eux. De là, nous avons proposé de reconsidérer la socialité du robot à l'aune de cet apparaître responsabilisant. Plus précisément, le robot est social en vertu de sa capacité à faire naître en nous le sentiment de responsabilité à son égard. Pour finir, nous avons conclu cette recherche sur l'idée que le sociobot serait une technologie responsabilisante non seulement dans le face-à-face avec lui, mais aussi au-delà de lui, en tant qu'il participerait, par un processus d'herméneutique matérielle, d'une culture de la responsabilité envers la vulnérabilité.

Mais qu'en est-il du versant immanquablement politique de l'altérité artefactuelle que nous n'avons pas traité dans ce travail ? Dans l'entretien qu'il nous accordait en juillet 2018, Dominique Lestel tenait les propos suivants :

« Se focaliser sur l'éthique des robots, comme on le fait beaucoup aujourd'hui, est trompeur parce que la question est aujourd'hui plus politique qu'éthique. En d'autres termes, la question fondamentale aujourd'hui est celle de savoir qui commande *avec* les robots, et comment on commande aux commandeurs » (De Sanctis, 2018).

Si notre recherche s'est concentrée sur l'une des nombreuses dimensions éthiques du « problème sociobotique » qu'est l'altérité artefactuelle, celle-ci n'en demeure pas moins indissociable d'importantes questions politiques que nous voudrions évoquer ici sous forme d'ouverture. Le problème ne nous semble donc pas « plus politique qu'éthique », mais politique *parce qu'éthique* et vice versa. A ce titre, la robotique sociale et l'altérité artefactuelle pose une question majeure : quel monde commun façonnons-nous en nous dotant de ces technologies responsabilisantes et quels intérêts sont-elles susceptibles de servir ? Comme le souligne Verbeek lui-même, la postphénoménologie est parfois présentée comme incapable de faire naître une pensée politique des techniques en raison de son intérêt pour « l'échelle micro des relations humains-technologies ». D'autres auteurs font état d'un échec du corpus à penser adéquatement la question du pouvoir (Verbeek, 2020). Verbeek s'en défend et montre comment la postphénoménologie peut développer son versant politique à partir de l'herméneutique matérielle proposée par Ihde : « la médiation technologique des interprétations et des cadres culturels a d'importantes conséquences politiques » (*ibid.*, p.142). Le philosophe prend notamment l'exemple de ce que Robert Rosenberger nomme « technologies contre les sans-abris »¹⁸¹ tels que les bancs publics équipés d'accoudoirs visant à empêcher leur utilisation comme lit. Les artefacts, comme l'a bien montré Don Ihde (1990), sont « multistables » en ceci qu'« ils n'ont pas d'essence fixe, mais peuvent être interprétés de multiples manières » (Verbeek, 2020, p.144). Ainsi,

« [u]n banc public peut aussi être un lit, une poubelle peut aussi être une source de nourriture et une borne à incendie, une source d'eau potable. Le design anti sans-abri cherche à réduire cette multistabilité en bloquant l'"interprétation matérielle" qui permet aux personnes sans domicile fixe d'être dans certains espaces » (*ibid.*).

Cet exemple montre bien que l'espace public est, à travers sa composante technique, soumis à des conflits d'interprétation dont l'enjeu est son appropriation par certains groupes humains au détriment d'autres. Ici, il est donc bien question de pouvoir en ce qu'une certaine herméneutique matérielle cherche à devenir exclusive en excluant les autres. Se produit alors un phénomène que Rosenberger qualifie d'« occlusion » :

¹⁸¹ Rosenberger, R. (2017). *Callous objects: Designs against the homeless*. University of Minnesota Press.

« les usagers dominants des technologies perdent la capacité d’appréhender d’autres stabilités que celle qui est la plus évidente pour eux et, ainsi, ne voient pas comment l’environnement matériel exclut d’autres formes d’usages et d’autres types d’usagers. Les accoudoirs sur les bancs publics semblent simplement permettre aux gens de s’installer plus confortablement et démarquer les différentes places où s’asseoir, mais ce faisant, ils occultent le fait qu’ils empêchent de s’allonger sur les bancs et de les utiliser comme des lits » (*ibid.*).

Dès lors que les techniques organisent nos perceptions et nos interprétations, elles incarnent de subtiles formes de pouvoir. Voilà pourquoi tout l’enjeu d’une herméneutique politique des technologies est de rendre visible la multistabilité des artefacts et, plus spécifiquement, les conflits de pouvoir pouvant exister entre chaque stabilité et les groupes sociaux qui s’y rattachent. Une première question de philosophie politique des techniques appliquées aux sociobots pourrait donc être : de quelles stabilités ces artefacts sont-ils porteurs et comment cohabitent-elles dans leur espace d’intervention ?

L’intégration du concept de multistabilité à une herméneutique politique des techniques permet de mettre en évidence l’idée selon laquelle les usages d’une même technologie sont le théâtre d’enjeux politiques dont nous n’avons généralement pas conscience dans nos rapports immédiats aux artefacts. Ce constat rend incontournable la question des intérêts que tel ou tel usage peut servir, au-delà du seul bénéfice individuel de l’utilisateur, mais pose également la question des limites de la postphénoménologie lorsque les problèmes politiques dépassent le cadre de la relation humain-technologie ou, plus précisément, lorsqu’une telle relation repose sur des bases à la fois phénoménologiquement inaccessibles et socialement sujettes à caution voire tout à fait condamnables ? Par exemple, que peut la théorie des médiations technologiques face aux conditions sociales de production et de fonctionnement des programmes d’intelligence artificielle que nous utilisons quotidiennement ? Comment la postphénoménologie peut-elle intégrer à ses analyses politiques le sort de celles et ceux qu’Antonio Casilli (2019) a nommé les « tâcherons du numérique », ces « petits travailleurs anonymes qui se tuent à la micro-tâche dans « l’arrière-boutique » pour permettre aux ingénieurs de vanter les prouesses de leur créature sur la scène publique » (Kaufmann & Glassey, 2020, p.13) ? Dans un autre registre, comment prendre en compte les saccages environnementaux – qui, en réalité, sont aussi toujours des saccages sociaux – qu’impliquent les modalités actuelles d’extraction des « terres rares » nécessaires au fonctionnement de nos

téléphones, tablettes et autres ordinateurs¹⁸² ? De façon plus générale, le problème est le suivant : comment la postphénoménologie peut-elle intégrer des questions éthiques et politiques qu'une technologie donnée pose en dehors du périmètre de ses médiations ? Le tâcheron numérique ne fait assurément pas partie de la médiation constituée par mon « assistant vocal » tout comme la précarité et l'aliénation du livreur Uber Eats n'est pas visible lorsque je commande un repas sur l'application¹⁸³. L'accès à ces réalités n'est possible qu'à partir d'un *savoir* issu d'une perception extérieure à la médiation (par exemple, la lecture du livre de Casilli ou le visionnage un documentaire sur le sujet). Ce savoir, il est vrai, vient peut-être modifier la relation que j'entretiens avec l'artefact en question voire la supprimer (en boycottant l'application ou en cherchant une alternative plus vertueuse par exemple), mais une difficulté demeure malgré tout : la postphénoménologie n'est pas réellement équipée pour affronter le problème majeur et désormais vital des conditions socio-écologiques de production des technologies. Son intérêt éthique et politique porte sur le design en tant qu'il matérialise des valeurs et produit des médiations qu'il est possible d'anticiper, mais pas sur le cadre social dans lequel il s'inscrit. Pourtant, le caractère éthiquement désirable d'une technologie est loin de se limiter aux médiations qu'elle suscite. A titre d'exemple, l'option « ne pas déranger » de mon téléphone incite peut-être à adopter une hygiène pratique permettant de lutter contre l'aliénation que l'artefact contribue lui-même à instaurer (parasitage permanent de l'action par des messages, des notifications etc.), mais elle ne dit rien des milliers d'hommes et d'enfants congolais qui, équipés d'instruments sommaires, creusent jours et nuits pour extraire le cobalt nécessaire à la batterie de ce même téléphone (Frankel, 2016). La médiation ne dit rien non plus de la rémunération de ces « creuseurs », 1 à 2 euros si la récolte journalière est bonne, ou de l'exposition des communautés locales à des métaux toxiques propices aux développements de maladies respiratoires et d'anomalies congénitales (*ibid.*). Ces problématiques transverses à l'industrie numérique n'épargnent par la robotique sociale qui, bien sûr, pose également des questions plus spécifiques. Nous pensons par exemple au sujet des « affordances émotionnelles et axiologiques » (Kaufmann & Glassey, 2020) impliquées par l'altérité des robots sociaux et qui, à ce titre, participent de leur fonctionnement. Si ces affordances font effectivement partie des médiations

¹⁸² Sur ces questions, voir, par exemple, Pitron, G. (2018). *La guerre des métaux rares*. Les Liens qui Libèrent.

¹⁸³ Outre la littérature désormais foisonnante sur le sujet, nous renvoyons à l'excellente série documentaire intitulée « Invisibles. Les travailleurs du clic », disponible en rediffusion sur le site de France.tv (<https://www.france.tv/slash/invisibles/>).

sociobotiques en raison du rôle essentiel qu'elles jouent au sein des relations d'altérité avec un robot, qu'en est-il des intérêts qui président à leur conception ? Cette question des intérêts est intimement liée à celle du pouvoir que les sociétés conceptrices d'artefacts cherchent à exercer sur les individus et la collectivité à travers leurs produits. Or, là aussi, la postphénoménologie n'appréhende le pouvoir qu'en tant qu'il s'exprime au sein des médiations. La problématique politique et, plus précisément démocratique, nous dit Verbeek,

« dépasse la question de savoir dans quelle mesure nous devrions limiter le pouvoir d'entreprises influentes telles que Google et Facebook, au motif qu'elles exercent un pouvoir sur la société qui échappe au contrôle démocratique. Dans la perspective postphénoménologique, tout processus de conception technique dans lequel des médiations technologiques voient le jour devient *de facto* un lieu d'exercice de la démocratie. Rendre visibles les médiations technologiques des pratiques et des interprétations humaines rend possible d'anticiper ces médiations lors de la conception de nouvelles technologies » (Verbeek, 2020, p.145).

Or, le fait que le problème « aille plus loin », « dépasse » (*reaches further*) la seule question du pouvoir des firmes, ne rend aucunement cette question caduque, loin de là. Verbeek a beau considérer l'ingénieur ou le designer comme un praticien de « l'éthique par d'autres moyens » (2005) en tant qu'il n'est rien moins qu'un concepteur de médiations technologiques, ce travail reste fondamentalement tributaire des intérêts de l'entreprise qui l'emploie (où des États qui, de fait, sont eux aussi insérés dans la compétition économique mondiale). De même, l'idée qu'« on n'échappe pas aux médiations technologiques », comme l'affirme à raison le philosophe (Verbeek, 2020, p.144), ne signifie pas que les médiations épuisent le problème politique : la genèse sociale dont elles procèdent est tout aussi fondamentale pour construire une politique technologique désirable. Le propos de Verbeek que nous venons de citer s'inscrit, en creux, dans un dialogue avec la théorie critique appliquée à la technique et, plus précisément, avec le travail du philosophe canadien Andrew Feenberg dont il faut dire quelques mots avant de continuer.

La contribution de Feenberg à la philosophie des techniques partage, avec la postphénoménologie, ses emprunts aux STS et aux travaux de Latour qu'elle articule à la théorie critique de l'École de Francfort. Son projet philosophique est entièrement tourné vers la déconstruction du déterminisme technique et de l'obscurantisme technocratique par le

biais d'un constructivisme critique. Pour Feenberg, outre la quiétiste anxiété qu'elles génèrent, les ontologies négatives de la technique comme celle proposée par Heidegger manquent le cœur social du phénomène. Toute technologie résulte d'une logique que l'auteur nomme « instrumentalisation » et qui peut être analytiquement décomposée en instrumentalisation « primaire » et « secondaire ».

« [L'instrumentalisation primaire] démarre le processus de création d'un monde en "démondisant" ses objets pour révéler leurs possibilités d'application. Elle les arrache à leur contexte initial et les livre à l'analyse et à la manipulation [...]. "Dé-mondiser" c'est réduire les éléments de la réalité pouvant réaliser des fonctions à leurs caractéristiques fonctionnelles et les situer dans un nouveau contexte où ils servent des fins quelconques » (Feenberg, 2014, p.155).

En somme, l'instrumentalisation primaire consiste en une décontextualisation des objets de notre environnement dont le but est de les fonctionnaliser. Il s'agit d'une logique réductionniste où l'objet ne vaut que par la ou les qualités permettant de lui attribuer une certaine finalité. On peut illustrer très simplement cette idée en citant l'exemple de la pierre utilisée pour ouvrir un coquillage : la pierre est « dé-mondisée », arrachée à son contexte d'origine pour remplir la fonction d'ouvre-coquillage en raison de sa forme, de son poids et de sa nature contendante. Jusqu'ici, l'analyse rappelle la pensée heideggérienne de l'arraisonement. Toutefois, Feenberg précise que, contrairement à son réflexe classique, la philosophie ne peut pas prétendre cerner la technique en la résumant à ce seul processus de décontextualisation instrumentale. A cette genèse principielle abstraite s'ajoute une genèse sociale qui donne à la technique sa traduction concrète dans le monde. Feenberg nomme cette sociogenèse « instrumentalisation secondaire ». Ainsi, contrairement à Habermas qui pense que le mélange du technique et du social est extrinsèque au phénomène car accidentel, Feenberg affirme que la conception technique est techniquement sous-déterminée et socialement surdéterminée :

« Les choix sociaux interviennent dans la définition du problème comme dans sa solution. La technique n'est donc pas "rationnelle" au bon vieux sens positiviste du terme, mais elle dépend de la société » (*ibid.*, p.146).

Et l'auteur d'ajouter :

« La sous-détermination [technique] du développement technologique laisse une place permettant à des intérêts sociaux et à des valeurs de contribuer à ce processus. À mesure que les éléments décontextualisés sont combinés, ces intérêts et valeurs assignent des fonctions, orientent des choix et assurent l'adéquation entre la technique et la société » (*ibid.*, p.196).

La voiture, par exemple, est intégrée à différents environnements, différents contextes enchevêtrés : elle est faite pour rouler sur un réseau routier spécifique (contexte technique), répond à la norme politique et morale de liberté ainsi qu'à des critères esthétiques et fonctionnels particuliers. Séparer la voiture de ces environnements, c'est manquer la matrice sociale qui justifie son existence : toute technologie procède fondamentalement d'une réticularité sociotechnique. Ce n'est donc pas la technique en tant que telle qui représente un danger, mais son mode particulier de déploiement via l'instrumentalisation secondaire. Voilà pourquoi toutes les marges de manœuvre pour rendre la technique bonne, désirable sont à chercher dans l'instrumentalisation secondaire et son encadrement. Nourri par les thèses critiques de l'École de Francfort, en particulier celles de Marcuse, Feenberg juge que le véritable ennemi n'est autre que la matrice capitaliste de la technique :

« Le capitalisme introduit une rationalisation d'une envergure jusque-là inconnue : il décontextualise et simplifie un énorme éventail d'éléments de la société et de la nature pour les intégrer dans un système de production et de distribution. Les choses traitées comme de la matière brute sont arrachées de leurs emplacements naturels, démontées et traitées de façon à ne présenter que leur aspect utile dans un contexte de production. Au cours du processus de production, elles acquièrent de nouvelles qualités qui conviennent au contexte humain auquel elles sont destinées dans la consommation. On transforme aussi les personnes : on les retire, par exemple du contexte du travail domestique traditionnel, pour les installer dans des usines. Il est évident qu'on ne peut pas les dépouiller de leurs aspects non productifs comme on le fait pour les arbres ou les minéraux, mais les règles de l'usine peuvent les obliger à ne déployer dans leur travail que leurs qualités productives » (*ibid.*, p.336-337).

En somme, le capitalisme génère une instrumentalisation secondaire qui n'est autre que la forme techno-économique de l'instrumentalisation primaire. Il réduit chaque chose au rôle fonctionnel qu'elle peut jouer dans la génération du profit.

A nouveau, le déploiement capitaliste de la technique ressemble fortement à l'arrondissement heideggérien à cette différence près que le triptyque technique-capitalisme-arrondissement n'est pas justifié ontologiquement, mais socio-historiquement. Cette modalité d'administration de la technique n'a donc rien de nécessaire. Feenberg développe d'ailleurs sa philosophie dans le but de substituer une contingence qu'il juge positive à une autre qu'il pourfend. Pour l'auteur, une *bonne* technologie est une technologie dont l'instrumentalisation secondaire est construite démocratiquement et de façon inclusive. Or, l'un des principaux reproches pouvant être adressés au système techno-capitalisme en place est précisément celui de son caractère oligarchique. Selon Feenberg, l'origine sociale des technologies est encryptée dans leur matérialité et leur fonctionnement qui demandent des connaissances et savoir-faire précis. Il nomme cet encryptage *code technique*. La technocratie capitaliste joue sur cette codification pour renforcer l'illusion du caractère déterministe et exclusif du progrès technologique qui l'immunise contre la critique. Conforme au projet rationaliste moderne où le social au sens large et la technique sont nettement séparés, la technocratie réserve ce second domaine aux experts et expulse le citoyen des processus de conception. La dimension politique du problème se double d'une dimension proprement existentielle car le public est réduit au rôle de spectateur-consommateur face une technique qui, pourtant, construit le monde, le dévoile au sens heideggérien (c'est, là aussi, un point de convergence avec la postphénoménologie). La technique est anthropologiquement constituante au sens où elle définit autant l'humain qui la produit et l'utilise que l'inverse.

« [L]a théorie critique considère les techniques comme un environnement plutôt que comme une collection d'outils. Nous vivons actuellement *avec* et même *dans* les techniques qui organisent notre mode de vie » (*ibid.*, p.168)

Et le philosophe de poursuivre :

« [e]n tant que monde, les techniques conditionnent ceux qui y habitent. De ce point de vue, on peut les comparer aux lois et aux mœurs qui, dans l'antiquité, exerçaient sur les citoyens une influence presque parentale en les formant pour les rendre conforme au modèle souhaité d'être humain » (*ibid.*, p.168-169).

Face à l'extrême rationalité instrumentale du mode d'administration technocapitaliste de la technique, Feenberg propose de développer une *rationalité démocratique* où

l'instrumentalisation secondaire serait pleinement intégrée à l'exercice de la citoyenneté et où, par conséquent, le public serait maître des évolutions technologiques qui le co-constituent.

Tout en reconnaissant le très grand mérite des analyses développées par Feenberg, Verbeek déplore qu'elles s'inscrivent dans « une dialectique de l'oppression et de la libération » (Verbeek, 2013, p.72) où l'émancipation à l'égard du pouvoir technocapitaliste dominateur s'obtiendrait à partir d'un « *combat* entre la technique et la société » (*ibid*, p.73). Inspirée par la seconde conception foucauldienne du pouvoir, la postphénoménologie affirme qu'en matière technique on n'échappe jamais au pouvoir puisqu'on n'échappe jamais aux médiations technologiques. La seule véritable émancipation possible n'est pas celle menant à une disparition du pouvoir, mais celle conduisant à développer une relation *libre* à celui-ci en donnant une forme désirable à nos existences technologiquement médiées. En d'autres termes, le problème n'est pas tant le pouvoir que sa perversion qu'est la domination. La dialectique de Feenberg resterait donc en partie prisonnière du modèle de pensée dichotomique propre à la modernité laissant croire à un « extérieur », un « en dehors » (*an outside of*) de la technique d'où il serait possible d'en combattre les effets néfastes. Cette critique de la critique nous semble toutefois peu convaincante et ce pour au moins deux raisons. Premièrement, Feenberg ne nous semble pas rattacher son propos à une quelconque sphère extérieure à la technique puisque, comme nous venons de le voir, il la pense comme un « environnement ». Son ambition consiste plutôt à proposer un changement de régime politique refonder démocratiquement le mode de production et d'administration de la technique pour changer notre façon d'être « *avec* et même *dans* les techniques », selon ses propres termes. Il n'est pas question pour Feenberg de libérer l'humain de « la » technique, mais de libérer la technique elle-même du « projet » – il reprend le terme à Marcuse – technocapitaliste consistant *in fine* à assimiler les instrumentalisation primaire et secondaire en vue de la maximisation du profit. Deuxièmement, la critique de Verbeek pointe en dernière instance davantage vers une complémentarité des corpus que vers une opposition, comme l'indique plus ou moins explicitement ce passage :

« Au lieu de développer des tactiques pour contrer les stratégies des pouvoirs dominants, les citoyens d'un monde technologique devraient développer des techniques pour donner une forme désirable à leurs existences technologiquement médiées et pour organiser leur

société technologique. Cela ne diminue en rien l'importance de s'opposer au pouvoir des grandes firmes si elles constituent une menace envers la société. Mais cela montre aussi qu'une focalisation exclusive sur la dialectique de l'oppression et de la libération éludent des dimensions sociales et culturelles déterminantes de la technologie » (Verbeek, 2013, p.78).

Là où l'ambition politique de la postphénoménologie consiste à interroger la désirabilité des médiations technologiques en réintroduisant l'antique question de la vie bonne dans nos délibérations démocratiques, celle de la théorie critique de la technique se concentre sur l'appropriation démocratique des choix technologiques qui produisent le monde dans lequel nous vivons. La complémentarité des deux approches se situe dans l'articulation entre la question du mode de *production* des techniques et celles des *relations* que nous entretenons avec elles. Car si la production et l'administration démocratique des techniques ne constitue pas une garantie absolue contre l'indésirabilité des relations humains-artefacts, la délibération quant au sens et au déploiement concret que peuvent prendre de bonnes médiations technologiques ne peut pas porter sur les angles morts de ces médiations, c'est-à-dire sur ce qui ne relève pas directement de leur conception. Concrètement, nous pensons que cette complémentarité peut être théoriquement exploitée à partir du croisement entre le concept de médiations technologiques et celui de code technique afin d'aborder la politique (mais aussi l'éthique) des techniques tant du point de vue de leur production, de leur genèse sociale que des relations que nous entretenons avec elles. Revenant à notre question sur les affordances émotionnelles et axiologiques impliquées par l'altérité des sociobots, ce croisement conceptuel permettrait d'étudier dans le détail les impacts existentiels des allotechniques sans occulter la question des intérêts sociaux que celle-ci peuvent servir et qui n'apparaissent pas clairement dans les médiations. Ce second point ouvre, par exemple, à une analyse des *business models* présents et à venir de la sociobotique, notamment en ce qui concerne son insertion dans l'économie des données. L'empathie sur laquelle joue structurellement la sociobotique ne fait-elle pas du robot social un outil de choix pour capter toujours plus massivement des données, nourrissant et renforçant ce que Shoshana Zuboff (2020) a nommé « capitalisme de surveillance » ? L'interpellation morale propre à l'apparaître des allotechniques et le sentiment de responsabilité qu'elle implique ne constitue-t-ils pas le cheval de Troie idéal de la rationalité instrumentale et du phénomène d'arraisonnement dont elle est intrinsèquement porteuse ?

BIBLIOGRAPHIE

Livres et articles

Akrich, M. (2010). Comment décrire les objets techniques ?, *Techniques & Culture*, 54-55, p.205-219.

Airenti, G. (2012). Aux origines de l'anthropomorphisme. Intersubjectivité et théorie de l'esprit. *in Robots étrangement humains. Gradhiva*, 15, p.34-53.

Alenljung, B., Lindblom, J., Cort, R. & Ziemke, T. (2017). User Experience in Social Human-Robot Interaction. *International Journal of Ambient Computing and Intelligence*. 8. 12-31.

Allen, C., Smit, I., Wallach. (2005). Artificial Morality: Top-down, Bottom-up, and Hybrid Approaches. *Ethics and Information technology*. Vol. 7, p.149-155.

Anders, G. (2002). *L'obsolescence de l'homme. Sur l'âme à l'époque de la deuxième révolution industrielle* [1956], coll. L'Encyclopédie des Nuisances, Éditions Ivrea.

Alač, M. (2009). Moving android: On social robots and body-in-interaction. *Social Studies of Science* 39(4): 491–528.

Alač, M. (2016). Social robots: Things or agents?, *AI & SOCIETY*, 31(4): 519–535.

Alač, M., Movellan, J., Tanaka, F. (2011). When a robot is social: Spatial arrangements and multimodal semiotic engagement in the practice of social robotics. *Social Studies of Science*, 41(6), p.126-159.

Anderson, M., Anderson, S-L. (2006). Machine ethics. *IEEE Intelligent Systems* 21(4): 10-11.

Aroyo, A., Kyohei, T., Koyama, T., Takahashi, H., Rea, F., Sciutti, A., et al. (2018). Will people morally crack under the authority of a famous wicked robot? in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (Nanjing: IEEE), 35–42.

Asaro, P-M. (2006). What should we want from a robot ethics ?. *International Review of Information Ethics*, 6(12):9–16.

Asimov, I. (2012). Le cycle des robots. 1- Les robots [1950]. *J'ai lu*, n°453.

Audétat, M. (dir) (2015). Sciences et technologies émergentes : pourquoi tant de promesses ?. Hermann.

Auger, J. H. (2014). Living with robots: A speculative design approach. *Journal of Human-Robot Interaction*, 3(1), 20–42.

Barbaras, R. (1999). Autrui. Quintette.

Bartneck, C., Forlizzi, J. (2004). A design-centred framework for social human-robot interaction. In: 13th IEEE International Workshop on Robot and Human Interactive Communication, ROMAN 2004, pp. 591–594. IEEE

Beaune, J-C. (1985). L’homme machinal et la machine humaine. *Les Études Philosophiques*, 1: 33-44.

Bentham, J. (2011). *Introduction aux principes de morale et de législation* [1789]. Paris. Coll. Analyse et Philosophie. Vrin.

Bergen, J. P., & Verbeek, P. P. (2021). To-do is to be: Foucault, Levinas, and technologically mediated subjectivation. *Philosophy & Technology*, 34(2), 325-348.

Biocca, F., Harms, C., and Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence. *Presence* 12, 456–480.

Bisconti Lucidi, P., Nardi D. (2018). Companion Robots: the Hallucinatory Danger of Human-Robot Interactions. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics and Society*, p.17-22.

Bontems, V. (2018). La machine respectueuse. L’éthique des techniques de Simondon à l’ère des robots. *Revue française d’éthique appliquée*, 5(1), p.22-33.

Borelle, C. (2018), « Sortir du débat ontologique. Éléments pour une sociologie pragmatique des interactions entre humains et êtres artificiels intelligents », *Réseaux, La Découverte*, n° 212, pages 207 à 231

Borelle, C. (2020). « La moralisation des robots sociaux par leurs utilisateurs », *Sociologie du travail* [En ligne], Vol. 62 - n° 1-2 | Janvier-Juin 2020, mis en ligne le 01 juin 2020.

Bovo, E. (2008). « L’exil des choses. Réflexions sur la technique à partir de Martin Heidegger et Emmanuel Levinas ». In Schmit, P-E & Chardel P-A. (dir.) (2008). *Phénoménologie et technique(s)*. Coll. Phéno. Le Cercle Herméneutique Éditeur.

Brey, P. (2006). The social agency of technological artifacts. In: Verbeek Peter-Paul, Slob Adriaan (eds) *User Behavior and Technology Development*. Springer, Dordrecht, pp.71–80.

Breazeal, C. (2002). *Designing sociable robots*, Cambridge, MIT Press, 2002.

Breazeal, C. (2003a). « Toward Sociable Robots », in *Robotics and Autonomous Systems*, vol. 42, p. 167-175.

Breazeal, C. (2003b). « Emotion and Sociable Humanoid Robots », *International Journal of Human-Computer Studies*, vol. 59 (1-2), p.119-155.

- Breazeal, C. (2009). Role of expressive behaviour for robots that learn from people", *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364 (1535), 3527.
- Breazeal, C. & Brooks, R. (2004). Robot emotion: a functional perspective. In *Who Needs Emotions: The Brain Meets the Robot*, ed. J. M. Fellous and M. Arbib, Oxford: Oxford University Press, 271-310.
- Bryson, J. (2010). Robots Should Be Slaves. In Y. Wilks (Ed.), *Close engagements with artificial companions: Key social, psychological, ethical and design issues*. Amsterdam: John Benjamins, 63-74
- Buber, M. (2012). *Je et tu* [1923]. Aubier.
- Burneleit, E., Hemmert, F., & Wettach, R. (2009). Living interfaces: The impatient toaster. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction (TEI)*, 21–22. Cambridge, UK.
- Bushman, B. J. (2002). Does Venting Anger Feed or Extinguish the Flame? Catharsis, Rumination, Distraction, Anger, and Aggressive Responding. *Personality and Social Psychology Bulletin*, 28(6), 724–731.
- Caeymaex, F., Despret, V., & Piéron, J. (2019). *Habiter le trouble avec Donna Haraway*. Éditions du Dehors.
- Canguilhem, G. (1966). *Le Normal et le pathologique*. Paris : Presses Universitaires de France.
- Canguilhem, G. (2001 [1955]). Le problème des régulations dans l'organisme et dans la société. *Écrits sur la médecine*, Paris, Seuil.
- Čapek, K. (2011), *R.U.R.* [1920]., Paris, La Différence.
- Casilli, A. (2019). En attendant les robots. Enquête sur le travail du clic. Seuil.
- Claverie, B., Le Blanc, B., Fouillat, P. (2013). La cobotique : La robotique soumise, *Communication & Organisation*, 44, p. 203-213, Presses Universitaires de Bordeaux.
- Chatila, R. (2014). Robotique et simplicité: modèles, architecture, décision et conscience. A. Berthoz, & J. Petit, *Complexité-Simplicité*. Paris: Collège de France.
- Citton, Y. (2014). Pour une écologie de l'attention. Paris : Seuil, coll. La couleur des idées.
- Coeckelbergh, M. (2009a). Personal Robots, Appearance and Human Good - A Methodological Reflection on Roboethics. *International Journal of Social Robotics*, 1(3), p.217-221.

- Coeckelbergh, M. (2009b). Virtual moral agency, virtual moral responsibility: on the moral significance of the appearance, perception, and performance of artificial agents ». *AI and Society*, Vol. 24, p.181-189, 2009b.
- Coeckelbergh, M. (2010a). Moral appearances: emotions, robots, and human morality. *Ethics and Information Technology*, 12: 235-241.
- Coeckelbergh, M. (2010b). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology*, 12:209-221.
- Coeckelbergh, M. (2010c). Artificial Companions: Empathy and Vulnerability Mirroring in Human-Robot Relations. *Studies in Ethics, Law, and Technology*, Vol.4: Iss 3, Article 2.
- Coeckelbergh, M. (2011a). Humans, Animals, and Robot: A Phenomenological Approach to Human-Robot Relations, *Philosophy and Technology*, 24(3): 269-278, 2011a.
- Coeckelbergh, M. (2011b). You, robot: On the linguistic construction of artificial others. *AI & Society*, 26(1), p.61-69.
- Coeckelbergh, Mark. (2012a). *Growing Moral Relations: Critique of Moral Status Ascription*. New York: Palgrave MacMillan.
- Coeckelbergh, Mark. (2012b). Care robots, Virtual Virtue, and the Best Possible Life. in Brey, P., Briggel, A., Spence, E. (eds). *The Good Life in a Technological Age*, New York/London: Routledge, p. 281-292.
- Coeckelbergh, Mark. (2016). Alterity Ex Machina: The Encounter With Technology as an Epistemological-Ethical Drama. In Gunkel, D. J., Marcondes Filho, C., Mersch, D. *The Changing Face of Alterity*, London and Maryland: Rowman & Littlefield International, Media Philosophy, p.181-196.
- Coeckelbergh, Mark. (2018). Why care about robots? Empathy, moral standing, and the language of suffering. *Kairos. Journal of Philosophy & Science*, 20(1), p.141–158.
- Coeckelbergh, M. (2021). Does kindness towards robots lead to virtue? A reply to Sparrow's asymmetry argument. *Ethics and Information Technology*.
- Corval, C., (2018). Les réceptions philosophiques de l'automate au XVIIIe siècle en France. Usages comparatifs et thématiques de la machine mouvante chez La Mettrie, Voltaire, Diderot, Rousseau et Vaucanson. Mémoire de Master 2 en Histoire de la philosophie (dir. Claire Crignon). Paris, Sorbonne Université.
- Constant, B. (2013). *Des réactions politiques* [1796]. Paris. Hachette Livre BNF.
- Damiano, L., & Dumouchel, P. (2018). Anthropomorphism in human–robot co-evolution. *Frontiers in psychology*, 9, 468.

- Danaher, J. (2020). Welcoming robots into the moral circle: A defence of ethical behaviourism. *Science and Engineering Ethics*; 26(4):2023–2049.
- Darling, K. (2016). Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects. In *Robot law*. Edward Elgar Publishing.
- Dautenhahn, K., & Billard, A. (1999). Bringing up robots or—the psychology of socially intelligent robots: From theory to implementation. In *Proceedings of the third annual conference on Autonomous Agents* (pp. 366-367).
- Dautenhahn, K. (2007). Socially intelligent robots: Dimensions of human-robot interaction. *Philosophical Transactions of the Royal Society B*, 362(1480), 679-704.
- Decety, J., & Cowell, J. M. (2014). The complex relation between morality and empathy. *Trends in cognitive sciences*, 18(7), 337-339.
- Deslandes, G., & Paltrinieri, L. (2017). Entretien avec Bernard stiegler. *Rue Descartes*, 91(1), 119-140.
- Demange, M., Lenoir, H., Pino, M., Cantegreil-Kallen, I., Rigaud, A. S., & Cristancho-Lacroix, V. (2018). Improving well-being in patients with major neurodegenerative disorders: differential efficacy of brief social robot-based intervention for 3 neuropsychiatric profiles. *Clinical interventions in aging*, 13, 1303.
- Dennett, D. (1990). *La Stratégie de l'interprète : le sens commun et l'univers quotidien*. Paris : Gallimard.
- Descola, P. (2005). *Par-delà nature et culture*. Paris : Gallimard.
- Dewey, J. (2011). *La formation des valeurs*. Paris : La Découverte.
- Dorrestijn, S. (2012a). Technical mediation and subjectivation: Tracing and extending Foucault's philosophy of technology. *Philosophy & Technology*, 25(2), 221-241.
- Dorrestijn, S. (2012b). *The Design of our own lives. Technical mediation and subjectivation after Foucault*, PhD dissertation, University of Twente.
- Dorrestijn, S. & Verbeek, P-P. (2013). Technology, wellbeing, and freedom: The legacy of utopian design. *International Journal of Design*, 7(3), 45-56.
- Dumouchel, P. & Damiano, L. (2016). *Vivre avec les robots. Essai sur l'empathie artificielle*. Paris. Seuil.
- Duffy, B. R., Rooney, C., O'Hare, G. M., & O'Donoghue, R. (1999). What is a social robot?. In *10th Irish Conference on Artificial Intelligence & Cognitive Science, University College Cork, Ireland, 1-3 September, 1999*.

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and autonomous systems*, 42(3-4), 177-190.

Grimaud, E., Taylor, A. C., Vidal, D., & Dufrêne, T. (2015). Qui est là? Présences-limites et effets de personne. *Persona. Étrangement humain*. Catalogue de l'exposition au Musée du Quai Branly. Paris. Actes Sud, pp. 11-17.

Feenberg, A. (2004). *(Re)penser la technique*, La Découverte/M.A.U.S.S.

Feenberg, A. (2014a). *Pour une théorie critique de la technique*, Lux.

Feenberg, A. (2014b). Vers une théorie critique de l'Internet. *tic&société*, 8(1-2).

Feil-Seifer, D., Mataric, M.J. (2011). Ethical principles for socially assistive robotics. *IEEE Robotics & Automation Magazine*, 18(1):24–31.

Fink, J. (2012). Anthropomorphism and human likeness in the design of robots and human-robot interaction. In *International conference on social robotics* (pp. 199-208). Springer, Berlin, Heidelberg.

Fischer, F. (2020). *Les normativités des technologies numériques : approche d'une éthique « by design »* [Thèse de doctorat, Université de Technologie de Compiègne]. thèses.fr. <https://www.theses.fr/2020COMP2583>

Floridi, L. & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and machines*, 14(3), 349-379.

Floridi, L. (2010). Information ethics. In *Cambridge Handbook of Information and Computer Ethics*, ed. Luciano Floridi, 77 – 100. Cambridge: Cambridge University Press.

Floridi, L. (2014). Artificial agents and their moral nature. In *The Moral Status of Technical Artefacts* (pp. 185-212). Springer, Dordrecht.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3-4), 143-166.

Forlizzi, J., DiSalvo, C., & Gemperle, F. (2004). Assistive robotics and an ecology of elders living independently in their homes. *Human-Computer Interaction*, 19(1-2), 25-59.

Forlizzi, J. (2007a). How robotic products become social products: an ethnographic study of cleaning in the home. In: *ACM SIGCHI/SIGART human-robot interaction*, p.129–136.

Forlizzi, J. (2007b). The product ecology: Understanding social product use and supporting design culture. *International Journal of Design*, 2(1), 11-20.

- Forlizzi, J., & Battarbee, K. (2004). Understanding Experience in Interactive Systems. Proceedings of the DIS2004, Cambridge.
- Foucault, M. (1984). *Histoire de la sexualité II. L'usage des plaisirs*. Paris. Coll. Tel. Gallimard.
- Friedman, B. & Kahn, P. H., Jr. (2003). Human values, ethics, and design. In Jacko, J. A. & Sears, A. (Eds.), *The human-computer interaction handbook* (pp. 1177–1201). Mahwah, NJ: Lawrence Erlbaum Associates.
- Frumer, Y. (2018). Cognition and emotions in Japanese humanoid robotics”, *History and Technology*, 34:2, p.157-183.
- Frumer, Y. (2020). The Short, and Strange Life of the First Friendly Robot. *IEEE spectrum*, 20/04/2020. <https://spectrum.ieee.org/robotics/humanoids/the-short-strange-life-of-the-first-friendly-robot>,
- Gallese, V. (2005). “Being like me”: Self-other identity, mirror neurons and empathy. In: S. Hurley and N. Chater (eds.), *Perspectives on Imitation: From Cognitive Neuroscience to Social Science* (Vol. 1). Cambridge, MA: MIT Press.
- Ganascia, J-G. (2017). *Intelligence artificielle. Vers une domination programmée ?*. Le Cavalier Bleu Éditions, coll. Idées Reçues.
- Giard, A. (2016a). Un désir d’humain. Les « love doll » au Japon. Les Belles Lettres.
- Giard, A. (2016b). La love doll au Japon : jeux imaginaires, incarnation et paradoxes. *Interrogations*, n° 23.
- Gibert, M., & Paris, M. (2010). L’imagination et les biais de l’empathie. In *Les ateliers de l'éthique/The Ethics Forum* (Vol. 5, No. 1, pp. 50-65). Centre de recherche en éthique de l’Université de Montréal.
- Grimaud, E. & Pare, Z. (2011). Le jour où les robots mangeront des pommes. Éditions Pétra.
- Grimaud, E. (2012). Androïde cherche humain pour contact électrique. *Les cinétiques de l’attachement en robotique* », *Gradhiva* [en ligne], vol. 15, n° 1.
- Grimaud, E. (2015). Les robots oscillent entre vivant et inerte. *Multitudes*, 58(1) : 45-58.
- Grimaud, E. & Vidal, D. (2012). Aux frontières de l’humain. Pour une anthropologie comparée des créatures artificielles. *Gradhiva*, vol. 15, n° 1.
- Guchet, X. (2015). Pour un humanisme technologique. Culture, technique et société dans la philosophie de Gilbert Simondon. PUF.
- Guchet, X. (2017). *Objet versus artefact*. Pour une philosophie des techniques orientée-objet, māj 0000, *Cahiers COSTECH* n°1.

- Guchet, X. (2022). *Du soin dans la technique : Question philosophique (Vol.6)*. ISTE Group.
- Gunkel, D. J. (2007). *The Machine Question: Ethics, Alterity, and Technology. Series in Philosophy/Communication, 4*, 120-150.
- Gunkel, D. J. (2012). *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. Cambridge: MIT Press.
- Gunkel, D. J. (2014). The rights of machines: caring for robotic care-givers. In: van Rysewyk SP, Matthijs P (eds) *Machine medical ethics*. Springer, Heidelberg, pp 151–166
- Gunkel, D. J. (2016). Another Alterity. Rethinking Ethics in the Face of the Machine. In Gunkel, D. J., Marcondes Filho, C., Mersch, D., *The Changing Face of Alterity*, London and Maryland: Rowman & Littlefield International, Media Philosophy, p.197-218.
- Gunkel, D. J. (2017). The Other question: can and should robots have rights?. *Ethics and Information Technology*, 20:87–99.
- Haraway, D. (2019). *Manifeste des espèces compagnes*. Paris : Flammarion.
- Hassenzahl, M. (2010). *Experience Design: Technology for all the right reasons*. San Rafael, CA: Morgan & Claypool.
- Hassenzahl, M., & Tractinsky, N. (2006). User experience – a research agenda. *Behaviour & Information Technology*, 25(2), 91-97.
- Heerink, M., Kröse, B., Evers, V. & Wielinga, B. (2008). The influence of social presence on acceptance of a companion robot by older people. *Journal of Physical Agents*, Vol. 2 No. 2, pp.33-40.
- Heidegger, E. (1985 [1927]). *Être et temps*. trad. E. Martineau. Paris : Authentica.
- Heidegger, E. (2013 [1958]). La question de la technique. In *Essais et Conférences*, Paris : Gallimard, coll. « Tel », 2013, p.9-48.
- Hegel, F., Muhl, C., Wrede, B. *et al.* (2009). Understanding social robots. *Advances in Computer-Human Interactions, ACHI'09*: 169-174.
- Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, 57(2), 243.
- Himma, K. E. (2009). Artificial agency, consciousness, and the criteria for moral agency: what properties must an artificial agent have to be a moral agent?. *Ethics Inf Technol* 11, 19–29.
- Hoffman, G., & Ju, W. (2014). Designing robots with movement in mind. *Journal of Human-Robot Interaction*, 3(1), 89–122.

- Houkes, W. & Vermaas, P. (2004). Actions versus Functions: A Plea for an Alternative Metaphysics of Artifacts. In *The Monist*, Vol.87, n°1, p.52-71.
- Houkes, W. & Meijers, A. (2006). The ontology of artifacts: the hard problem. *Studies in History and Philosophy of Science*, 37, 118–131.
- Ihde, D. (1990). *Technology and the lifeworld. From Garden to Earth*. Indiana University Press.
- Ihde, D. (1998). *Expanding hermeneutics: visualism in science (studies in phenomenology and existential philosophy)*. Northwestern University Press, Evanston
- Jonas, H. (1995 [1979]). *Le principe responsabilité. Une éthique pour la civilisation technologique*. Paris : Cerf.
- Jonas, H. (1998). *Pour une éthique du futur*. Paris : Rivages Poche / Petite Bibliothèque.
- Jones, R. A. (2017). What makes a robot social. *Social Studies of Science*, 47, p.1-24.
- Kahn, P. H. (Jr), Freier, N. G., Friedman, B., Severson, R. L. & Feldman, E. (2004). Social and moral relationships with robotic others?. Proc. 13th Int. Workshop Robot Hum. Interact. Commun. (ROMAN) (2004) pp. 545–550
- Kant, E. (1992 [1785]). *Fondements de la métaphysique des mœurs*. Paris. Vrin.
- Kaufmann, L., & Glassey, O. (2020). Mon nom est personne. Éprouver l'(a)moralité des humanoïdes. *SociologieS*.
- Key, B. (2015). Fish do not feel pain and its implications for understanding phenomenal consciousness. *Biology and Philosophy* 2015;30(2):149–165.
- Kleinpeter, E. (2015). Le Cobot, la coopération entre l'utilisateur et la machine. *Multitudes*, 58, (1), p. 70-75.
- Kroes, P. & Meijers, A. (2002). Reply to critics. *Techné: Research in Philosophy and Technology*, 6(2), 34-43.
- Kroes, P. & Meijers, A. (2006). Introduction: the dual nature of technical artifacts. *Studies in History and Philosophy of Science*, 37, 1–4.
- Kudina, O. & Verbeek, P-P. (2019). Ethics from Within: Google Glass, the Collingridge Dilemma, and the Mediated Value of Privacy. *Science, Technology, & Human Values*, Vol.44(2), 291-314.
- Laplane, L. & Solary, E. (2017). Identité des cellules souches normales et cancéreuses. *Médecine/science*, 33, 899–904.

Latour, B. (1991). *Nous n'avons jamais été modernes*. Paris : La Découverte.

Latour, B. (1992). Where are the missing masses, sociology for a few mundane artefacts. In *Shaping Technology-Building Society. Studies in Sociotechnical Change*. Bijker, W. & Law, J. (ed.), MIT Press, Cambridge Mass., p. 225-259.

Latour, B. (1994a). On Technical Mediation – Philosophy, Sociology, Genealogy. *Common Knowledge*, Vol.3, n°2, p.29-64.

Latour, B. (1994b). Une sociologie sans objet ? Remarques sur l'interobjectivité. *Sociologie du travail*, vol. 36, n°4, p.587-607.

Latour, B. (2000). La fin des moyens. *Réseaux. Communication-Technologie-Société*, 18(100), 39-58.

Latour, B. (2007), *Changer de société, refaire de la sociologie*, Paris, La Découverte.

Lazaric, N. (1992). Origines et développement de la robotique. *Revue d'économie industrielle*, vol. 61. p.54-67.

Lenay, C. (2010). C'est très touchant. La valeur émotionnelle du contact. *Intellectica*, 53(1), 359-397.

Lestel, D. (2021). *Machines insurrectionnelles: Une théorie post-biologique du vivant*. Fayard.

Levillain, F., Zibetti, E. (2017). Behavioral Objects: The Rise of the Evocative Machines. *Journal of Human-Robot Interactions*, Vol.6, No.1, p.4-24.

Levinas, E. (1982). *Éthique et infini. Dialogue avec Philippe Nemo*. Coll. *biblio essais*. Le Livre de Poche.

Levinas, E. (1990). *Totalité et Infini. Essai sur l'extériorité* [1961]. Coll. *biblio essais*. Le Livre de Poche.

Marzano, M. (2016). *La philosophie du corps*. Coll. *Que sais-je ?*. PUF.

Marion, G. (2017). L'émergence de la valeur d'usage et l'« agentivité » des objets matériels. *Revue française de gestion*, 2017/4, n° 265, p. 71-92.

Moor, J. H. (2006). The Nature, Importance and Difficulty of Machine Ethics. *IEEE Intelligent Systems*, 18-21.

Morizot, B. (2020). *Raviver les braises du vivant. Un front commun*. Paris. Coll. *Domaine du Possible*. Actes Sud / Wild Project.

Munier, B. (2014). Homme et machine : la phase du miroir. *Hermès*, vol. 68, p. 143-148.

Neely, E-L. (2014). Machines and the moral community. *Philosophy & Technology*. 27(1):97–111

Nevejans, N. (2017). *Traité de droit et d'éthique de la robotique civile*. LEH Éditions.

Oudeyer, P. Y. (2016). Des ordinateurs aux robots : les machines en informatique. <http://www.pyoudeyer.com/oudeyerMachinesRobots123Codez.pdf>

Peshkin, M. & Colgate J. E. (1999). « Cobots ». *Industrial Robot*, vol. 26, n° 5, p. 335-341.

Prairat, E. (2012). La responsabilité. *Le Télémaque*, (2), p.19-34.

Quéré, L. (1988). Sociabilité et interactions sociales. *Réseaux. Communication-Technologie-Société*, 6(29), 75-91.

Rosenthal-von der Pütten A. M., Krämer, N. C., Hoffmann, L., Sobieraj S. & Eimler, S. C. (2013). An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics*; 5 (1):17–34.

Rouvroy, A., & Berns, T. (2013). Gouvernamentalité algorithmique et perspectives d'émancipation. *Réseaux*, (1), 163-196.

Sabanovic, S., Bennett, C. C., Chang, W-L. & Huber, L. (2013). PARO robot affects diverse interaction modalities in group sensory therapy for older adults with dementia. *IEEE Int Conf Rehabil Robot 2013*; 2013:6650427.

Sabanovic, S. & Chang, W-L. (2016). Socializing robots: constructing robotic sociality in the design and use of the assistive robot PARO. *AI & Society*, 31, 537–551.

Sakka, S., Gaboriau, R. *et al* (2016). Rob'autism: how to change autistic social skills in 20 weeks. In *5th International Workshop on Medical and Service Robots*.

Sarrica, M., Brondi, S. & Fortunati, L. (2019). How many facets does a “social robot” have? A review of scientific and popular definitions online. *Information Technology & People*.

Sartre, J-P. (1976). *L'Être et le Néant* (1943). Paris : Gallimard.

Sciamma, D. (2012a). Vivre avec des robots : designer la relation. *Annales des Mines - Réalités industrielles*, février 2012(1), p.103-108.

Sciamma, D. (2012b). Demain le design. De l'interaction à la relation. *Interfaces numériques*, 2012, vol. 1, n°1.

Schaer, R. (2020). *Répondre du vivant*. Le Pommier.

Sebbah, F. D. (2006). Visage de clone. *Les études philosophiques*, (3), 353-366.

Sharkey, A. & Sharkey, N. (2010). Granny and the robots: Ethical issues in robot care for the elderly. *Ethics and Information Technology* 14, 27-40.

Sharkey, A. & Sharkey, N. (2020). "We need to talk about deception in social robotics!". *Ethics and Information Technology*, 23, 309–316.

Shevlin, H. (2021). How Could We Know When a Robot Was a Moral Patient?. *Cambridge Quarterly of Healthcare Ethics*, 30(3), 459-471.

Shibata, T. (2012), "Therapeutic seal robot as biofeedback medical device: qualitative and quantitative evaluations of robot therapy in dementia care", *Proc. IEEE* 100, 2527–2538.

Shibata, T., & Coughlin, J. F. (2014). Trends of robot therapy with neurological therapeutic seal robot. *PARO. Journal of Robotics and Mechatronics*, 26, 418–425.

Simondon, G. (2012). *Du mode d'existence des objets techniques* [1958]. Aubier.

Simondon, G. (1983). Trois perspectives pour une réflexion sur l'éthique et la technique. In *Sur la technique. (1953-1983)*. Paris : PUF, « Hors collection », 2014, p. 337-351.

Singer, P. (2012). *La libération animale* [1975]. Payot. Coll. Petite Bibliothèque Payot.

Sparrow, R. & Sparrow, L. (2006). In the hands of machines? The future of aged care. *Minds & Machines* 16, p.141–161.

Sparrow, R. (2020). Virtue and vice in our relationships with robots: Is there an asymmetry and how might it be explained? *International Journal of Social Robotics*.

Strum, S. & Latour, B. (1987), Redéfinir le lien social : des babouins aux humains, in Akrich, M., Callon, M. & Latour, B. (dir.) (2006). *Sociologie de la traduction*. Presse des Mines, p.71-86.

Suzuki, Y., Galli, L., Ikeda, A., Itakura, S., Kitazaki, M. (2015). Measuring empathy for human and robot hand pain using electroencephalography. *Scientific Reports*; 5:15924.

Takayama, L., Ju, W. & Nass, C. (2008). Beyond dirty, dangerous and dull: what everyday people think robots should do. In *Proceedings of the 3rd ACM/IEEE international conference on human robot interaction, HRI '08*. ACM, New York, NY, USA, p. 25-32.

Tisseron, S. (2015). *Le jour où mon robot m'aimera. Vers l'empathie artificielle*. Paris, Albin Michel.

Tisseron, S. (2017). *Empathie et manipulations : les pièges de la compassion*. Paris, Albin Michel.

Tronto, J. (2009). *Un monde vulnérable. Pour une politique du "care"*. Paris : La Découverte.

- Turkle, S. (2006). A nascent robotics culture: New complicities for companionship. *AAAI Technical Report Series*.
- Turkle, S. (2007). Authenticity in the age of digital companions. *Interaction Studies*, 8(3), 501517.
- Turkle, S. (2015). Seuls ensemble. De plus en plus de technologies, de moins ne moins de relations humaines. L'Échappée.
- Turkle, S. (2018). En parlant aux machines, nous perdons notre humanité. *Hermès, La Revue*. 80.
- Turkle, S., Breazeal, C., Dasté, O., & Scassellati, B. (2006). First encounters with Kismet and Cog: Children respond to relational artifacts. In Messaris, P. & Humphreys, L. (Eds.), *Digital media: Transformations in human communication*. New York: Peter Lang Publishing.
- Tzafestas, S. G. (2016). *Roboethics: A Navigating Overview*. Heilberg: Springer.
- Van Wynsberghe, A. (2020). Designing robots for care: Care centered value-sensitive design. In *Machine Ethics and Robot Ethics* (pp. 185-211). Routledge.
- Verbeek, P-P. (2005). *What Things Do. Philosophical reflections on technology, agency and design*. Penn State Press.
- Verbeek, P-P. (2010). Accompanying Technology: Philosophy of Technology after the Ethical Turn. *Technè: Research in Philosophy and Technology*, 14(1), p.49-54.
- Verbeek, P-P. (2011). *Moralizing Technology. Understanding and designing the morality of things*. The University of Chicago Press.
- Verbeek, P-P. (2013). Resistance is Futile: Toward a Non-Modern Democratization of Technology", *Technè : Research in Philosophy and Technology*, 17:1, p.72-92.
- Verbeek, P-P. (2015). "Beyond Interaction: A Short Introduction to Mediation Theory.", *Interactions* 12 (3): 26-31.
- Verbeek, P-P. (2020). Politicizing postphenomenology. In *Reimagining philosophy and technology, reinventing Ihde* (pp. 141-155). Springer, Cham.
- Vernon, D. (2014). *Artificial Cognitive Systems: A Primer*. Cambridge, MA: MITPress.
- Vial, S. (2013). *L'être et l'écran. Comment le numérique change la perception*. Paris : PUF.
- Vial, S. (2014a). Ce que le numérique change à autrui : introduction à la fabrique phénoménoteknique de l'altérité. *Hermès, La Revue*, n°68, p. 151-157.

Vial, S. (2014b). *Court traité du Design*. Paris : PUF.

Vidal, D. (2012). Vers un nouveau pacte anthropomorphique ! Les enjeux anthropologiques de la nouvelle robotique. *Gradhiva*, (15), 54–75.

Vidal, D. (2016). *Aux frontières de l’humain, dieux, figures de cire, robots et autres artefacts*. Paris. Alma.

Vidal, D. & Gaussier, P. (2014). Un robot comme personne. Ontologies comparée et expérimentale au musée du quai Branly. *Terrain*, n° 62, p.152-165.

Wada, K., Shibata, T., Saito, T., Sakamoto, K., & Tanie, K. (2005). Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged. In *Proceedings of the 2005 IEEE international conference on robotics and automation* (pp. 2785-2790). IEEE.

Winner, L. (1980). Do Artifacts Have Politics?. *Daedalus*, Vol. 109, No. 1, Modern Technology: Problem or Opportunity?, p.121-136.

Worms, F. (2016). Un moment du vivant ?. Dans : François, A. & Worms, F. (dir.), *Le moment du vivant* (pp. 15-25). Paris : Presses Universitaires de France.

Wright, J. (2018). Technowelfare in Japan: Personal Care Robots and Temporalities of Care. Ph.D. Thesis, Hong Kong Institute for the Humanities and Social Sciences, University of Hong Kong.

Zuboff, S. (2020). L’âge du capitalisme de surveillance. Zulma.

Articles de media en ligne

Ackerman, E. (2019). This “Useless” Robot Wants to Succeed Where Others Failed. *IEEE Spectrum*, 19/09/2019. <https://spectrum.ieee.org/kiki-social-home-robot>

Ackerman, E. (2020), “Meet Moxie, a Social Robot That Helps Kids with Social-Emotional Learning”, *IEEE Spectrum*, 28/04/2020. <https://spectrum.ieee.org/automaton/robotics/home-robots/moxie-a-social-robot-for-childhood-development>

Ackerman, E. (2021). SoftBank Stops Making Pepper Robots, Will Cut 165 Robotics Jobs in France. *IEEE Spectrum*, 29/06/2021. <https://spectrum.ieee.org/softbank-stops-making-pepper-robots-will-cut-165-robotics-jobs-in-france>

Arpin-Simonetti, E. (2020). Le grand remplacement robotique n’aura pas lieu : entrevue avec Antonio A. Casilli. *Relations*, (808), 17–22.

Brigand, M. (2015). Pepper, le robot humanoïde qui accompagne les clients de Carrefour. Le

Figaro, Tech & Web, 21/10/2015. <http://www.lefigaro.fr/secteur/high-tech/2015/10/21/32001-20151021ARTFIG00110-pepper-le-robot-humanoide-sommelier-et-conseiller-culinaire.php>

Chodorge, S. (2020). « Trois entreprises françaises travaillent sur un respirateur low-cost ». *L'usine nouvelle*, 10/04/2020 <https://www.usinenouvelle.com/editorial/l-image-du-jour-trois-entreprises-francaises-travaillent-sur-un-respirateur-low-cost.N951976>

Crowe, S. (2018). Social robot maker Jibo sells off assets. *The Robot Report*. 27/11/2021. <https://www.therobotreport.com/jibo-assets-acquired-sqn/>

De Sanctis, J. (2018). Dominique Lestel : « La notion de « robot » est une catégorie-faute-de-mieux ». *Matrice* [www.julien-desanctis.com], 05/07/2018 <https://julien-desanctis.com/dominique-lestel-la-notion-de-robot-est-une-categorie-faute-de-mieux/>

De Sanctis, J. (2019a). Conversation avec Véronique Aubergé : « Le robot social est un selfie de l'homme moderne ». *The Conversation*, 06/10/20 <https://theconversation.com/conversation-avec-veronique-auberge-le-robot-social-est-un-selfie-de-lhomme-moderne-124048>

De Sanctis, J. (2019b). Des robots sociaux servent de « prothèses » à des enfants autistes, entretien avec Sophie Sakka. *Digital Society Forum*, Orange, 16/10/2019 <https://digital-society-forum.orange.com/fr/les-actus/1253-apaiser-l39autisme-en-parlant-34a-travers34-un-robot>

De Sanctis, J. (2019c). « L'IA forte, c'est le décide », entretien avec Dominique Sciamma. *Digital Society Forum*, Orange, 11/12/2019 <https://digital-society-forum.orange.com/fr/les-actus/1271-34l39ia-forte-c39est-le-decide34>

De Sanctis, J. (2020a). « L'avenir de la robotique sociale : assistance ou surveillance ? ». *The Conversation*, 22/01/2020 <https://theconversation.com/lavenir-de-la-robotique-sociale-assistance-ou-surveillance-125962>

De Sanctis, J. (2020b). « Décryptage d'une rhétorique roboticienne au temps du Covid-19 », *Digital Society Forum*, Orange, 23/04/2020 <https://digital-society-forum.orange.com/fr/les-actus/1293-decryptage-dune-rhetorique-roboticienne-au-temps-du-covid-19>

Dzieza, J. (2019). "Beat the machine": Amazon warehouse workers strike to protest inhumane conditions. *The Verge*, 16/07/2019. <https://www.theverge.com/2019/7/16/20696154/amazon-prime-day-2019-strike-warehouse-workers-inhumane-conditions-the-rate-productivity>

Dzieza, J. (2020). How hard will the robots make us work?, *The Verge*, 27/02/2020. <https://www.theverge.com/2020/2/27/21155254/automation-robots-unemployment-jobs-vs-human-google-amazon>

Frankel, Todd. C. (2016). The Cobalt Pipeline. Tracing the path from deadly hand-dug mines in Congo to consumers' phones and laptops. *The Washington Post*, 30/09/2016
<https://www.washingtonpost.com/graphics/business/batteries/congo-cobalt-mining-for-lithium-ion-battery/>

Guizzo, E. (2019). "Boston Dynamics' Spot Robot Dog Goes on Sale", *IEEE Spectrum*, 24/09/2019. <https://spectrum.ieee.org/boston-dynamics-spot-robot-dog-goes-on-sale>

Halftermeyer, N. (2020). « Coronavirus : robots et drones bientôt au cœur de la crise française ? », *Frenchweb.fr*, 18/03/2020 <https://www.frenchweb.fr/coronavirus-robots-et-drones-bientot-au-coeur-de-la-crise-francaise/396203>

Hoffman, G. (2019). Anki, Jibo, and Kuri: What We Can Learn from Social Robots That Didn't Make It. *IEEE Spectrum*, 01/05/2019 <https://spectrum.ieee.org/anki-jibo-and-kuri-what-we-can-learn-from-social-robotics-failures>

JapanGov (2019). A Machine That Does Nothing? Cuddly LOVOT May Be The Most Innovative Japanese Robot Yet. *Forbes*. 06/02/2019.
<https://www.forbes.com/sites/japan/2019/02/06/a-machine-that-does-nothing-cuddly-lovot-may-be-the-most-innovative-japanese-robot-yet/?sh=9220e931fec5>

Lecher, C. (2019). How Amazon automatically tracks and fires warehouse workers for "productivity", *The Verge*, 25/04/2019.
<https://www.theverge.com/2019/4/25/18516004/amazon-warehouse-fulfillment-centers-productivity-firing-terminations>

Lecomte, E. (2014). « Voici Pepper, le robot émotif d'Aldebaran », *Sciences et Avenir* 06/06/2014. https://www.sciencesetavenir.fr/high-tech/robot/voici-pepper-le-robot-émotif-d-aldebaran_92724

Lewis, J. (2005). Robots of Arabia. *Wired*. vol. 13, no. 11, pp. 188–195.
<https://www.wired.com/2005/11/camel/>

Mao, B. (2018). Catherine Maladou : « La plupart du temps on se comporte comme des robots ». *Usbek & Rica*, n°21, p. 76 <https://usbeketrica.com/fr/article/la-plupart-du-temps-on-se-comporte-comme-des-robots>

Sharkey, N. (2018). Mamma mia it's Sophia: A Show Robot or a Dangerous Platform to Misperceive?. *Forbes*, 17/11/2018.
<https://www.forbes.com/sites/noelsharkey/2018/11/17/mama-mia-its-sophia-a-show-robot-or-dangerous-platform-to-misperceive/?sh=60329a4e7ac9>

Stiegler, B. (2011), « "Le grand désenchantement". Un entretien avec le philosophe Bernard Stiegler », *Le Monde*, 21 février.
<https://www.lemonde.fr/blog/fredericioignot/2011/02/21/nous-vivons-un-extreme-desenchantement-un-entretien-avec-le-philosophe-bernard-stiegler/>

Vanderborght, B. (2019). Robotic Dreams, Robotic Realities: Why Is It So Hard to Build Profitable Robot Companies?. *IEEE Spectrum*, 21/03/2019.
<https://spectrum.ieee.org/robotic-dreams-robotic-realities>

Vincent, J. (2016). Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. *The Verge*. 24/03/2016. <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>

Études et rapports

CERNA (2014), *Éthique de la recherche en robotique*, rapport n°1, consulté le 22/04/2020 sur http://cerna-ethics-allistene.org/digitalAssets/54/54000_38704_Avis_robotique_livret.pdf

Grinbaum, A., Devillers, L., Adda, G., Chatila, R., Martin, C., Zolynski, C., & Villata, S. (2021). *Agents conversationnels: Enjeux d'éthique* (Doctoral dissertation, Comité national pilote d'éthique du numérique; CCNE).

Émissions de radio

Sebbah, F-D. (invité). (2016, 16 janvier). « Emmanuel Levinas et les préoccupations de l'heure » [Podcast audio]. *Répliques*. France Culture
<https://www.franceculture.fr/emissions/repliques/emmanuel-levinas-et-les-preoccupations-de-lheure>

Latour, B. (invité). (2022, 21 mars 2022). « Épisode 1/4 : Bruno Latour : "Il ne faut jamais simplifier le réel" » [Podcast audio]. *Les chemins de la philosophie*. France Culture.
<https://www.radiofrance.fr/franceculture/podcasts/les-chemins-de-la-philosophie/les-chemins-de-la-philosophie-du-lundi-21-mars-2022-6989359>

Videos

Boston Dynamics (2015). Introducing Spot Classic (previously Spot) [Vidéo]. Boston Dynamics. YouTube. <https://www.youtube.com/watch?v=M8YjvHYbZ9w>

Boston Dynamics (2016). Atlas, The Next Generation [Vidéo]. Boston Dynamics. YouTube. <https://www.youtube.com/watch?v=rVlhMGQgDkY>

Carr-Brown, D. (2018). L'irrésistible ascension d'Amazon. Allemagne : ARTE/RBB. Pumpernickel Films.

Damiano, L. (2022). Vers l'éthique synthétique. Une proposition de la philosophie des sciences [Vidéo]. Strate, École de design. YouTube.
<https://www.youtube.com/watch?v=yFNwconvDjU>

Le Mouton Numérique. (2018). « Pourquoi nos robots deviennent-ils empathiques ? ». [Vidéo]. Le Mouton Numérique. YouTube. <https://www.youtube.com/watch?v=Kz0Tn5LL-ec&t=4061s>

TEDx Talks. (2019, 26 avril). « Il est temps de passer des logiciels aux émotiels ». [Vidéo]. YouTube. <https://www.youtube.com/watch?v=dYSW9alpQaE>

Liens hypertextes des licences photographiques

arcticpenguin - <https://www.flickr.com/photos/arcticpenguin/>

Héctor García - <https://www.flickr.com/photos/torek/>

Stephen Chin - <https://www.flickr.com/photos/steveonjava/>

Jiuguang Wang - <https://www.flickr.com/people/32366606@N00>

Pierre Metivier - <https://www.flickr.com/photos/feuilllu/>

Paul Allais - <https://www.flickr.com/photos/hokutosuisse/>

Loz Pycok - <https://www.flickr.com/photos/blahflowers/>

Kārlis Dambrāns - <https://www.flickr.com/people/65265630@N03>

Cmglee - <https://commons.wikimedia.org/wiki/User:Cmglee>

Ars Electronica - <https://www.flickr.com/photos/arselectronica/>

Elisfkc - <https://www.flickr.com/people/127662106@N04>

Inhabitat - <https://www.flickr.com/photos/inhabitat/>

