# Immersion dans les mondes virtuels et émergence de l'intentionnalité

Benoît Virole

2015-2021

**Résumé**

Ce texte est celui d'une conférence faite au congrès de la revue *Perspectives Psychiatriques*≪ L'intelligence artificielle au défi de l'intersubjectivité ≫, le 13 Mars 2015. Nous insistons sur l'imprévisibilité comme propriété fondamentale de l'attribution d'une intentionnalité.

**Mots-clefs**

Psychanalyse Sciences cognitives Intelligence artificielle intentionnalité cyberpsychologie

*Introduction*

Quelles sont les propriétés dont doit être doté un système d'intelligence artificielle (IA) pour qu'un sujet humain lui attribue une intentionnalité ? Cette question est une autre façon de poser le problème de la légitimité de l'intitulé ≪ intelligence artificielle ≫ en opposant la bêtise apparente d'un système automatique exécutant des instructions fixées, à l'intelligence d'une entité dont le comportement dénote une intention sous-jacente, c'est-à-dire la recherche de la réalisation d'un but indépendamment des conditions effectives de la réalisation de ce but. Un système doté d'une intentionnalité est capable de s'adapter aux variations d'un contexte de réalisation. Cette question est essentielle dans la recherche associant robotique et sciences humaines. Les robots et les avatars numériques présentent des enjeux industriels et sociétaux importants. Ne serait-ce, par exemple, que pour la création de bornes interactives numériques utilisant des personnages virtuels. Il existe toutes sortes de recherche sur l'encodage et le décodage des expressions émotionnelles, sur la modélisation des interactions linguistiques, sur celle du raisonnement, etc. L'attribution à ces entités artificielles d'une capacité intentionnelle donnant l'illusion d'une interactivité humaine est un challenge important. Il est d'autant plus important qu'il existe une contradiction profonde entre intentionnalité et système artificiel. Pour Husserl, la construction de l'intersubjectivité nécessite la perception de l'intentionnalité, agie dans le corps (≪ *Leib* ≫) de l'autre[1]. Elle implique la résonnance kinesthésique avec le corps charnel de l'autre, résonnance considérée comme le fondement de l'empathie (résonnance émotionnelle). La prise de conscience de l'existence d'une mentalisation chez l'autre (≪ théorie de l'esprit ≫) nécessite l'intuition anticipatrice des mouvements de l'autre. L'intersubjectivité, fondement de la relation au monde social et à la communication des états mentaux, semble ainsi, en première approximation, tributaire de la reconnaissance de l'intention *incarnée* dans le corps de l'autre. Comment ce processus de résonnance kinesthésique avec la ≪ chair ≫ de l'autre peut-il se produire avec des avatars numériques dénués de toute sensibilité charnelle ? Nous nous proposons d'apporter une contribution, modeste, à ce domaine de recherche en essayant d'identifier les propriétés d'attri-

1. Husserl distingue la posture kinesthésique qui relève du ≪ *Leib* ≫ et la posture anatomique qui relève du ≪ *Körper* ≫, cf. Husserl E., *Sur l'intersubjectivité*, (tome I et II), (trad. Nathalie Depraz) Puf, 2001.

bution d'une intentionnalité à des personnages virtuels en exploitant les données cliniques recueillies à partir de la situation des psychothérapies utilisant la médiation des jeux vidéo.

### *Le retour de l'intelligence artificielle*

Rappelons que le terme d'*intelligence artificielle* désigne des techniques permettant à des systèmes informatiques de réaliser des opérations cognitives de haut niveau similaires ou approchant celle de l'intelligence humaine. Trois techniques principales ont été développées[2] :

1. Les systèmes experts basés sur l'écriture de règles d'inférences et permettant la genèse de raisonnement artificiel comparable à celui des déductions d'un expert humain dans un domaine considéré ; ces systèmes nécessitent l'encodage d'instruction dans un langage de programmation symbolique. Ils permettent la reproduction de raisonnements déductifs de haut niveau, sont moins habiles pour reproduire des comportements de catégorisation et de se modifier eux mêmes par apprentissage.
2. Les réseaux de neurones artificiels (connexionnisme) permettant d'opérer des catégorisations perceptives et de générer, par émergences, des conduites intelligentes d'optimisation. Toutes les opérations logiques de base (et, ou, ou exclusif) peuvent être réalisées par des assemblages d'unités reproduisant les processus de seuil existant dans les neurones biologiques. Ces réseaux sont dépendants de processus d'apprentissage (supervisé ou non), lorsque on boucle les sorties sur les entrée, des processus d'émergences fonctionnelles nouvelles peuvent être observés.
3. Les systèmes multi-agents et les algorithmes génétiques utilisant la puissance d'itération de calcul pour optimiser les solutions à des contraintes complexes, de la même façon que la sélection naturelle performe des organismes adaptés à leur environnement.

Moteurs d'une grande aventure intellectuelle et scientifique jusqu'à la fin du siècle dernier (1960-2000), ces techniques ont marqué un palier du fait de la croissance exponentielle de la puissance des processeurs et des capacités de stockage qui permettent la manipulation de connaissances stockées, sans avoir besoin de processus sophistiqués. Aujourd'hui, le développement de la robotique et des mondes virtuels, tels ceux intégrés aux jeux vidéo, sollicite à nouveau des approches inspirées de l'intelligence artificielle pour générer des univers interactifs où le comportement d'un avatar système est piloté par des scénarios évolutifs tenant compte du contexte[3]. Le sujet humain utilisant ces mondes virtuels rencontre ces avatars pilotés par ces systèmes d'IA qui vont interagir avec lui et adopter des comportements différents selon les actions du sujet. Il en résulte un phénomène psychologique spontané chez le sujet humain : il leur attribue une *intention.*

### *Cadre et technique de la psychothérapie*

Ce phénomène d'attribution d'intention est très aisément observable dans les thérapies à médiation par jeux vidéo. Rappelons le cadre. Le patient, généralement un enfant ou un adolescent, est assis côte à côte avec le psychothérapeute devant un écran d'ordinateur où défilent les images d'un jeu vidéo. Patient et thérapeute vivent une expérience conjointe d'immersion (une co-immersion) dans ce monde virtuel. La plupart du temps, le patient manipule numériquement un avatar (généralement anthropomorphe, parfois sans apparence comme dans le cas d'un pointeur) et interagit avec d'autres avatars et objets. Dans les jeux en ligne, ces avatars sont dirigés par d'autres joueurs. Dans notre dispositif thérapeutique, l'avatar patient interagit exclusivement avec des avatars dirigés par le programme du jeu. Sur le plan psychothérapeutique, cette situation permet la réalisation symbolique de fantasmes, en particulier narcissiques, de défenses, l'expression d'émotions et est à la source d'une verbalisation. Tous ces éléments deviennent des matériaux pour l'interprétation et la conduite de la thérapie qui suit les voies habituelles de la psychothérapie psychodynamique (frustration contrôlée, régression, maniement

2. Il existe d'autres techniques : réseaux bayésiens, réseaux d'automates, théorie des jeux, algorithmes génétiques, etc.

3. Cf. les jeux *Créatures (*1998) intégrant des réseaux de neurones formels pour piloter les créatures Norms capables d'apprendre des comportements ; le jeu *Halo : combat Evolved* (2001) utilisant des arbres de décision ; *FEAR* (2005) avec planificateur de scripts sensibles au contexte.

du transfert, interprétation). Mais à ces dimensions classiques de la thérapie d'orientation psychanalytique s'ajoute une dimension particulière : l'avatar est investi comme étant un objet-soi, non pas dans le sens d'un investissement narcissique d'une image de soi, même idéalisé, mais dans le sens d'un objet incarnant les réalisations intentionnelles. Cet investissement s'accompagne d'une très faible dépense énergétique musculaire réelle, donnant ainsi une forme de plaisir par épargne. La perception du mouvement de son propre avatar génère par effet miroir (décharges corolaires) un investissement énergétique moteur qui n'est pas déchargé musculairement mais dans un affect de plaisir.

Ce cadre est propice à l'observation de phénomènes d'attribution par le patient de propriétés intentionnelles à l'avatar piloté par le système. C'est cet aspect, tangentiel par rapport aux buts de la psychothérapie, qui nous intéresse directement. La situation la plus exemplaire est celle où le sujet dirige son avatar dans un monde virtuel qui contient un autre avatar dirigé par le système et qui cherche à attaquer, à nuire, ou à capturer l'avatar sujet. Dans la réalité informatique, la plupart du temps les mouvements de l'avatar système sont pilotés par un ensemble de règles d'inférences mais de plus en plus de jeux intègrent des méthodes plus complexes utilisant les ressources de l'intelligence artificielle. Ces interactions entre avatar-sujet et avatar-système sont banales dans les jeux vidéo (par exemple : *Ray Man*, *Tomb Raider*, *GTA*...). Le sujet déplace son avatar, réalise des actions, et l'avatar-système (ennemi, monstre,...) se déplace et agit en tentant compte des mouvements de l'avatar sujet. Cette situation génère chez le sujet des verbalisations qui sont les indices d'une attribution d'intention à l'avatar système : « Il veut m'obliger à aller dans cette direction » , « Il croit que je n'ai pas vu ce qu'il cherche à faire, mais on va l'avoir », « Qu'est-ce qu'il cherche ? », « Je sais ce qu'il me veut », « Pourquoi me fait-il cela ? ». Toutes ces phrases et questions révèlent l'existence chez le sujet de la croyance en l'existence d'une intention présidant aux mouvements réalisés par l'avatar du système et dirigée vers le sujet. L'étude linguistique est particulièrement instructive car le sujet va parlé de l'avatar système en utilisant une forme pronominale réflexive et des verbes dont le sémantisme est celui d'une intentionnalité « il me (cherche) » ; « Qu'est-ce qu'il me veut (vouloir) ». La variabilité et l'éthique de la situation psychothérapeutique ne permettent pas l'établissement de statistiques mais nous avons suffisamment de recul et recueilli suffisamment de données cliniques pour affirmer qu'il existe trois moments logiques cette attribution d'intentionnalité.

*L'attribution ontologique*

Le premier moment est celui de l'immersion. Le sujet attribue à la situation virtuelle un indice de réalité suffisant, mais momentané, pour induire une conduite cognitive adaptée. En d'autres termes, « il joue le jeu » « il est dans le jeu » . Cette attribution transitoire de réalité n'est pas triviale. Elle n'est pas liée directement à la qualité de simulation de la réalité – des jeux pauvres sur le plan graphique sont fortement immersifs et des jeux à la splendeur graphique, ne se sont pas – mais à la qualité du couplage entre la réponse de son avatar à l'intention du sujet (en terme de rapidité, d'interactivité, de « jouabilité » pour reprendre le terme utilisé par les joueurs. L'aspect « dynamique » de l'acte virtuel est à la source du sentiment d'appropriation d'un avatar qui semble répondre avec célérité aux intentions d'actions émises par le sujet. Cet aspect est fondamental car nous savons aujourd'hui que la représentation consciente d'une intention *suit* le déclenchement de l'action et non pas le *précède.* La décision consciente n'est pas la cause du mouvement mais sa conséquence. Les mouvements mentaux intentionnels conscients ne sont pas les causes de nos actions mais les marqueurs réflexifs d'une action déjà engagée (Libert, 1985)[4]. Le couplage n'est donc pas celui de la réponse de l'avatar à la décision consciente de réaliser un mouvement. Il est celui de l'observation d'un mouvement réalisé par l'avatar par un déterminisme inconscient et qui sera secondairement marqué par un indice d'intentionnalité consciente. Il ne faut pas confondre ce couplage avec les simples phénomènes réflexes. Cette secondarité de l'intentionnalité consciente se produit pour des

4. Libert Benjamin (1985) « Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action » , *The Behavorial and Brain Sciences*, 8 : 529-566.

| *Pronoms* | *Fonctions* |
|---|---|
| *Je* | Désigne parfois le sujet patient seul, parfois son avatar dans le jeu par décentrement. |
| *Tu* | Adresse du patient à son propre avatar (en cas de désaccord entre la réalisation et l'attention) ou à l'avatar ennemi |
| *Il* | Dénomination de son propre avatar avec détachement du lien d'appropriation intentionnelle ou dénomination de l'avatar système |
| *On* | Dénomination englobant la plupart du temps patient et thérapeute conjugués dans l'acte intentionnel performé par l'avatar |
| *Nous* | Dénomination englobant patient et thérapeute mais avec scission de l'avatar considéré alors comme extérieur |
| *Vous* | Adresse du patient au thérapeute |
| *Ils* | Dénomination par le patient des objets ou avatars multiples du système |

Tableau 1 – Les différents usages des pronoms personnels dans les verbalisations du patient au cours de l'immersion dans les jeux vidéo.

actions dont la cinétique de développement est beaucoup plus longue que les boucles réflexes médullaires.

*L'intégration de l'imprévisible*

Le second moment est ainsi celui de la surprise, de l'inattendu. Notons d'abord l'importance de l'imprévisible dans la construction du sentiment de réalité. La détection de phénomènes aléatoires, non prévisibles, hasardeux génère un sentiment de réalité beaucoup plus intense que celle de la simulation concrète de la réalité. Des mondes virtuels très réalistes, dotés d'une sophistication multisensorielle, mais prévisibles par leur automatisme, se révèlent moins immersifs que des applications basiques sur le plan graphique mais qui présentent une imprévisibilité dans les objets virtuels. Le sujet attend du système un comportement qui ne se produit pas et un autre comportement, imprévisible, survient à la place, générant une mise en alerte réflexive du sujet. L'avatar système se positionne par exemple à une place inattendue obligeant le sujet à réfléchir sur une nouvelle stratégie. Il peut aussi échouer par une stratégie perdante, faire une erreur dans une trajectoire, dont le sujet a observé précédemment qu'il la connaissait pourtant parfaitement. Ou bien encore, l'avatar se détourne d'un but standard du jeu (centré sur le sujet) pour un autre but, inconnu du sujet. Toutes ces situations mettent le sujet dans un situation de surprise, prémisse à une réflexion. La conséquence systématique est alors l'attribution au système d'une intention. L'immersion est alors consolidée. Le sujet *croît* (momentanément) qu'il est en présence d'un sujet intentionnel.

*L'appréhension cognitive de la logique de l'autre*

Le troisième moment est celui de la déduction de la logique de l'autre. Il nécessite un processus d'intrusion cognitive dans la pensée de l'autre, processus dont la psychanalyse a fourni une conceptualisation utile avec la notion *d'introjection projective*. Le sujet attribue à l'avatar système une intériorité cognitive, une intentionnalité comportant des buts cachés, des méthodes inconnues, qui vont nécessiter de la part du sujet une activité réflexive se concrétisant ensuite dans de nouvelles stratégies ou hypothèses. Le comportement du sujet, ses décisions d'actions, vont intégrer les hypothèses sur les raisons qui président aux actions de l'autre (l'avatar système), y compris la possibilité que l'autre réalise des erreurs. Nous assistons alors à une forme d'intersubjectivité, certes réduite par rapport à l'intersubjectivité humaine aux

seuls aspects cognitifs, mais néanmoins intéressante par son coté embryonnaire.

## Conclusions

L'étroitesse de notre situation d'observation limite la portée des conclusions que l'on peut en tirer. Cependant, nous sommes en droit de proposer trois éléments de réflexion.

1. Le premier concerne le domaine de l'intelligence artificielle. Comment pourrait-on améliorer les avatars numériques et les robots, pour générer chez le sujet humain l'illusion d'être en présence d'une intelligence intentionnelle ? Notre expérience nous invite à considérer que *l'imprévisibilité* est bien une des propriétés fondamentales de l'attribution de l'intentionnalité humaine. C'est bien la présence du hasard, et d'une certaine façon l'erreur, qui donne le sentiment du vivant. Les systèmes-experts à base de règles d'inférences, même ceux fonctionnant en chaînage arrière, paraissent moins intelligents qu'un réseau de neurones de type carte de Kohonen, enrichi par un bruit (hasard) et bifurquant de façon imprévisible et ceux-ci paraissent encore moins intelligents qu'un système à algorithme génétique qui assume la nécessité de la variabilité du hasard pour le génération de comportements adaptés aux contextes mouvants.
2. Le second élément de réflexion concerne la psychologie humaine. Le dispositif thérapeutique avec les jeux vidéo invite à rehausser l'importance de la situation de la co-attention mère (ou partenaire) enfant devant un objet tiers doté d'une autonomie comportementale (jouet, animal, processus physique). Cette situation (dont le prototype se situe vers le huitième mois de vie) active un phénomène de couplage des pensées, une forme des synchronisation des actes cognitifs. La construction de la théorie de l'esprit prend source dans ce moment très particulier où l'écart entre la préconception et le mouvement de l'objet tiers génère une déstabilisation cognitive induisant l'attribution d'une intention autonome à l'objet.
3. Le troisième concerne la possibilité de réaliser des thérapies avec des avatars thérapeutes mettant au chômage technique les psychothérapeutes humains. Il ne s'agit pas ici de science fiction. Les relations étonnantes entre les enfants autistes, les robots et les avatars virtuels pourraient bien anticiper un nouvel âge où des thérapeutes numériques, programmés par des experts humains, montrerons une compétence, une disponibilité et une patience infinie. Bien sûr, il paraît difficile de leur attribuer des aptitudes au transfert analytique et à la compréhension de la complexité de la vie psychique. On pourrait aussi sourire de l'idée d'interprétations artificielles, bien que, parfois, on les observe *in situ* dans les cadres thérapeutiques les plus humains ! Mais il est possible que ces avatars thérapeutes, capable de comprendre des énoncés linguistiques, de lire dans les traits du visage les indices des émotions réprimées puissent donner en retour des réponses émotionnelles calibrées, offrir dans leur réponse une frustration soigneusement dosée, dans un réceptacle bienveillant enclenchant une croissance psychique. Encore faudrait-il que ces robots ≪ thérapeutes ≫ puissent être capables de quelques erreurs opportunes, de lapsus, d'incertitudes et d'une imprévisibilité charmante, en d'autres termes qu'ils deviennent envers et contre tout diablement humains.

*Publications en référence à l'intelligence artificielle*

Virole B., Siboni J. ≪ Neuropsychologie et Intelligence Artificielle ≫ , ANAE *Approche Neuropsychologique des Apprentissages chez l'Enfant*, Vol2., pp.171 à 176, 1990.

Virole, B., *Semantic links between single words in schizophrenia, An artificial intelligence approach*, Étude Inserm, imagerie fonctionnelle et schizophrénie, sous la direction de Jean-Luc Martinot, 1995 disponible sur www.benoitvirole.com.

Virole B., *Sciences cognitives et Psychanalyse*, Presses Universitaires de Nancy, 1994.

Virole B., Réseaux de neurones et psychométrie, Étude prospective des applications des réseaux de neurones formels dans le traitement des données psychométriques. *Editions du Centre de Psychologie Appliquée*, 2001, disponible sur www.benoitvirole.com.

*Publications sur le virtuel en psychothérapie*

Virole B., Radillo A., *Cyberpsychologie*, Paris, Dunod, 2010.

Virole B., ≪ La technique des jeux vidéo en psychothérapie ≫ , *Subjectivation et empathie dans les mondes numériques*, S. Tisseron ed., Dunod, 2013.

Pour citer cet article :

https ://virole.pagesperso-orange.fr/IAPSY.pdf (2015).