

**INTERACTION N'EST PAS FICTION ?
LE CAS DE LA RENCONTRE HOMME-ROBOT.**

Manuel Rebuschi & Marion Renauld

(Archives Henri-Poincaré, MSH Lorraine, Université de Lorraine)

Résumé : L'interaction humain-robot relève en partie de la simulation, de la part du robot mais également de celle de l'agent humain. Le jeu de faire-semblant à l'œuvre quand je fais comme si le robot était un être sensible doté d'une intériorité évoque une célèbre analyse des fictions par K. Walton. L'interaction humain-robot est-elle fictionnelle ? Se sépare-t-elle radicalement des interactions jugées plus sérieuses, d'humain à humain ? L'article explore cette question en distinguant finalement différents types de faire-semblant, selon que la finalité du jeu de l'interaction lui est interne ou le transcende.

Mots-clefs : Interaction homme-robot · Comportement, Action, Interaction · Fiction · Faire-semblant · Description vs. Interprétation · Intentionnalité

Introduction : une question de faits ?

Dans les interactions humaines, nous projetons des intentions sur les comportements d'autrui dans le but de les comprendre. Par exemple, si une femme me tend un papier, je suppose qu'elle a l'intention de me le donner, ou de me le faire lire, et je peux prendre ou non ce papier, et le prendre de telle et telle manière, en fonction de ma reconnaissance de son intention et de mon envie d'y répondre. Mes suppositions sont justifiables par des conventions partagées, explicites ou implicites selon le contexte et le comportement plus ou moins habituel de la femme, mais aussi par des considérations sur la nature ontologique de cette femme – ici, la reconnaissance d'un être humain, de la même espèce que moi. Que celle-ci soit dotée de conscience, en particulier, je le suppose plus que je ne ferais semblant de le croire. Si cette femme était déguisée en clown, je pourrais peut-être faire-semblant de croire que ce clown a l'intention de me donner un papier, et cette fiction pourrait orienter ma réponse pour construire une interaction semblable à un jeu de rôle, dans lequel moi-même je ferais semblant de remercier le clown. Dans ce cas, je suppose donc moins que je n'imagine une situation fictive.

Mais que se passe-t-il quand c'est un distributeur automatique, par exemple un horodateur, qui me tend un papier ? Cette machine me le délivre plus qu'elle ne me le tend, car je ne lui prête pas d'intention, je ne lui suppose pas d'intériorité et l'effet en est, par exemple, que je ne la remercie pas pour son geste. Et parce que sa réaction semble être automatique, ou parce que rien ne m'incite, en somme, à la traiter autrement que comme un système dépourvu de finalité, je n'ai donc pas l'impression d'interagir avec cette chose. J'interagis moins avec elle (avec ça) qu'avec mon ordinateur, par exemple, qui pourrait me saluer lorsque je l'allume. Ou plutôt, je fais comme si mon ordinateur me saluait, parce qu'en réalité il ne fait qu'afficher sur l'écran les lettres de mon nom, ou peut-être même émettre le son de mon nom, comme n'importe quel autre.

La situation semble passer un seuil lorsqu'il s'agit de mon chien, revenant vers moi en tenant le journal dans sa gueule, et que je pourrais remercier en projetant sur ses gestes une certaine intention de me faire plaisir, une certaine émotion de contentement ou même un sentiment d'amour réciproque. Je suppose alors que son comportement n'est pas purement mécanique, ni seulement pavlovien, mais comporte une dimension intentionnelle, peut-être à tort. On peut ici distinguer une interaction sérieuse d'une autre fictionnelle, en fonction des croyances que j'ai sur la nature ontologique de mon animal : je peux faire-semblant de croire qu'il a l'intention de me faire plaisir en m'apportant le journal, si je ne crois pas que mon chien soit réellement doté de sentiments, ou bien je le remercie sincèrement parce que je crois que mon chien a vraiment des émotions. Je peux aussi hésiter sur la question et projeter sur l'animal des fictions qui me permettent peut-être d'interagir plus efficacement, ou de m'amuser plus pleinement.

Et si mon chien était un robot ? Pouvons-nous vraiment parler d'interaction et suis-je justifiée à projeter des intentions et des émotions sur cette machine ? Et pourquoi suis-je effectivement amenée à faire cela, c'est-à-dire à faire comme si le robot canin avait envie de me faire plaisir, était heureux, ou pas, de porter jusqu'à moi le journal dans sa gueule ? Mais lorsque je fais comme si ce robot canin, comme son homologue vivant, avait effectivement l'intention de me faire plaisir, il semble que cette simulation ne puisse pas être autre chose qu'une feinte (ludique ?), en ce sens que je sais que ce n'est qu'un robot. A l'instar des jeux de faire-semblant avec des poupées ou des soldats de

plomb, au théâtre, devant un film ou à la lecture d'une histoire fictionnelle, ici le jeu de faire-semblant s'impose, en somme, en vertu de ma connaissance de la nature ontologique de ce sur quoi je projette telle ou telle croyance simulée : nous faisons comme si la poupée avait une personnalité, ou comme si l'histoire était vraie, tout en étant sûr que ce n'est pas le cas.

Maintenant, la différence entre les jeux de faire-semblant classiques (susnommés) et ceux qui impliquent un robot semble tenir à la dimension interactive de l'échange entre lui (ça) et moi : à partir du moment où le robot réagit d'une façon qui paraît singulière et circonstancielle, je peux avoir l'impression d'être engagée dans une situation dynamique avec le robot. Et lorsque j'agis dans le cadre de cette interaction, je ne fais pas seulement semblant d'agir. Par exemple, le robot me tend un papier, je prends le papier en estimant que c'est un geste d'offrande, je remercie le robot et le robot penche un peu la tête sur le côté, ou « frétille », j'interprète cela comme un signe de contentement, je peux en venir à le juger cordial, et poursuivre l'échange en lui tendant de nouveau le papier, et le robot pourrait le prendre, etc. Le robot et moi agissons au sein d'un ensemble d'actions qui ne semblent ni mécaniques, ni automatiques, mais appropriées et « réfléchies ». En ce sens, il semble qu'en vertu du sentiment d'interaction, la fiction n'est plus totale, mais limitée et contrainte par l'action en cours.

L'hypothèse serait donc la suivante : *plus il y a interaction, moins il y a fiction* (faire-semblant)¹. Pourtant, je suis sûre d'interagir avec cette machine, mais je ne crois pas réellement que le robot est un être autonome, qu'il possède une conscience et qu'il ressent des émotions – même à avoir quelques doutes quant à son degré d'autonomie. Nos interactions avec les robots ne paraissent donc pas très claires : nous personnalisons la machine, nous lui prêtons des attributs anthropomorphes ou zoomorphes sous forme de croyances feintes, mais en même temps nous participons réellement à un échange – voire à une rencontre – sur le plan comportemental. Nous faisons comme si l'objet manifestait une subjectivité, alors que nous savons que ce n'est pas le cas, ou alors que nous hésitons sur les capacités et la nature ontologique de l'objet (du sujet ?) avec lequel nous

1 Et moins il y a interaction visible, plus il y a sentiment de jeu (c'est le cas des drones qui, commandés à distance, peuvent tuer des hommes en un geste que l'individu aux manettes considère comme lointain, presque fictionnel, parce que désincarné).

interagissons ; mais dans les deux cas, nous pouvons en partie interagir sérieusement et effectivement avec la machine.

Le but de cet article est de clarifier les rapports entre interaction et simulation, et plus précisément la nature et les présupposés (croyances, motivations) de tels jeux de faire-semblant interactionnels. La fiction s'arrête-t-elle là où commence l'interaction ? Ou plutôt, l'interaction commence-t-elle là où commence la projection d'intentions et d'émotions ? Le faire-comme-si s'étirole-t-il quand s'ouvre le faire-avec, ou le faire-avec suppose-t-il de faire-comme-si ? Plus largement, si toute interaction suppose une dimension interprétative (des intentions en plus des descriptions), est-ce que toute interprétation comporte une part fictionnelle (en plus de la part rationnelle) ? Dans ce cas, nos étranges interactions avec les robots ne seraient qu'un cas particulier de l'interaction humaine, une façon de montrer non seulement la dualité intrinsèque (interpréter/décrire) des comportements en général, mais également la dimension fictionnelle des interprétations engagées dans une interaction.

Précisons que le développement de cette réflexion prend appui sur un certain nombre d'expériences menées par un groupe de recherche interdisciplinaire et mettant en jeu une lampe robotisée face à des sujets. Partant de ce qu'en disent ces derniers, à savoir globalement qu'ils ont affaire à une interaction réelle, il s'agit alors d'analyser le contenu et le statut des hypothèses que ceux-ci entretiennent à propos de ce dispositif lorsqu'ils estiment interagir avec. Quel est le degré d'adhésion – feinte ou authentique – des sujets à cette croyance, notamment, d'une forme de subjectivité reconnue à la lampe ? Les deux options rivales d'une nouvelle forme d'animisme ou d'une pure simulation sont ainsi mises en défaut et dépassées en faveur d'une étude plus approfondie de différents types de faire-semblant et d'interaction selon leurs finalités – bien plutôt que selon le type d'objet en cause.

1. Les jeux de faire-semblant « classiques » : une frontière étanche entre fiction & action.

L'objectif de cette première section est de poser le cadre conceptuel qui nous permettra d'appréhender l'interaction avec un dispositif robotisé. Il s'agit notamment de fixer cer-

taines notions telles que la fiction, la simulation, mais aussi simplement l'interaction, qui sont mobilisées dans ce type de contexte. Nous ne partons évidemment pas de rien. L'approche de Walton (1990) propose une analyse des fictions en termes de simulation ou de jeu de faire-semblant, et c'est elle qui constitue ici notre point de départ². On s'intéresse ensuite à la distinction entre action et comportement, et finalement entre différents types d'interactions.

1.1. Fiction et non-fiction

Lorsqu'un auteur produit un récit fictionnel, il fait comme s'il racontait une histoire vraie, et quand nous lisons le récit fictionnel, nous faisons comme s'il était vrai. Mieux : l'auteur fait comme s'il était le narrateur, parfois le témoin de telle et telle scène, et nous, lecteurs, participons à son jeu de rôle en faisant comme s'il fallait le prendre au mot. Ce jeu du *comme si* ou du faire-semblant s'étend aux acteurs, qui jouent leur rôle *sérieusement*, se font passer pour un autre, adoptent ses croyances et désirs, parlent et se comportent comme s'ils étaient cet autre. La caractéristique de ces jeux de faire-semblant est qu'ils reposent sur une suspension de la croyance et du doute : en participant, comme auteur, acteur ou lecteur-spectateur à une fiction, il est entendu que l'on est dans un jeu d'imitation (*mimesis*) auquel il n'est pas question d'adhérer sans recul, et qu'il ne s'agit pas d'évaluer sérieusement.

Le jeu de l'imitation est au cœur de la conception de Walton (1990), qui s'inspire des jeux d'enfants où l'on s'improvise, par exemple, chasseurs d'ours, et où différents éléments de l'environnement sont mobilisés comme s'ils étaient autres qu'eux-mêmes. Walton évoque cette situation tout à fait ordinaire où des souches d'arbres dans la forêt sont utilisées comme si elles étaient des ours : il faut alors s'en méfier, ne pas les approcher, en bref adopter le comportement qui serait approprié si ces souches étaient des ours. Ce jeu transforme ainsi certains objets matériels en signes : la souche n'est plus

2 Nous privilégions ici l'analyse waltonienne notamment en vertu de sa systématisme, mais surtout parce qu'elle n'est pas focalisée sur la fiction linguistique et s'intéresse en particulier à la participation des joueurs au sein de jeux plus ou moins autorisés de faire-semblant. Il serait possible de raffiner les problèmes que posent la définition de la fiction. cf. Renauld (2014)

une simple souche se suffisant à elle-même, mais, grâce au jeu, sert à désigner autre chose, sa présence devenant ici le signe de la présence d'un ours.

Le jeu des enfants avec les souches n'est plus alors à considérer comme une simple activité physique, où l'on courrait et grimperait aux arbres *pour rien*. Si on agit de la sorte, c'est parce que les joueurs et leur environnement ont pris *un sens* particulier au jeu. On court et on grimpe aux arbres *pour* échapper au danger, ou *pour* attaquer ensemble un ours. Chaque joueur et chaque élément de l'environnement est susceptible³ de devenir signe et de dénoter autre chose que lui-même, jusqu'à ce coin de forêt en réalité dédié au jeu qui devient, fictionnellement, une forêt pour chasser les ours. Le jeu peut être vu comme plaçant un *calque* sur ce coin de forêt, qui conduit les participants à saisir des signes là où à l'origine il n'y en avait probablement aucun, ou des signes d'autres choses n'ayant rien à voir avec le jeu (comme un panneau interdisant de faire du feu).

Si le jeu de faire-semblant produit des signes, il faut alors, en tant que joueur, les comprendre. La convention qui veut que les souches représentent des ours est relativement simple, et les joueurs vont très rapidement comprendre leur environnement de cette manière. Mais il peut y avoir des cas-limites où la compréhension n'est pas immédiate. Ainsi, telle grosse racine qu'un des joueurs prend pour un ours tandis que d'autres le refusent : on est alors conduit à un conflit d'interprétations. L'interprétation est solidaire de tout jeu de faire-semblant dès lors qu'il devient un peu subtil, que ce soit par des cas-limites comme celui-ci, ou plus généralement quand on a affaire à une œuvre fictionnelle complexe, dont le sens et la portée ne peuvent pas être saisis immédiatement.

Les jeux de faire-semblant sont généralement démarqués des actions et attitudes authentiques par une frontière étanche. La fiction repose en effet classiquement sur un cadre conventionnellement délimité et reconnu : un roman ou un film de fiction ne sont pas vendus comme des reportages, et une pièce de théâtre n'est clairement pas la présentation d'interactions réelles sur une scène. La répartition des tâches est également conventionnellement fixée. Pour les auteurs/acteurs : on joue tel rôle ; pour les lecteurs/

3 Tout ne devient pas signe, et même la plupart des objets de l'environnement sont inchangés : tel arbre pourra rester simplement un arbre, qu'un prétendu chasseur pourra utiliser en tant qu'arbre pour se cacher.

spectateurs : on fait semblant d’y croire. La séparation entre rôles actifs et passifs est généralement bien délimitée aussi : si le spectateur est mobilisé dans une pièce participative, c’est qu’il prend également un rôle d’acteur. Dans le jeu de l’imitation des enfants ou dans les jeux de rôles pour adultes, tout le monde participe activement à la fiction. On a alors affaire à des *interactions fictionnelles*⁴.

Dans la suite de la section, nous nous penchons sur les définitions de l’action et de l’interaction avant de distinguer interaction réelle et interaction fictionnelle.

1.2. Du comportement à l’action

La distinction entre comportement et action est classique en philosophie. Pour en donner une première idée, on peut dire qu’une action, c’est un comportement avec de l’intentionnalité. L’intentionnalité n’est pas à comprendre ici au sens ordinaire (« cet homicide a été commis de façon intentionnelle ») qui renvoie à un but fixé par l’agent, mais au sens technique utilisé en philosophie : est intentionnel ce qui est à *propos* de quelque chose, et en général à propos de quelque chose d’autre que soi. Les souvenirs, les désirs, mais aussi les portraits, les films et les mots sont réputés intentionnels, tandis que les maux de tête, les pierres, les voitures, ne le sont pas. L’intentionnalité au sens ordinaire est ainsi un cas particulier de l’intentionnalité qui nous intéresse. Une *action* est un comportement intentionnel dans ce sens élargi : un comportement qui n’est pas un simple comportement sans finalité ni sens, mécanique ou automatique, mais qui est à *propos* de quelque chose.

Cette caractérisation se veut neutre relativement aux conceptions divergentes défendues sur l’intentionnalité. Que l’on soit réaliste ou antiréaliste, réductionniste (« naturaliste ») ou non concernant l’intentionnalité, on peut maintenir que l’action est intentionnelle par définition. Le rôle de l’intentionnalité dans l’action, antérieur et causal, ou contemporain et rationnel, dépend étroitement de la conception choisie de l’intentionnalité. Nous ne voulons pas ici adopter une conception particulière mais examiner ce qui distingue, en pratique, l’action du comportement.

4 Walton admet qu’il existe des cas plus compliqués où la frontière entre le monde réel et le monde du jeu semble plus poreuse, mais l’effet en est toujours une extension du domaine de la fiction, et non un patchwork ou une modification du fictionnel en réel. cf. Walton (1978).

Le comportement est comme la pierre, la voiture ou le mal de tête : il est, mais il n'est à propos de rien. Pour reprendre un exemple célèbre de Wittgenstein : quand *mon bras se lève*, c'est un simple comportement, qui peut être décrit de manière behavioriste ou physicaliste, éventuellement avec un peu d'anatomie et de biochimie. Le fait que mon bras se lève est un événement du monde physique qui affecte mon corps, dont on peut en principe donner une explication en termes de causes également inscrites dans le monde physique, et qui peut produire des effets dans le monde physique. Par contraste, si on comprend *je lève le bras* comme une action, c'est que mon comportement a alors un auteur – un sujet, un agent : *je* – et qu'il est alors à propos de quelque chose : je lève le bras pour attraper une bouteille, demander la parole, dire que j'ai mal... ou pour lever le bras. Mon action n'est pas seulement un événement qui affecte mon corps, elle a un ou plusieurs *sens* qu'il s'agit de comprendre, éventuellement d'interpréter, relativement à un contexte. Expliquer le comportement exige d'en trouver les *causes*, comprendre l'action, donc son sens, exige d'en exprimer des *raisons*.

La distinction entre action et comportement se manifeste en premier lieu dans le discours sur l'action et le comportement. Notre discours (notre vocabulaire, nos questions...) porte-t-il sur les aspects physiques, physiologiques, réflexes, neuronaux, sur les causes et les effets ? Il s'intéresse alors au comportement. Porte-t-il sur les aspects mentaux, psychologiques, sociaux, sur le sens et les enjeux ? Il est alors concerné par l'action. On a ainsi deux modes de discours qui nous permettent de délimiter les deux domaines. Y a-t-il des *faits* (intentionnels) qui nous obligent à adopter le mode de discours intentionnel, c.-à-d. à parler en termes d'actions ? À l'inverse, l'occurrence d'une action requiert-elle le mode de discours intentionnel, donc le langage ? Répondre à ces questions suppose d'adopter une conception précise de l'intentionnalité. Pour ce qui nous intéresse, nous nous bornerons à considérer les deux modes de discours comme les indicateurs de la distinction entre action et comportement. La distinction est ici comprise *a minima*, comme une dualité conceptuelle.

1.3. Différents types d'« interaction »

La dualité du comportement et de l'action se retrouve lorsqu'on a affaire à plusieurs comportements simultanés et corrélés. Notre vocabulaire est cependant limité : la cooccurrence de comportements corrélés est généralement simplement qualifiée d'*interaction*. Afin de marquer la dualité de l'action et du comportement dans le cas de l'interaction, on emploiera la graphie *Interaction* avec une majuscule pour les cas intentionnels, et *interaction* sans majuscule pour les cas indifférenciés qui ne présupposent pas d'intentionnalité. De même qu'une action est un comportement intentionnel, une *Interaction* est donc une interaction intentionnelle.

L'interaction verbale est toujours une *Interaction*. En effet, l'usage d'une langue naturelle présuppose l'intentionnalité : si nous échangeons des mots et des phrases, c'est parce qu'ils ont un sens, et les locuteurs utilisent ce sens en contexte pour signifier quelque chose à leurs interlocuteurs. L'interaction verbale est ainsi intentionnelle de bout en bout. Si nous interagissons avec une machine programmée pour poser des questions et nous satisfaire de certaines réponses comme un distributeur de billets de train, ce n'est pas la machine elle-même qui manipule la langue et le sens, mais ce sont ses concepteurs et propriétaires qui l'utilisent pour cela : l'interaction homme-machine est *in fine* une *Interaction* homme-homme médiatisée par la machine.

Tournons-nous vers les interactions non verbales, qui ne mobilisent pas de langue naturelle ou artificielle, c'est-à-dire pas de langage symbolique. Ces interactions peuvent mobiliser d'autres signes : les signes de « langages du corps » s'il en existe qui soient non symboliques, ou d'autres signes naturels. Les signes naturels sont les supports de formes minimales d'intentionnalité : les cernes indiquent l'âge d'un arbre, la fumée indique le feu. Pour les partisans de la naturalisation comme Dretske, ces signes offrent une base à partir de laquelle il faut reconstruire l'intentionnalité dans son ensemble. Ici nous nous contentons de les distinguer des symboles qui sont des signes arbitrairement sélectionnés pour indiquer ce dont ils sont le signe. Les animaux produisent des signes naturels qui peuvent *exprimer* certains états internes : les oreilles en arrière et le poil hérissé d'un chien sont un signe de colère ou d'agressivité, tel cri est un signe de douleur. Certains animaux peuvent comprendre ces signes naturels dans l'interaction.

D'autres, au contact de l'homme, peuvent également comprendre certains symboles, mais il s'agit d'un cas particulier qui les fait entrer dans une forme d'interaction verbale.

L'interaction peut également constituer elle-même un signe. Ainsi en va-t-il des actions conjointes ou coordonnées, comme la danse de salon, le jeu au ballon ou la promenade familiale dans les bois. Qu'elle soit verbale ou non, l'Interaction est ainsi caractérisée par le fait qu'elle mobilise des signes, qu'elle est porteuse de sens pour les agents qui interagissent comme pour des observateurs extérieurs. L'Interaction suppose par conséquent la compréhension des signes de l'interaction. L'intercompréhension repose généralement sur un jeu de langage partagé, donc sur une forme de vie commune ; elle mobilise aussi l'empathie, la projection de soi dans l'autre ou réciproquement.

Quand la compréhension n'est pas immédiate, il faut interpréter. Cela se produit quand le sens de l'action d'autrui n'est pas saisissable parce qu'on manque d'un jeu de langage partagé, comme avec quelqu'un qui maîtrise mal notre langue naturelle, ou avec un danseur contemporain dont on ne maîtrise pas les codes. Mais cela peut être aussi parce que l'action est complexe et qu'on ne la maîtrise pas totalement, ou parce qu'on a affaire à un comportement verbal extraordinaire qui fait que l'interaction déraile.

1.4. De l'interaction sérieuse à l'interaction simulée

Comme on l'a indiqué plus haut, la frontière entre fiction et non-fiction est généralement réputée étanche. À première vue, c'est également le cas pour l'Interaction, qui est sérieuse ou simulée, de façon claire et nette. Soit on Interagit authentiquement avec autrui, soit elle et moi jouons des rôles et ce sont ces rôles qui interagissent. Nous y sommes ou nous faisons semblant. Il y a bien entendu le cas hybride de la duperie, où l'un joue un rôle sans que l'autre le sache, où l'Interaction est simulée d'un côté et sérieuse de l'autre, le menteur jouant avec la disparité des cadres de référence (Goffman 1974). Ce type d'Interaction se comprend aisément à partir des deux types fondamentaux, sans constituer en tant que tel une énigme.

Les jeux collectifs de faire-semblant consistent à réinterpréter les signes de l'Interaction en les dotant d'un sens fictionnel. Nous pouvons danser réellement un tango, et il s'agit d'une Interaction véritable ; nous pouvons danser un tango dans une représenta-

tion de théâtre, et il ne s'agit plus d'une danse véritable mais d'un tango simulé. Ce qui distingue *in fine* les deux Interactions ne se situe pas sur le plan de l'interaction, qui est identique, mais sur celui du sens qui lui est donné.

2. Les projections-personnifications : ou comment s'émouvoir d'une lampe.

Dans cette section nous présentons un ensemble d'expériences conduites dans un cadre interdisciplinaire et mettant aux prises des sujets humains avec une lampe robotisée. L'idée peut sembler étrange dans un article à vocation philosophique, d'autant qu'il ne s'agit pas ici de philosophie expérimentale où les intuitions de philosophes sont confrontées à celles du commun. Cela peut cependant être justifié ainsi : les interactions homme-robot étendent le champ des interactions et expériences humaines en général ; à ce titre, elles constituent l'occasion de tester notre vocabulaire et notre cadre conceptuel. Faut-il les étendre pour les adapter ? Faut-il en changer ? À l'opposé d'expériences de pensée où nous repoussons les limites du possible pour raisonner *a priori* sur nos concepts, les expériences qui nous concernent ici donnent lieu à la production de comportements et de locutions intentionnelles qui offrent un éclairage utile sur la robustesse de nos conceptions d'autrui. Quand j'interagis avec une lampe, est-ce que je fais semblant d'*Interagir* ? Telle est la question.

2.1. Appréhender l'interaction non-verbale

Le groupe PsyPhINe (pour psychologie, philosophie, informatique et neurosciences) s'est constitué au début des années 2010 à l'université de Lorraine par la réunion de chercheurs travaillant dans les disciplines citées, auxquelles se sont ajoutées d'autres spécialités : robotique, sociolinguistique et anthropologie. Le projet initial du groupe était de confronter les approches disciplinaires de notions comme l'intelligence, l'intentionnalité ou la cognition, en se focalisant principalement sur des interactions non verbales entre l'homme et la machine. À partir de 2015, le groupe a défini et mis en œuvre plusieurs protocoles expérimentaux qui ont permis d'observer des dizaines de sujets en interaction avec une lampe robotisée. Les sujets étaient invités à réaliser une tâche à l'aide de jouets en bois et le dispositif expérimental était construit de telle sorte qu'une

interaction puisse être instaurée avec une lampe robotisée posée en face d’eux – ou du moins, de telle sorte qu’on puisse parler des comportements corrélés de la lampe et du sujet comme d’une interaction.

À rebours de nombreuses recherches qui se concentrent sur les robots anthropomorphes, l’objectif du groupe est d’observer la manière dont une interaction peut être établie entre un humain et un objet non-anthropomorphe et ouvertement artificiel, même s’il est clairement zoomorphe⁵. L’idée est également de se concentrer sur l’interaction non-verbale, étant entendu que l’interaction verbale, comme il en existe avec de nombreuses machines aujourd’hui, présuppose une forme d’intentionnalité. S’il est avéré qu’une interaction n’émerge pas à n’importe quelle condition⁶, ce que le groupe a pu vérifier après quelques échecs, l’interaction, quand elle est manifeste, soulève de multiples questions. Est-elle objectivement mesurable ? Dépasse-t-elle l’interaction quotidienne avec une machine à café ? Peut-elle être qualifiée d’Interaction au sens où nous l’avons défini plus haut, à savoir, être considérée comme intentionnelle ?

Plusieurs facteurs ont été relevés comme pouvant être pertinents pour mesurer la qualité de l’interaction : le regard du sujet, indifférent à la lampe ou à l’inverse se concentrant sur elle, voire sur les zones pointées par la lampe quand celle-ci était supposée l’aider, ce qui ressemble à une sorte d’attention conjointe ; les « tours de parole », respectés ou non, c’est-à-dire l’alternance des gestes de l’un et de l’autre avec l’attente d’une réaction une fois son propre geste effectué ; et, dans une moindre mesure, l’expression d’émotions par le sujet, en réaction au comportement de la lampe, expression qui n’est bien entendu pas nécessairement destinée à la lampe mais qui peut l’être.

Ce qui est toutefois du plus grand intérêt pour notre étude, c’est la manière dont on parle de ce qui se passe ; plus précisément, ce que disent les sujets à propos de leur interaction avec la lampe, comment ils la qualifient au travers de questionnaires mais aussi au cours d’entretiens libres, et ce qu’en disent aussi les observateurs, qu’ils soient impliqués dans la recherche ou observateurs extérieurs. Car s’il fait consensus que la

5 La lampe a un corps et une tête, donc la structure d’un animal ; elle est dotée de « comportements », qui sont des mouvements coordonnés du corps et de la tête, ce qui renforce l’analogie.

6 Nous reviendrons par la suite sur cette question de la nécessité ou non de conditions ontologiques à l’interaction.

lampe *n'est pas* un être vivant avec lequel le sujet Interagit véritablement, les questions de la nature de l'interaction, de l'octroi ou non d'une intériorité à la lampe, offrent quant à elles matière à débat.

2.2. Quelques faits d'expériences

Plusieurs éléments sont ressortis de ces expériences. Tout d'abord, une convergence sur l'occurrence d'une interaction. Dans l'un des protocoles intitulé Kapl'IA, conduit en 2016 avec 43 sujets, ces derniers avaient pour tâche de réaliser une construction libre en Kapla (jeu de bois) avec pour objectif d'allumer la lampe robotisée. L'objectif étant sans relation directe avec la tâche à réaliser, le sujet était implicitement invité à être attentif aux mouvements de la lampe, voire à entrer en interaction avec elle. De fait, plus de 80% des sujets ont affirmé penser que la lampe réagissait à leurs actions et qu'il y avait eu interaction (Gendre & Schneider 2016).

Or les sujets étaient divisés en trois groupes, chacun des groupes étant soumis à une condition distincte : deux des groupes avaient en face d'eux une lampe pilotée par un humain, qui visionnait la scène par l'intermédiaire d'une caméra et pouvait ainsi adapter ses actions, tandis que les membres du troisième groupe voyaient se dérouler devant eux une succession de mouvements entièrement prédéfinie, donc totalement indépendante de leur action. Il est notable que les membres du dernier groupe n'ont pas été proportionnellement moins nombreux à considérer que la lampe ajustait son comportement au leur, alors que ce n'était pas le cas : il s'agit d'une manifestation du biais d'agentivité. Ce constat est renforcé par l'observation que ceux, parmi les sujets, qui humanisent la machine (par la mention d'émotions, la comparaison explicite avec un être humain, etc. au cours de l'entretien qui suivait l'expérience) ont tous considéré que la machine réagissait à leur action.

La qualification des comportements de la lampe a été l'objet d'une autre expérience, intitulée Qualcom (Becker *et al.* 2019). Les participants étaient confrontés à la lampe robotisée hors de toute situation interactionnelle et informés de sa manipulation par un humain. Si les attributions (hors contexte) n'ont généralement pas permis de dégager des consensus ni même des accords intersubjectifs probants sur la signification des

comportements de la lampe, les participants ont néanmoins très souvent employé à son sujet des termes relevant du registre des émotions (« surprise », « peur », « curiosité »...) ou lui ont attribué des capacités d'attention et de perception. Dans les entretiens enregistrés, de nombreux participants ont assimilé la lampe à un animal, ou invoqué des capacités d'agir spécifiques aux animaux, voire à l'humain, quand ils ne l'ont pas traitée comme partenaire dans une interaction, tandis que d'autres n'ont pas cessé de la considérer comme une simple machine ou un robot.

Le statut de la lampe est également en question. S'agit-il d'un dispositif autonome ou d'un objet dont les mouvements sont pilotés à distance ? Les avis sont généralement partagés, sauf dans une expérience qualitative (avec peu de sujets) intitulée *Decide*, où tout le monde a tranché en faveur de la première hypothèse (André & Boniface 2018). La question est certainement décisive pour évaluer la nature de l'interaction : interaction avec un artefact dans le premier cas, interaction avec un être humain (donc Interaction) médiatisée par cet artefact dans le second. L'indécision relativement à cette question, relevée dans plusieurs expériences, semble n'avoir pourtant pas interféré avec la manière dont les sujets sont entrés (ou pas) en interaction avec la lampe. Et quand les participants ont tous jugé qu'ils avaient affaire à un artefact comme dans *Decide*, cela ne les a absolument pas empêchés de parler de la lampe comme d'un animal, de lui attribuer des émotions et des intentions.

2.3. Quelle valeur attribuer à ce discours ?

Les différentes expériences conduites par le groupe PsyPhINe ont produit des situations où des sujets ont simultanément cru qu'ils interagissaient avec un dispositif robotisé autonome et attribué des émotions et intentions à ce dispositif. Que signifie ce décalage ? Devons-nous considérer que les sujets font *comme s'ils* interagissaient, tout en interagissant simplement avec la lampe ? S'agit-il d'Interaction purement fictionnelle ?

Une hypothèse formulée par plusieurs anthropologues (Vidal 2012) est que de telles situations conduisent à une nouvelle forme d'animisme étendu aux machines. Cela résoudrait en effet la tension mais au prix d'une attribution de croyances que nous pensons erronée, et qui ne coïncide pas avec ce que les sujets disent eux-mêmes des ma-

chines. Généralement, en Europe occidentale dans la seconde moitié des années 2010, les personnes adultes *ne croient pas* que les robots sont des êtres dotés d'une intériorité à l'égal des humains ni même des animaux. Si l'animisme est une croyance, on ne peut pas considérer qu'elle soit partagée par les sujets rencontrés au cours des expériences.

Il faut par conséquent envisager une autre hypothèse. Les sujets qui parlent de la lampe dans un entretien juste après l'expérience ne sont plus dans la situation expérimentale à faire comme s'ils interagissaient avec un être vivant. Après coup, ils sont en face d'un interlocuteur avec qui ils évoquent le plus sérieusement du monde, dans le cadre d'une recherche scientifique, ce qu'ils viennent de vivre. Pourquoi le vocabulaire employé est-il plus propre à un être vivant qu'à une machine ? Le *comme si* de l'interaction s'étend-il incidemment après le moment de l'expérience ? Ou aurions-nous seulement affaire à des formes inoffensives de métaphores, pour les cas d'attribution d'intériorité, contrastant avec des croyances littérales concernant l'orientation téléologique d'un dispositif programmé ?

3. Les jeux de faire-semblant ambigus : frontière poreuse entre fiction et interaction.

Dans cette section, nous questionnons la nature des rapports entre un humain et un dispositif robotisé, rapports qui paraissent relever d'une dualité paradoxale, ou d'une forme d'inconséquence rationnelle : nous savons que les robots ne sont pas des êtres vivants et pourtant cela ne nous empêche pas d'interagir avec eux d'une façon comparable à notre comportement vis-à-vis d'un être vivant. Plus étonnant encore, c'est peut-être précisément parce que nous tenons compte de la dimension artificielle de ces créations que nous en venons à tisser une relation émotionnellement forte avec eux ! Ce type de phénomène s'apparente à ce qui a lieu dans nos rapports avec les êtres de fiction, des jouets pour enfants aux personnages de romans, de films ou de jeux-vidéo : nous sommes manifestement engagés dans des projections d'intentions et de sentiments, malgré la conscience de la non-existence des supposés agents⁷. La question ici se complique du fait de la réactivité des dispositifs robotiques : si nous ne pouvons pas vrai-

⁷ C'est là le cœur du problème philosophique posé par le « paradoxe de la fiction », à propos duquel on se demande comment il est possible de ressentir des émotions envers des personnages dont on sait pourtant qu'ils n'existent pas. cf. Radford (1975).

ment dire que nous interagissons avec Cendrillon en écoutant son histoire, ou en animant soi-même une poupée inerte, une machine Cendrillon, capable de nous répondre, de se déplacer ou encore de présenter des mimiques sur un visage articulé, semble moins fictive, plus autonome et donc plus à même d'être traitée comme un être intentionnel. Quels sont les liens entre fiction, interaction et Interaction ? La fictionnalité d'une projection s'arrête-t-elle là où commence l'Interaction, ou celle-ci se nourrit-elle toujours davantage du jeu de faire comme si ?

3.1. Une interaction manifeste mais asymétrique

On peut partir de l'idée que nous avons affaire à une interaction entre humain et machine robotisée dès lors que nous constatons des corrélations dans leurs comportements. Mais il faut se demander si la cooccurrence de ces derniers est fortuite, mécanique ou intentionnelle, pour savoir si nous pouvons en outre parler d'Interaction. Il semble même discutabile de parler des actions d'une machine, et non seulement de ses comportements, puisque cela suppose déjà de leur prêter une forme d'intentionnalité. À tout le moins, il est probable ici que nous devons assumer une certaine gradation entre divers objets quand il s'agit d'opérer ce type de distinction : du soldat de plomb, un objet inerte, aux robots actuels capables de converser, en passant par les figures et figurines de jeux vidéo, une description en termes de causes peut apparaître de moins en moins éclairante et appeler une description en termes de raisons, quand bien même cela s'inscrit dans le cadre général d'un faire comme si partagé.

S'il y a bien interaction manifeste entre un humain et un dispositif robotisé, celle-ci reste cependant asymétrique. En effet, le point de vue de l'homme face à la machine est différent de celui de la machine face à l'homme : ce n'est jamais la machine qui projette des intentions, des émotions ou des scénarii interprétatifs sur les faits et gestes de l'humain, mais toujours l'inverse⁸. Il s'agit même pour un programme de redéfinir les actions humaines en termes de comportements afin de les intégrer au système robotisé comme des données lisibles, recevables. En ce sens, il y aurait, d'un côté, interaction robotique, et de l'autre, Interaction humaine. Une question subsidiaire consiste à se de-

⁸ Si certaines machines sont programmées pour « reconnaître les émotions » des sujets humains, les projections d'intentions sont alors dérivées de celles des ingénieurs.

mander comment décrire le face-à-face entre deux machines : nous dirions probablement que, de leur point de vue, ce n'est qu'une suite de réactions conjointes, tandis que de notre point de vue, il pourrait s'en dégager l'impression d'un échange humain⁹.

Et c'est effectivement au niveau des impressions que nous avons que quelque chose se joue de l'ambiguïté de nos rapports avec les machines robotisées. Dit autrement, c'est le sentiment d'une agentivité de la lampe, par exemple, qui nous pousse à lui prêter une intériorité, une personnalité, un pouvoir de décision. Ce sentiment semble issu de l'observation de cette corrélation de comportements qui soutient l'interaction, voire du caractère aléatoire et mystérieux, disons imprévisible, de ses réactions, qui motiverait l'idée d'une véritable Interaction. Mais comment s'ordonnent justement ce qu'on peut appeler les faits interactifs et les fictions interprétatives ?

3.2. Une Interaction simulée mais engagée

Tentons de décrire la façon dont les choses peuvent s'être passées dans le cas des expériences Kapl'IA. Le sujet commence à construire une figure avec les bâtons de bois, et la lampe, à s'animer. Le sujet se demande s'il y a un lien entre ce qu'il fait et ce que fait la lampe. En admettant qu'il y en ait un, le sujet prête de plus en plus d'attention à la lampe, jusqu'à certaines intentions pour expliquer ou simplement qualifier son comportement. Dans ce scénario, c'est parce que la lampe semble interactive que le sujet se voit motivé pour faire comme s'il avait affaire à une personne ou un être vivant. Mais on peut aussi imaginer le contraire : le sujet est déjà en train de s'émouvoir de la lampe, et même de s'adresser à elle, avant même d'avoir constaté quelque causalité entre son action et les gesticulations de celle-ci. Dans ce cas, c'est parce que le sujet a décidé d'emblée d'entrer dans un jeu de faire-semblant avec un dispositif robotisé – voire avec un quelconque objet – que l'interaction démarre et s'enrichit.

Il n'est pas facile de trancher entre ces deux lectures. D'après les expériences, les sujets peuvent avoir l'impression d'interagir avec la lampe même quand celle-ci effectue en réalité une succession de mouvements entièrement prédéfinie, donc totalement indépendante de leur action, ce qui tend à soutenir la seconde description. En même temps,

⁹ Voir exemple les discussions entre robots, « Les discussions incroyables de deux enceintes intelligentes », par G. Brandy, *Slate*, 7 janvier 2017 (<https://www.slate.fr/story/133691/google-home-discussion>).

si la lampe ne faisait rien qui n'attire ni n'intrigue le sujet, ce dernier n'aurait alors probablement aucune raison de la personnifier, et il semble que ce soit parce que les robots sont aujourd'hui de plus en plus interactifs que les humains se montrent de plus en plus portés à ressentir des émotions à leur égard.

Peut-être faut-il encore distinguer ici deux moments du rapport humain/machine pour préciser le type de fiction en jeu : pendant l'interaction se développe une Interaction simulée mais engagée, tandis qu'après l'interaction, pendant la restitution verbale de celle-ci, l'engagement semble diminuer même si la simulation persiste au niveau du discours. Dit autrement, à l'instar de ce qui a lieu quand nous lisons un roman ou regardons un film, nous sommes plongés dans la fiction lorsque nous interagissons avec la machine, nous sentant ainsi co-acteurs d'un commun jeu de faire-semblant, alors que nous ne faisons qu'user d'un vocabulaire – herméneutique – lorsque nous décrivons notre expérience *a posteriori*. La personnification n'a pas la même portée : dans un cas, elle reflète l'ambiguïté de nos actions, dans l'autre, elle est purement pratique. On peut alors se demander ce qui nous motive à nous engager dans un jeu de faire-semblant, ou à en sortir, en un mot ce qui semble altérer ou renforcer la conscience que nous avons du caractère artificiel du dispositif robotique.

3.3. Une fictionnalisation limitée mais utile

L'originalité des expériences menées par PsyPhINe a notamment consisté à éviter de faire entrer la dimension anthropomorphique du dispositif comme une condition de possibilité de l'Interaction homme/machine. Cependant, il est clair que la ressemblance de la lampe avec un animal, ou un visage, n'a pas ôté toute légitimité à cette hypothèse – révélée dans le discours des sujets à propos de leur expérience. On ne peut pas se défaire si facilement de la force de l'illusion mimétique, fonctionnant presque comme une cause mécanique du processus de personnification à l'œuvre dans nos Interactions avec des objets inertes ou des êtres inexistantes : tout se passe comme si on ne pouvait pas s'empêcher de trouver vivant tout ce qui, de près ou de loin, s'apparente à des formes humaines ou animales (ajoutez deux yeux à n'importe quel support et vous aurez soudain tendance à en faire un sujet).

Toutefois, il n'est pas certain que la ressemblance mimétique renforce toujours le sentiment d'une Interaction avec un être doué d'intériorité. En effet, pour peu que la marionnette, le robot ou la machine ressemble mal à son avatar réel, la feinte semble alors échouer et travailler au contraire à accentuer l'artifice jusqu'à rendre l'interaction maladroite, voire impossible (Mori 1970). En fait, quelque chose se joue davantage dans la dimension cinétique du dispositif : le fait d'être en mouvement, et les types de mouvements opérés par la machine donnent du crédit, ou non, à la fictionnalisation en cours¹⁰. En tout cas, si la feinte est ici plus ou moins totale et réussie, il semble qu'elle ne soit que le résultat d'une perception pour ainsi dire non décidée, d'un trompe-l'esprit causalement produit par l'ingéniosité humaine des fabricants. Cela n'explique pas, notamment, la raison pour laquelle nous prêtons des intentions et des émotions à une lampe même après la fin de l'interaction réelle, c'est-à-dire en-dehors d'une illusion vécue.

L'hypothèse la plus plausible est ainsi de considérer que nous nous engageons dans un jeu de faire-semblant, non pas tant en vertu de motifs plastiques, mais en vertu de motifs cognitifs. Personnifier la lampe, conjecturer des intentions cachées derrière ses actions ou des émotions derrière ses maladresses, ce serait chercher à faire sens, à interpréter pour pouvoir agir, réagir et interagir. En effet, s'il ne semble pas nécessaire de faire comme si la machine à café était un être efficace et besogneux, ou serviable et poli, pour comprendre comment la faire fonctionner, ou comment elle fonctionne, cela paraît différent dans le cas de la lampe. De deux choses l'une : ou la lampe est un objet qui dysfonctionne, et il n'y a donc rien à comprendre, mais seulement quelque chose à réparer, ou bien il s'agit d'un dispositif interactif, et alors il s'agit d'interpréter les signes de ses comportements pour tisser un certain rapport rationnel entre ses gestes et les nôtres. Dans cette perspective, la fiction fait partie des outils de l'interprète qui, selon un principe de charité (Quine 1960) élargi aux êtres qui n'existent pas, tente de projeter une certaine forme de rationalité sur un dispositif connu pour être seulement causal.

10 Cf. par exemple *Hors Cadre* de Samuel Bianchini et Didier Bouchon, 2015 : « Un cadre est accroché à un mur. Il est vide et sobre : aucun contenu, aucun fond, en bois brut, sans moulure ou autre ornement. Mais, par intermittence, il bouge, subtilement, se tord sur lui-même. Ponctuellement, ses mouvements se font plus brusques, violents même, comme s'il subissait des décharges motrices, involontaires, telles celles à l'œuvre dans les crises d'hystérie. Le cadre est comme un corps soumis à des pulsions qui le dépassent, qu'il n'est pas en mesure de cadrer. Il se donne à saisir par ses seuls mouvements et non par une quelconque représentation qu'il pourrait contenir. » (<http://dispotheque.org/fr/hors-cadre>)

Ainsi pouvons-nous estimer que les projections fictives en jeu dans l'interaction homme/machine sont motivées par le besoin de comprendre, et sans doute également par le désir d'Interagir. En résumé, d'un côté, plus il y a fiction, plus il y a Interaction ; mais de l'autre, plus il y a interaction, moins il y a fiction. L'utilité cognitive de la fiction pourrait s'effacer, finalement, derrière la force conative d'un robot de plus en plus autonome. Une question demeure en suspens : pourquoi désirons-nous à ce point l'Interaction avec des robots ?

4. Les projections-intentions : ou pourquoi la compréhension d'une action ne se réduit pas à sa dimension descriptive.

La qualification de l'interaction entre un humain et un robot semble soulever des difficultés inextricables. En interagissant, nous ferions comme si le robot était un être doté d'une intériorité, donc nous serions dans un jeu de faire-semblant. Ce jeu devrait cependant nous assurer qu'il n'y a pas de véritable Interaction mais tout au plus une simulation, de notre côté tout autant que de celui de la machine. Or cette simulation, si elle en est une, semble bien requise pour produire une interaction sérieuse, donc non simulée, de la part du sujet humain. Pour sortir de ce cercle, nous proposons ici de distinguer différents types de faire-semblant, puis différents types d'interaction qui signifient différents niveaux d'engagement. Grâce à ces distinctions, nous aborderons à nouveaux frais la question de l'interprétation et de ce qui la distingue de la description.

4.1. Différents types de faire-semblant : nécessaires ou cognitivement utiles ?

Nous avons évoqué le faire-semblant dans plusieurs contextes. Dans le jeu, tout d'abord, à la suite de Walton, quand des enfants font comme si des souches figuraient des ours. Dans l'interaction avec une machine, d'autre part, lorsque des sujets font comme s'ils avaient affaire à un être capable d'intentions et d'émotions. Il y a certainement une continuité entre ces deux situations, mais probablement pas d'identité : tandis que le jeu enfantin relève bien de la fiction interactive, le jeu de l'adulte avec un robot paraît impliquer une interaction authentique, quoique pas tout à fait ordinaire.

Il est une différence fondamentale entre les deux exemples : avec la chasse aux ours, on envisage des possibilités contrefactuelles et tenues pour telles. Les enfants ne croient pas que les souches sont des ours, ils sont même convaincus du contraire, mais font comme si elles en étaient. Dans le cas de l'interaction avec un robot, la situation est évidemment moins claire. Considérons la possibilité que la lampe ait l'intention de m'indiquer une zone sur l'espace de l'expérience : cette possibilité est-elle contrefactuelle ? Cela n'est pas forcément le cas. L'intention est peut-être bien présente, quoique de façon dérivée – ce serait le manipulateur de la lampe, ou l'ingénieur qui l'a conçue, qui seraient à l'origine de cette intention et qui l'exprimeraient à l'aide de la lampe. Ou peut-être n'est-elle pas plus ni moins présente que lorsqu'un chien m'indique un chemin. Dans une veine quinienne (Quine 1960), on pourrait considérer qu'il n'y a pas de fait pour trancher la question : ni le chien, ni la lampe n'ont les mots pour exprimer leurs intentions. Que la lampe ait cette intention ou non, cela peut être ni vrai ni faux, mais simplement une attribution plus ou moins utile et pertinente de ma part.

Le dénominateur commun aux deux jeux de faire-semblant est la suspension de nos croyances. Dans le premier cas, nous ferions semblant de croire à ce que nous tenons pour faux, dans le second, nous ferions comme si quelque état de chose était réalisé, indépendamment de nos croyances quant à sa réalité. Ce qui distingue les deux situations semble lié à la fonction du jeu de faire-semblant. Dans le premier cas, on nage en pleine fiction, et postuler que les souches sont des ours est constitutif de la conduite du jeu – que l'on cesse de faire semblant et le jeu s'arrête. Dans le second cas, le postulat d'une intériorité à la lampe n'est pas une condition nécessaire à l'interaction ; en revanche, comme nous l'avons souligné dans la section précédente, ce postulat répond à un motif cognitif : c'est en prêtant des intentions et des émotions à la machine que nous pouvons donner du sens à son comportement, le traiter comme une action et par suite, développer une Interaction avec elle. Le faire-semblant de l'interaction avec les robots relève ainsi de ce que Dennett (1987) appelle la « stratégie de l'interprète » : cette posture consiste à prêter de l'intentionnalité à un système pour faciliter la prédiction de son comportement¹¹.

11 Pour Dennett, l'usage de cette stratégie n'est justifié que lorsqu'elle facilite les prédictions. Pour prédire le fonctionnement d'un thermostat, la description fonctionnelle suffit (et sa description physique en cas de

4.2. Différents types d'interaction (à nouveau) : finalité interne ou externe ?

On peut distinguer les interactions humaines suivant plusieurs traits : immédiates (en face à face) vs. médiatisées (par le téléphone, l'écrit, un écran) ; dans le réel vs. dans le virtuel (dans un jeu vidéo multijoueurs) ; rationnelles (comme les échecs) vs. corporelles (comme la danse). Il semble toutefois que ces distinctions n'aient pas d'impact particulier sur le type de faire-semblant mobilisé. En revanche, la *finalité* de l'interaction en a certainement un : ce que les sujets projettent dans l'interaction, ce qu'ils en attendent ou le sens qu'ils lui donnent, vont déterminer leur posture.

On peut alors distinguer le théâtre ou le jeu à composante fictionnelle¹², où la finalité est contenue dans l'interaction, des interactions dont la finalité est inscrite au-delà du jeu interactif. Dans les situations du premier type, le jeu de faire-semblant est celui des enfants mimant la chasse à l'ours : on y fait comme si ce que l'on tient pour faux était vrai à seule fin de faire fonctionner la fiction constitutive du jeu. Au théâtre ou dans un jeu de rôles, on fait comme si on incarnait tel ou tel personnage, avec son histoire, ses états mentaux... en sachant que ce n'est pas le cas. Il s'agit ainsi de cas singuliers où l'interaction est entièrement conduite dans un cadre fictionnel.

À l'autre pôle, il faut considérer toutes les activités conjointes dont la finalité excède le cadre de l'interaction. Ainsi en va-t-il des tâches de production ou de construction, où l'interaction débouche sur un objet : en construisant une cabane, on ne se contente pas de faire comme si on jouait tel ou tel rôle, on agit aussi dans le monde¹³. Cela vaut pour les situations où la cabane est construite dans le cadre d'un jeu de rôles : les enfants qui construisent un abri fictif contre les ours le font en construisant une cabane réelle. Cela vaut aussi pour les situations où la cabane est construite dans le monde virtuel d'un jeu vidéo : un objet virtuel est bel et bien produit, qui fait office de cabane dans le monde fictionnel du jeu (Chalmers 2017). La finalité de l'interaction déborde ainsi le seul cadre fictionnel en impactant le réel. Dès lors, le comme si doit inclure des hypothèses qui ne

dysfonctionnement) et la projection d'intentionnalité n'est d'aucune utilité.

12 Autrement dit, on exclut les jeux sans simulation (*mimicry*) ni vertige (*ilinx*), pour reprendre les catégories de Caillois (1958), comme les dames, les échecs, les jeux mathématiques...

13 On pourrait l'étendre à n'importe quelle interaction ayant des effets sur l'environnement, destructions comme constructions.

sont pas contrefactuelles mais concernent les intentions d'autrui relativement à l'action conjointe dans le monde. Si untel me tend une branche, c'est probablement pour que je la pose à l'endroit où il en manque une dans la cabane-abri contre les ours que nous construisons conjointement.

La résolution de problèmes est également un type d'action qui ne contient pas la totalité de sa finalité. D'une certaine manière, on peut y voir un cas particulier d'une tâche de construction dont le résultat est abstrait plutôt que dans le monde. On peut attacher à ce type d'action le jeu d'échecs, les jeux mathématiques ou d'autres jeux de réflexion. À la finalité interne au jeu (gagner la partie) s'ajoute une finalité externe qui est de l'ordre de la résolution de problème : confrontée à une configuration de l'échiquier, une joueuse va tenter de déterminer une stratégie gagnante. L'interaction suppose alors de projeter des états mentaux à l'adversaire analogues à celles dont on dispose en tant que joueuse : une connaissance de la configuration, une maîtrise des règles, une capacité d'anticipation et de calcul. C'est par cette stratégie de l'interprète que l'on joue aux échecs. Le faire semblant n'est plus ici celui de la fiction interactive, il est ce qui sous-tend le raisonnement en vue de la résolution de problèmes.

Les performances artistiques ont également une finalité qui les dépasse. Si l'interaction sous-jacente peut être fictionnelle ou sérieuse, ce qui compte pour la personne qui agit est alors sa portée esthétique, qui va au-delà de l'interaction elle-même et suppose des niveaux d'interprétation additionnels. D'autres exemples peuvent concerner l'utilisation de mondes ludiques et fictionnels en vue d'apprentissage : des simulateurs de vol aux dits jeux sérieux. Le faire-semblant des agents n'est alors pas seulement celui qui consiste à jouer un rôle dans le dispositif, il comporte également une motivation esthétique ou cognitive, selon les cas, qui le démarque nettement du *comme si* fictionnel.

Pour résumer, les types de faire-semblant paraissent dépendre étroitement de l'articulation de l'interaction à sa finalité. S'il s'agit d'une finalité purement interne, on a un faire-semblant qui l'est également, du type de celui à l'œuvre dans le jeu de la chasse aux ours ou dans la pratique d'un jeu vidéo sans ambition d'apprentissage ou de découverte. Quand la finalité va au-delà du cadre de l'interaction, le faire-semblant est alors mobilisé

dans une stratégie cognitive ou pour des motivations esthétiques ou autres¹⁴. Ce qui fonde cette distinction est probablement à chercher du côté de l'engagement du sujet : minimal, voire nul dans un cadre strictement fictionnel, il prend de l'importance lorsqu'on en sort, à proportion de la portée de l'action au-delà de ce cadre. Quand la finalité de l'interaction la dépasse, elle devient au moins en partie sérieuse.

Conclusion : une question de choix ?

Revenons à notre question initiale, qui était de qualifier l'interaction homme-robot : s'agit-il d'une authentique Interaction ? En interagissant avec un robot, le sujet n'est pas tout à fait comme dans le jeu de la chasse aux ours qui est un jeu assumé de faire-semblant fictionnel. Il est bien plutôt dans une interaction dont la finalité ou la portée va au-delà : il peut interagir sérieusement avec l'objet robot, que cela soit pour une construction conjointe ou pour une performance. Le sujet est engagé et le faire-semblant a une motivation cognitive évidente. Confronté à un objet doté de comportements et apparemment sensible à l'environnement, le sujet est enclin à adopter la stratégie de l'interprète : prêter des intentions à cet objet pour prédire son comportement, voire produire un comportement coordonné. Pour interagir sérieusement avec l'objet, il est naturel et cognitivement pertinent de faire comme si cet objet manifestait une subjectivité.

Dans la typologie des interactions et des jeux de faire-semblant exposée précédemment, nous avons sciemment évité la question de la nature des agents interagissant. Le fait est que l'engagement du sujet et le type de faire-semblant mobilisé ne paraissent pas en dépendre : que je joue au théâtre avec une actrice ou avec un robot, je serai dans un jeu purement fictionnel ; que je joue aux échecs avec un humain ou un programme, je serai dans un jeu rationnel non fictionnel. L'interaction avec les robots comporte toutefois une asymétrie fondamentale entre agents qui conduit potentiellement à un décalage de cadres : l'appréhension de la situation par l'agent humain n'est sans doute pas la même que celle par le robot. Si l'agent humain adopte une forme ou une autre de faire-sem-

14 Il faut ainsi nuancer ce que l'on a présenté plus haut dans la section en assimilant le jeu d'acteurs au théâtre, au jeu de la chasse à l'ours. Le théâtre, comme toute pratique artistique, a des finalités externes. Il en va de même de la production et réception de fictions littéraires, dont le jeu de faire-semblant répond également à des motifs esthétiques et cognitifs.

blant, le robot est quant à lui attelé à la production d'un comportement adapté : peut-on dire qu'il est également dans le faire-semblant ?

Avec les robots, on est finalement confronté à la question de l'applicabilité de nos concepts intentionnels ou mentaux dans des situations limites. À cet égard, on peut soutenir qu'il n'y a pas de bonne ou de mauvaise réponse à cette question, mais que cela est affaire de décision (Putnam 1964). Suivant le principe de la double norme de Quine (1960), l'interprétation ne doit pas être traitée comme une description. Si l'Interaction comporte une dimension comportementale objective (béhavioriste) qui peut être décrite et quantifiée, elle comporte aussi une dimension intentionnelle qui relève de l'interprétation et suppose une projection dans la scène de l'interaction. La « vérité » des interprétations n'est donc pas réductible à la vérité d'une description. Ce que met en lumière l'interaction avec une machine, c'est cette dualité intrinsèque des Interactions en général.

Ainsi, le *comme-si* n'est pas seulement l'affaire de l'agent humain interagissant avec un robot. Il est également du côté de l'observateur de l'interaction. Le caractère intentionnel ou non n'est pas inscrit dans l'interaction elle-même, il est assumé ou non par celles et ceux qui l'observent. Dès lors que l'agent humain adopte la stratégie de l'interprète de manière justifiée, il paraît pertinent de le suivre et d'utiliser des catégories intentionnelles pour analyser la situation. De l'interaction, on passe alors à l'Interaction.

Remerciements

Les auteur·es ont eu l'occasion de présenter leur travail au *Colloque Cathy Dufour "Intelligences Artificielles"* (Vandœuvre-les-Nancy, novembre 2018). Elle et il tiennent à remercier les participant·es à ce colloque pour leurs questions et suggestions, ainsi que le groupe PsyPhINe (MSH Lorraine) et un·e rapporteur·e anonyme pour leurs critiques et commentaires sur une version antérieure de cet article.

Références

- André V., Boniface Y., « Quelques considérations interactionnelles autour d'une expérience robotique », in WACAI (Workshop sur les Affects, Compagnons Artificiels et Interactions), île de Porquerolles, 13-15 juin 2018. (<https://tinyurl.com/ybcvren5>)

- Becker J., André V., Dutech A., « Qualcom : une expérience sur la qualification des comportements d'une lampe robotique », in *Techniques & Culture*, Varia, 2019. (<http://journals.openedition.org/tc/10771>)
- Caillois R., *Les jeux et les hommes*, Paris, Gallimard, 1958.
- Chalmers D., « The Virtual and the Real », in *Disputatio*, IX, 46, 2017, p. 309-352.
- Dennett D., *The Intentional Stance*, Cambridge, Mass., MIT Press, 1987.
- Gendre M., Schneider C., « Lampe et interaction. Projet Kapl'IA. Rapport de projet tutoré », Master SCA, Université de Lorraine, 2016. (<https://tinyurl.com/yde6edfg>)
- Goffman E., *Frame Analysis: An Essay on the Organization of Experience*, New York, Harper and Row, 1974.
- Mori M., « The Uncanny Valley », in *Energy*, 7(4), 1970, p. 33-35.
- Putnam H., « Robots: Machines or Artificially Created Life? », in *The Journal of Philosophy*, 61, 21, 1964, p. 668-691.
- Quine W. V. O., *Word and Object*, Cambridge, Mass., MIT, 1960.
- Radford C., « How Can We Be Moved by the Fate of Anna Karenina? », in *Proceedings of the Aristotelian Society*, Supp. vol. 49, 1975, p. 67-80, trad.fr. « Comment pouvons-nous être émus par le destin d'Anna Karénine ? », in J.-P. Cometti, J. Morizot et R. Pouivet (dir.), *Esthétique contemporaine, Art, représentation et fiction*, Paris, Vrin, 2005, p. 327-345.
- Renauld M., *Philosophie de la fiction. Une approche pragmatiste du roman*, Préface de Th. Pavel, Rennes, Presses Universitaires de Rennes, coll. « Aesthetica », 2014.
- Vidal D., « Vers un nouveau pacte anthropomorphique ! », in *Gradhiva*, 15, 2012, p. 54-75.
- Walton K. L., « How remote are fictional worlds from the real world? », in *Journal of Aesthetics and Art Criticism* 37 (1), 1978, p. 11-23.
- Walton K. L., *Mimesis as Make-Believe*, Harvard University Press, 1990.