



HAL
open science

Utiliser les agents virtuels pour lutter contre les discriminations sociales

Morgane Peirola, Magalie Ochs

► **To cite this version:**

Morgane Peirola, Magalie Ochs. Utiliser les agents virtuels pour lutter contre les discriminations sociales. WACAI 2021, Centre National de la Recherche Scientifique [CNRS], Oct 2021, Saint Pierre d'Oléron, France. hal-03377551

HAL Id: hal-03377551

<https://hal.science/hal-03377551>

Submitted on 14 Oct 2021

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.

Utiliser les agents virtuels pour lutter contre les discriminations sociales

Morgane Peirolo

Laboratoire Parole et Langage
Aix-Marseille Université
Marseille, France
morgane.peirolo@gmail.com

Magalie Ochs

Laboratoire d'informatique et des systèmes
Aix-Marseille Université
Marseille, France
magalie.ochs@lis-lab.fr

RÉSUMÉ

Les agents virtuels peuvent être utilisés pour promouvoir des comportements pro-sociaux. Dans cet article, nous explorons le potentiel des agents virtuels et de la réalité virtuelle pour lutter contre les discriminations sociales à travers l'induction d'empathie chez l'utilisateur. Nous présentons différents systèmes développés dans cette perspective et discutons les facteurs d'influence, les outils de mesures des performances de ces systèmes, ainsi que les limites actuelles de ces travaux.

MOTS CLÉS

Discrimination, environnements virtuels, réalité virtuelle, stéréotype, préjugé, empathie, informatique pro-sociale, agents virtuels, agents conversationnels animés.

INTRODUCTION

Dans nos sociétés contemporaines, les discriminations restent omniprésentes. Ce phénomène se décline sous plusieurs formes, les plus connues étant les discriminations raciales, sexistes et religieuses. D'autres formes moins médiatisées, mais tout aussi stigmatisantes existent, telles que l'âgisme (discrimination liée à l'âge), la discrimination liée aux handicaps ou encore la glottophobie, définie par Blanchet (2016) comme une discrimination linguistique. Bien qu'il existe une littérature détaillée de leurs mécanismes d'action, ces comportements perdurent et sont ancrés dans le tissu social au point qu'il semble impossible de les modifier (Guimond, 2000). Les méthodes traditionnelles de lutte contre les attitudes discriminantes font état de cette résistance au changement : il semble que les campagnes de sensibilisation ont un effet limité sur la prise de conscience collective des problèmes liés aux discriminations. À l'heure actuelle, de nouvelles techniques de réalité virtuelle de sensibilisation se tournent vers la mise en situation des participants : ces derniers, placés dans la peau de personnages virtuels discriminés (*body ownership*), sont alors à même de remettre en question leurs comportements et leurs schémas cognitifs.

Dans le domaine de l'e-éducation et de la formation, un intérêt croissant est apparu autour de la « formation par la

simulation » dans des environnements virtuels. De nombreux environnements de réalité virtuelle ou mixte ont été réalisés pour le développement de compétences techniques chez l'utilisateur. Plus récemment, un certain nombre de travaux de recherche ont été menés autour de dispositifs permettant de simuler une interaction sociale avec un Agent Conversationnel Animé (ACA) et ainsi entraîner ses propres compétences sociales (« *Virtual Agents for Social Skills Training (VASST)* » (Bruijnes, et al., 2019)). Les compétences sociales, partie intégrante de l'intelligence sociale, se définissent comme la capacité d'un individu à gérer son comportement, verbal et non-verbal, pour construire une relation sociale avec autrui (Albrecht, 2006).

Du point de vue de la machine, cette intelligence socio-émotionnelle est d'autant plus importante que l'utilisateur lui-même à une propension à adopter un comportement social face à ces entités artificielles : c'est le paradigme *CASA* (« *Computers Are Social Actors* »). Les individus ont de plus tendance à anthropomorphiser les systèmes interactifs, même caricaturaux, en leur attribuant un genre, une ethnie et une personnalité. Les comportements sociaux des utilisateurs, qui savent qu'ils interagissent avec une machine mais qui se comportent comme si celle-ci était humaine, sont déclenchés automatiquement et de manière inconsciente (Nass et Moon, 2000). D'autres recherches ont confirmé ces comportements sociaux des utilisateurs dans le contexte d'interaction avec des agents conversationnels animés (Hoffmann *et al.*, 2009, Kopp *et al.*, 2005).

Ainsi, les systèmes interactifs humanoïdes peuvent permettre d'améliorer ces compétences interactionnelles (Aylett, 2007). Dans le projet eCUTE (Hall et al., 2011), les enfants et les jeunes adultes sont confrontés à différents scénarios d'interactions avec des personnages virtuels pour les sensibiliser aux différences culturelles et développer leur empathie. La plateforme FearNot ! (Aylett, 2007) a été développée pour lutter contre les comportements de harcèlement scolaire à l'école. Elle repose sur la propension des individus à ressentir de l'empathie envers les systèmes interactifs humanoïdes. En jouant le rôle de la victime à travers un personnage virtuel et en développant ainsi de l'empathie pour les victimes de harcèlement scolaire, les harceleurs changent leur comportement. En questionnant les biais automatiques et implicites des individus, l'empathie s'avère ainsi être un facteur important dans la lutte contre les discriminations.

Dans cet article, nous discutons de **l'utilisation des agents virtuels et de la réalité virtuelle dans la remédiation des comportements discriminants**. L'article est organisé de la manière suivante : dans une première section, nous décrivons les fondements théoriques de la formation des stéréotypes et préjugés, ainsi que l'intérêt de l'empathie dans la lutte contre la discrimination. Dans une deuxième section, nous explorons l'apport de l'informatique pro-sociale dans la sensibilisation à la discrimination, en particulier grâce aux interactions entre individus et agents virtuels. Enfin, nous présentons dans une troisième section un aperçu des méthodes utilisées pour mesurer l'impact de tels systèmes interactifs sur les comportements discriminants.

I. FONDEMENTS THÉORIQUES

1. De la représentation sociale aux comportements discriminants

A. Le rôle de la représentation sociale dans la construction de la réalité

La vie quotidienne nous offre un monde déjà construit : tout y est structuré, hiérarchisé selon des normes sociales que les individus intègrent sans les questionner. Selon Schutz (1974), cette construction offre une connaissance partiellement valide de la réalité, toujours biaisée : elle permet à l'individu d'agir rapidement, sans accorder trop de réflexion à ses actes. De cette approximation naissent les représentations sociales. Les représentations sociales sont des structures qui permettent à l'individu de comprendre et de classer la réalité pour lui donner un sens et justifier ses propres comportements (Moscovici, 1972). Celles-ci sont construites à travers un ensemble d'éléments tels que les attitudes, les valeurs, les opinions, etc. Cette grille conceptuelle à travers laquelle nous appréhendons la réalité et notamment le rapport à autrui est par définition restrictive. S'y expriment les difficultés à accepter l'étrange, le différent, ce qui est exclu de la norme et se trouve inclassable : c'est sur cela que se construisent les idéologies discriminantes. Ces images font partie des représentations sociales et sont fondées sur une catégorisation simplificatrice de la réalité sociale, qui donne lieu à des comportements de rejet (Allport, 1954). En s'encrant dans l'imaginaire commun, ces images sont ensuite relayées et enrichies par les interactions que tissent les acteurs sociaux. C'est ainsi que Goffman (1988) définit l'interaction : non pas comme la simple rencontre de deux personnes mais comme « *un certain type d'ordre social* ».

B. Stéréotype, préjugé, discrimination : définitions

Selon une approche traditionnelle, on définit généralement la triade suivante : le *stéréotype*, le *préjugé* et la *discrimination*. Selon Fiske (1998), le *préjugé* en serait la composante affective, le *stéréotype* la composante cognitive et la *discrimination* la composante comportementale. Contrairement au stéréotype, qui se définit par une croyance à l'encontre d'un groupe distinct, le préjugé est généralement chargé d'un affect. Ce qui les distingue est essentiellement la dimension consensuelle du stéréotype, qui se

veut descriptif d'un groupe (par exemple "il y a peu de femmes en filières scientifiques"), tandis que le préjugé renvoie à une évaluation chargée d'affect (par exemple "les femmes sont mauvaises en sciences"). La discrimination est quant à elle la manifestation active de ces schémas conceptuels et affectifs. Elle s'exprime par une mise à l'écart ou un traitement différentiel des membres d'un groupe donné (Baron & Branscombe, 2011).

Une des principales difficultés liées à la remédiation des comportements discriminants repose sur le choix de stratégies adéquates pour induire des changements durables dans les attitudes et les comportements. Le risque majeur est de créer un phénomène de résistance à la persuasion (McGuire, 1964) : l'individu tend à résister à un argumentaire persuasif par la réfutation. Cette résistance est par ailleurs d'autant plus grande que le comportement ou le schéma de pensée initial est ancré dans des pratiques sociales et partagé par des groupes de référence. Pour pallier ce problème, une stratégie consiste à se fonder sur le phénomène d'empathie.

2. L'empathie comme stratégie persuasive

A. Définition

L'empathie se définit comme la capacité à s'identifier à autrui en adoptant son point de vue à travers la compréhension de ses états cognitifs et affectifs (Dunning et Nordgren, 2013). Les capacités empathiques sont donc doubles : une capacité cognitive à s'approprier l'état mental d'autrui, par nature subjectif, et une capacité affective à partager son état émotionnel (Decety, 2004).

B. Réduire les discriminations grâce à l'empathie ?

Selon Janezic (2015), « *l'empathie est une variable médiatrice de persuasion* ». En effet, un nombre croissant d'études montre que la création de réponses empathiques chez le spectateur est un levier efficace de prévention de la discrimination (Abrams et Maibach, 2008). Ce phénomène s'explique par trois facteurs : la dimension cognitive et émotionnelle de l'empathie, le sentiment d'implication et la prise de perspective. Nous détaillons ces trois facteurs ci-dessous.

a. Ressources empathiques

Par sa double composante cognitive et affective, l'empathie est un ressort important dans les stratégies de sensibilisation et de lutte contre les discriminations. (Batson et Ahmad, 2009). En effet, le traitement cognitif et émotionnel de l'information que requiert l'empathie permet un ancrage en mémoire plus profond que celui produit par des émotions non empathiques (Shamay-Tsoory, 2010).

b. Le sentiment d'implication

Le recours à l'empathie permet d'augmenter le sentiment d'implication de l'individu vis-à-vis d'une problématique (Campbell et Babrow, 2004). Ce sentiment d'implication est, quant à lui, un facteur motivationnel important : en stimulant les ressources empathiques d'un individu, celui-ci est ainsi plus susceptible de se sentir impliqué et mobilisera plus de ressources pour lutter contre ces comportements.

c. *La prise de perspective*

La prise de perspective est la capacité à se mettre à la place d'autrui. Selon Stephan et Finlay (1999), ce comportement permet d'atténuer la différence perçue entre l'individu et l'autre. Ainsi, se mettre dans la peau de personnes discriminées confronte le sujet à la réalité des comportements discriminatoires et le motive à améliorer ses attitudes envers l'exogroupe (groupe constitué d'individus externes à son groupe de référence, l'endogroupe). (Upshaw, Kaiser et Sommerville, 2015).

C. *La détection de la discrimination par un observateur*

À notre connaissance, peu de travaux se concentrent sur le rôle de l'observateur dans la détection et la remédiation de la discrimination. Il est cependant intéressant de questionner l'implication du témoin dans le processus de sensibilisation. Le terme d'observateur distant désigne un observateur extérieur à la scène : il n'est ni discriminateur ni victime, mais témoin. Ce statut particulier de l'observateur a un double conséquence : en étant extérieur à l'interaction, l'observateur est moins impliqué dans la situation ; cette distanciation permet une prise de recul et un détachement émotionnel suffisant pour stimuler son empathie et par là même le pousser à l'action en produisant un comportement pro-social (Reny Delisle, 2013). Dans cette perspective, aujourd'hui certaines formations interpersonnelles visent à entraîner les témoins de scènes de discrimination à réagir. Cependant, comme présenté dans la section II.2, aujourd'hui dans les systèmes interactifs, les utilisateurs sont la plupart du temps placés dans le rôle de la victime de discrimination plutôt que dans celui du témoin.

II. AGENTS VIRTUELS ET RÉALITÉ VIRTUELLE : APPORTS

Comment les agents virtuels peuvent-ils amener les utilisateurs à réduire leurs stéréotypes et générer des comportements empathiques chez l'utilisateur ? Cette problématique est au cœur des travaux de l'informatique pro-sociale. Dans cet article, nous proposons un état des lieux sur la remédiation de la discrimination en environnement virtuel. Nous y décrivons les récents travaux portant sur l'induction d'empathie grâce aux agents virtuels et définissons plus précisément l'utilité de l'immersion en réalité virtuelle dans la lutte contre la discrimination.

1. Améliorer l'humain à travers les machines : l'informatique pro-sociale

Les systèmes interactifs humanoïdes peuvent être utilisés pour promouvoir les coopérations entre individus et plus généralement les comportements prosociaux (Paiva *et al.*, 2018). Les questionnements sur l'utilisation de l'Intelligence Artificielle pour le bien de la société sont récents. Différents événements sont

organisés dans ce sens au niveau mondial (par exemple, le workshop *AAAI WS : « AI and Operations Research for Social Good »* en 2017, ou le « *AI for Good Global Summit* » à Genève en 2017). Ainsi émerge un nouveau courant de recherche, l'*informatique prosociale*, qui se définit comme « *l'informatique visant à soutenir et à promouvoir des actions qui profitent à la société et aux autres* » (Paiva *et al.*, 2018). Le comportement des systèmes interactifs humanoïdes peut en effet influencer le comportement humain. Par exemple, des travaux de recherche ont montré que le simple regard d'un robot peut déclencher un comportement altruiste chez l'humain (Burnham et Hare, 2007). L'empathie, un élément essentiel au comportement prosocial, a été largement explorée par plusieurs chercheurs en informatique afin d'être intégrée dans les systèmes interactifs humanoïdes (Ochs *et al.*, 2008; Paiva *et al.*, 2017).

2. Les systèmes virtuels comme vecteurs d'empathie

La recherche en informatique pro-sociale permet à la fois de promouvoir les comportements pro-sociaux mais aussi d'améliorer la compréhension de leurs mécanismes (Paiva *et al.*, 2018). L'un de ces mécanismes est l'*amorçage conceptuel*. En 2013, Rosenberg *et al.* ont proposé à des participants d'incarner des superhéros virtuels capables de voler. Selon leurs analyses, les pouvoirs virtuels ont mené les participants à produire plus de comportements pro-sociaux une fois sortis de la simulation. L'une des explications serait que l'amorçage implicite de concepts généralement associés aux superhéros (force, courage, empathie) inciteraient les individus à adhérer et produire les comportements qui s'y rapportent. Ces recherches montrent que l'incarnation de personnages virtuels dans un environnement virtuel peut permettre d'activer des concepts chez les utilisateurs qui auront une influence directe sur leurs comportements dans le monde réel. Dans une perspective de remédiation de la discrimination et en se fondant sur les résultats de ces recherches, nous pouvons supposer que l'incarnation de l'utilisateur dans un agent virtuel pourrait permettre l'activation de concepts stéréotypiques et donc une prise de conscience des comportements discriminants..

Plusieurs travaux de recherche ont d'ores et déjà exploré l'utilisation des agents virtuels pour lutter contre les discriminations ethniques, liées à l'âge ou encore au genre. Nous en présentons quelques exemples ci-dessous.

A. *Discriminations raciales*

Les travaux fondateurs de la lutte contre la discrimination proviennent de la prise de conscience de stéréotypes et de biais produits envers les groupes racisés tels que la communauté afro-américaine.

En 2009, Groom *et al.* ont proposé à des participants d'incarner un personnage virtuel blanc ou noir. Les résultats montrent que, contrairement à l'hypothèse d'une réduction des

biais raciaux implicites suite à l'incarnation d'un personnage noir, l'incarnation a réactivé les stéréotypes raciaux.

Les travaux de Peck *et al.* (2013) contredisent pourtant ces résultats : après avoir virtuellement incarné des avatars noirs («*body ownership*»), les participants présentent des niveaux de biais raciaux implicites inférieurs à ceux avant la simulation. L'une des critiques adressées aux recherches de Groom *et al.* (2009) concerne la durée et la qualité de l'expérience d'incarnation : contrairement à la condition de Peck *et al.* (2013) où l'incarnation concernait le corps entier des participants, seuls les mouvements de tête étaient synchronisés avec le participant; les durées d'incarnation étaient également différentes (environ 65 secondes contre 11 minutes 30). L'immersion induite par la simulation occupe donc une place importante dans le processus de réduction des biais implicites (Section 3).

B. Discriminations par l'âge

L'*âgisme* est un terme proposé par Butler (1969) pour décrire les discriminations liées à l'âge. Du fait de l'accroissement de la longévité humaine, celles-ci sont de plus en plus fréquentes.

Selon Galinsky *et al.* (2005), l'une des composantes centrales de la prise de perspective est le *chevauchement soi-autrui* (*self-other overlap*). Ce chevauchement illustre la proximité perçue par un individu entre ses représentations mentales et celles d'autrui. En 2016, Oh *et al.* ont montré qu'incarner une personne âgée dans un environnement virtuel permet de réduire le *chevauchement soi-autrui*, et concourt ainsi à renforcer le désir de communiquer avec des membres de l'exogroupe. En revanche, les résultats ne montrent pas de réduction des biais négatifs. Cet effet peut s'expliquer par l'utilisation d'une mesure explicite des préjugés (*explicit ageism*), soumettant les participants à un biais de désirabilité sociale : conscients de la stigmatisation sociale des attitudes discriminantes, ces derniers sont plus enclins à fournir des réponses biaisées.

Il existe encore très peu d'études portant sur la remédiation de l'âgisme grâce à des agents virtuels, ce qui rend difficile toute comparaison avec d'autres mesures. L'utilisation de mesures implicites des discriminations semblent cependant être une alternative intéressante pour pallier les biais de désirabilité sociale.

C. Discriminations de genre

Le sexisme est fondé sur des discriminations spécifiques au sexe, et par extension au genre. En 2019, Lopez *et al.* ont mené des travaux dans des environnements virtuels immersifs pour étudier le rôle de l'incarnation dans la réduction des biais implicites liés au genre. Contre toute attente, les résultats montrent une augmentation des biais sexistes chez les participants ayant incarné un avatar féminin. Comme dans le cas de Groom (2009), l'immersion semble avoir entraîné une réactivation des stéréotypes.

En 2018, Aitamurto *et al.* ont testé les implications des discriminations de genre dans un système virtuel constitué d'une Split-Sphere 360° : le participant se retrouve immergé dans une capsule vidéo scindée en deux. D'une part, la simulation le place

dans la peau d'un homme; d'autre part, dans la peau d'une femme. Les résultats montrent que l'immersion dans deux mondes parallèles crée un sentiment de malaise chez le participant, exprimé par le sentiment de *Fear Of Missing Out (FOMO)*, c'est-à-dire la peur de rater une information importante durant la simulation. Cependant, l'un des avantages de ce dispositif est le contraste frappant qu'offre les deux vidéos en simultanée. Comme le dit l'un des participants à la fin de l'expérience : «Être capable de voir deux scènes fortement distinctes en même temps accentue les différences entre les perspectives de la femme et celle de l'homme» (*«Being able to see two highly contrasting scenes at the same time. It emphasized the differences between the perspective of the man and the perspective of the woman.»*).

Les travaux de recherche utilisant les agents virtuels pour lutter contre les discriminations restent encore très peu nombreux. Aujourd'hui la réalité virtuelle immersive incarnée (VRI) offre la possibilité de remplacer le corps d'une personne par un corps virtuel et donner ainsi l'illusion et le sentiment subjectif d'avoir un autre corps. L'utilisateur incarne le corps d'un avatar qui imite ses mouvements en temps réel. Des travaux récents ont montré le pouvoir de cette illusion d'avoir un autre corps en réalité virtuelle sur le développement de l'empathie (Bertrand *et al.*, 2018) et sur les stéréotypes ethniques (Peck *et al.*, 2013; Banakou *et al.*, 2016). Cette technique de réalité virtuelle pourrait être utilisée pour d'autres types de discrimination tels que les discriminations de genre ou la grossophobie.

3. Induction de comportements empathiques : leviers d'influence

La capacité d'un agent virtuel ou d'un environnement de réalité virtuelle à remédier à la discrimination peut dépendre de différents facteurs. En particulier, la *présence*, la *co-présence* et la *crédibilité* de la simulation sont des facteurs particulièrement influents que nous détaillons ci-dessous.

L'immersion est le fait d'être «*absorbé dans un monde virtuel en quittant le monde réel*» (Ris *et al.*, 2018). Le participant a la sensation de vivre une expérience dans un environnement différent de celui où il se situe physiquement : il s'agit du *sentiment de présence*. Dans le cadre des environnements virtuels, la présence se définit comme la sensation d'être ancré dans la simulation alors même que le corps y est physiquement absent. Comme le rapportent les participants, le sentiment d'immersion est fortement lié au sentiment de présence.

Youngblut (2003) définit la *co-présence* comment le sentiment d'appartenir à un groupe d'individus (êtres humains ou agents virtuels) immergés dans un environnement virtuel. La *co-présence* se distingue de la présence sociale par la dimension cognitive de cette dernière : en plus de sentir la présence d'un autre individu (réel ou virtuel), on attribue à ce dernier un certain nombre d'états cognitifs et affectifs, comme des émotions ou des intentions.

La *crédibilité* des agents virtuels (en termes de réalisme et de comportements verbal et non-verbal) joue également un rôle

important dans le sentiment de co-présence. En 2005, Bailenson *et al.* ont montré qu'un décalage entre différents niveaux de réalisme d'un agent virtuel était corrélé à une baisse du sentiment co-présence éprouvé par les participants : par exemple, un agent virtuel aux réactions comportementales adéquates, mais à l'aspect visuel rudimentaire provoquera un désengagement dans la simulation.

Comme le soulignent Schutte et Stilinović (2017), la *co-présence* et la *présence sociale* sont corrélées au sentiment d'immersion et par là-même à l'induction de comportements empathiques. Au plus leurs niveaux sont élevés, au plus l'individu plongé dans l'environnement virtuel aura l'impression d'interagir dans une scène de la vie réelle.

Selon Held et Durlach (1992), le degré d'immersion est directement lié à la richesse de la simulation. L'intégration de la modalité haptique (le toucher) dans des environnements virtuels partagés par plusieurs individus (Basdogan *et al.*, 2000) permet d'améliorer les performances des participants et participe au sentiment d'unité dans l'environnement virtuel (« *sense of togetherness*»). L'intégration du "toucher" pourrait être utilisé dans les travaux de lutte contre la discrimination en environnement virtuel pour intensifier l'expérience immersive et ainsi amplifier ses effets.

III. VERS UN CHANGEMENT DURABLE DES COMPORTEMENTS ?

La question centrale des travaux de lutte contre la discrimination concerne la viabilité des comportements pro-sociaux : est-il possible de changer durablement les comportements ? Comment maintenir un comportement pro-social hors de la simulation ? Les travaux de Herrera *et al.* (2018) ont permis de montrer qu'une expérience de mise en situation en réalité virtuelle (*Immersive Virtual Environment, IVE*) produit une augmentation de l'empathie à long terme plus importante qu'avec une méthode traditionnelle, où il est demandé au participant d'imaginer la mise en situation (*Mental Simulation, MS*).

Dans cette dernière partie, nous nous concentrons sur les différents types de mesures utilisées pour quantifier l'impact des systèmes interactifs sur les comportements et l'état cognitif des participants. Ces mesures sont de trois types : *mesures de l'empathie*, *mesures des facteurs d'influence* et *mesures des comportements discriminatoires*.

1. Mesures de l'empathie

Fortement dépendante de la dimension immersive des environnements virtuels et de la fluidité des interactions avec les agents virtuels, l'empathie se pose comme une mesure fiable de la modification des attitudes discriminantes. Les méthodes d'évaluation traditionnelles de cette dernière sont centrées sur l'observation et les questionnaires auto-évaluatifs (Carey *et al.*, 2017). Parmi ces méthodes, la plus couramment utilisée est l'Indice

de Réactivité Interpersonnelle (*Interpersonal Reactivity Index, IRI*).

L'IRI est un test développé par Mark H. Davis (1980) pour mesurer l'empathie sous différents aspects à travers un ensemble de questions. Selon l'auteur, l'empathie se construit à travers un ensemble de facteurs reliés les uns aux autres et qui se divisent en quatre parties : l'adaptation contextuelle (*perspective-taking*), c'est-à-dire la capacité à adopter un point de vue extérieur ; le souci empathique (*empathy concern*), soit la capacité à éprouver des sentiments de compassion pour autrui ; la détresse personnelle (*personal distress*), qui s'exprime par la tendance à ressentir de la détresse au contact de la détresse d'un autre individu ; enfin, le fantasme (*fantasy*), qui renvoie à la capacité à adopter les états mentaux et émotionnels de personnages fictifs. Ce dernier aspect de l'empathie est particulièrement important lorsque l'utilisateur interagit avec des agents virtuels : une faible prédisposition à une empathie *fantasy* peut empêcher l'induction d'empathie chez l'utilisateur.

2. Mesures des facteurs d'influence

L'induction de comportements pro-sociaux tels que l'empathie est sensible à de nombreux paramètres. Au sein d'un environnement virtuel se pose la question de l'immersion du participant : il semble qu'une immersion importante soit liée au niveau de crédibilité de l'environnement virtuel. En 2018, Shin a montré qu'il existe une forte corrélation entre le degré d'immersion du participant dans la simulation et sa capacité à produire des comportements empathiques. Nous recensons ici trois types de questionnaire permettant de mesurer le sentiment d'immersion du participant : le *simulator sickness questionnaire*, le questionnaire de présence et le questionnaire de co-présence.

A. Simulator sickness questionnaire

Le cyber-malaise (*cyber-sickness*) s'apparente au mal des transports et est produit suite à l'immersion dans un environnement virtuel. Ce malaise provient d'un conflit entre les systèmes vestibulaire, visuel et proprioceptif : le système visuel perçoit un mouvement alors que le corps est quasiment immobile, créant ainsi un décalage perceptif (Stanney *et al.*, 1999). Le cyber-malaise a un impact négatif sur l'expérience virtuelle du participant, notamment en diminuant son sentiment d'immersion (Weech *et al.*, 2019). L'une des méthodes de mesure du cyber-malaise est le *Simulator Sickness Questionnaire (SSQ)*. Le questionnaire se compose d'indices du niveau de nausée, de désorientation et de ressenti oculomoteur. La somme des indices donne un score total de sévérité du cyber-malaise.

B. Questionnaire de présence

Les questionnaires de présence (*Presence Questionnaire, PQ*) sont composés d'un ensemble de dimensions permettant de mesurer l'immersion du participant : s'y retrouve la *présence spatiale* (sentiment d'habiter physiquement l'espace virtuel),

l'*implication* (attention mobilisée pendant l'expérience) ou encore le *réalisme perçu* (mesure subjective du réalisme de la simulation). Selon Schubert (2003), ces trois dimensions sont essentielles à l'ancrage du participant dans l'environnement virtuel et participent à son immersion. Plusieurs questionnaires subjectifs ont été développés pour mesurer le sentiment de présence de l'utilisateur comme l'*IGroup Presence Questionnaire* (IPQ) incluant 14 questions (Schubert et al., 2001).

C. Questionnaire de co-présence

Les questionnaires de co-présence mesurent le sentiment d'appartenance à un groupe dans un environnement virtuel. La version de Poeschl et Doering (2015) se décompose en quatre facteurs : la réaction du participant face aux agents virtuels, la réaction des agents virtuels perçue par le participant, le ressenti des possibilités d'interactions et la co-présence des autres participants (humains ou virtuels). Les mesures de présence et co-présence sont souvent regroupées et analysées comme des facteurs interdépendants. Dans (Bailenson et al., 2005) est proposé un questionnaire pour mesurer le sentiment de co-présence d'un utilisateur envers un agent virtuel.

3. Comportements discriminants : outils de mesure

Les mesures des comportements discriminatoires s'effectuent généralement sous forme d'échelles ou de questionnaires auto-évaluatifs. Ces mesures tentent de minimiser les facteurs d'influence provenant des normes culturelles, telles que la désirabilité sociale. Elles reposent pour cela sur des évaluations implicites. Nous proposons d'illustrer certaines d'entre elles à travers leur utilisation en environnement virtuel.

A. Test d'association implicite

Le test d'association implicite (*Implicit-Association Test*, IAT) se base sur le postulat que les individus répondent généralement de manière moins objective lorsque les sujets abordés sont enclins à susciter un biais de désirabilité sociale ou que les sujets ne sont pas conscients de leurs opinions et comportements. Créé en 1998 par Greenwald, McGhee et Schwartz, il permet de mesurer les stéréotypes négatifs à travers les associations automatiques d'idées (généralement implicites) et sert de base explicative aux comportements discriminants. Le test repose sur une classification de concepts et une évaluation de la rapidité d'association de paires de concepts. Prenons l'exemple d'une triade de concepts « bon », « vieux », « jeune ». Le participant doit associer ces catégories par paires : selon la théorie des associations implicites, si les catégories « vieux » et « bon » sont reliées moins vite que les catégories « jeune » et « bon », le participant a tendance à avoir des attitudes plus négatives envers les personnes âgées. Dans les travaux de Groom (2009), les auteurs utilisent l'IAT pour mesurer les biais raciaux des participants lorsqu'ils incarnent un avatar ayant une couleur de peau différente de la leur. Enfin, un test distinct a été créé pour mesurer les associations sexistes implicites

: le Gender IAT. Ce test est utilisé comme mesure indirecte du sexisme en réalité virtuelle dans (Salmanowitz, 2018).

B. Distance interpersonnelle

Les mesures comportementales sont une alternative robuste aux questionnaires. Elles permettent de mesurer l'influence sociale dans les environnements virtuels en se libérant des biais subjectifs que le participant entretient concernant ses propres attitudes. Pour prendre en compte des indices spatiaux, la distance interpersonnelle semble être une mesure fiable des attitudes implicites. L'hypothèse centrale est qu'une attitude discriminatoire envers un individu distinct s'exprimera par une augmentation de la distance interpersonnelle entre cet individu et le participant. Placés dans un monde virtuel, les êtres humains interagissent avec leur environnement de façon écologique : plusieurs études (Bailenson, 2003) montrent en effet que l'espace personnel autour des participants est respecté. Peu de participants traversent leurs partenaires virtuels, alors même que les lois physiques de la simulation le permettent.

C. Échelle d'argument racial

Développée par Saucier et Miller (2003), l'échelle d'argument racial (*Racial Argument Scale*, RAS) permet de mesurer indirectement les préjugés des participants envers une minorité discriminée. Cette échelle se base sur une évaluation de la pertinence d'arguments en faveur de conclusions positives ou négatives envers les individus noirs. Les études montrent que l'échelle d'argument racial est une mesure consistante qui présente une forte corrélation avec des mesures d'autoévaluation du racisme (Saucier et Miller, 2003; Heitland et Bonher, 2011). L'échelle d'argument racial est notamment utilisée dans la mesure des discriminations en simulation virtuelle dans (Groom, 2009).

D. Échelle moderne de racisme

L'échelle moderne de racisme (McConahay, 1986) a été conçue pour mesurer la dimension cognitive des stéréotypes raciaux. Elle se base sur des croyances entretenues par des personnes blanches envers des individus racisés et distingue deux ensembles de croyances : les croyances modernes et dépassées (Swim et al., 1995). En effet, les auteurs considèrent que la nature des préjugés est soumise au contexte historique dans lequel les attitudes et croyances se forment.

IV. CONCLUSION

Le recours à une intelligence artificielle (incarnée par des agents virtuels) pour pallier un comportement humain (la discrimination) semble presque paradoxal. Les travaux en informatique pro-sociale prouvent pourtant que ces systèmes interactifs ont un impact positif sur l'induction de comportements pro-sociaux. Les travaux montrent que la simulation de situations de discrimination, en plaçant l'utilisateur dans le rôle d'une victime à travers l'incarnation d'un personnage virtuel, permet de réduire les stéréotypes par l'induction d'empathie. C'est par ailleurs le

sentiment accru d'immersion qui distingue l'utilisation d'agents virtuels d'autres méthodes de sensibilisation à la discrimination. En poussant le fictionnel jusqu'aux limites de la réalité, l'interaction avec une intelligence virtuelle pro-sociale permet un engagement plus fort de l'utilisateur que ne le permet le recours à des fictions moins immersives (telles que le théâtre, le cinéma ou encore la littérature).

Différents facteurs, propres à l'environnement virtuel mais aussi à l'utilisateur, peuvent influencer l'expérience immersive et ainsi les effets de tels systèmes : nous y retrouvons le sentiment de présence et de co-présence, mais aussi la propension de l'utilisateur à ressentir de l'empathie envers des personnages fictifs. Pour quantifier l'effet des systèmes sur les comportements discriminants des utilisateurs ou sur ses stéréotypes, plusieurs mesures subjectives et objectives peuvent être utilisées telles que les tests d'associations implicites, les échelles ou encore la distance interpersonnelle dans les simulations virtuelles. Malgré le potentiel que peut offrir de tels systèmes pour lutter contre les discriminations sociales dans le monde réel, les travaux de recherche sur ce sujet restent encore très peu nombreux. Certains types de discrimination ne sont quasiment pas explorés (comme la grossophobie) et les facteurs d'influence sont encore peu connus : qu'en est-il de l'influence de la personnalité, de l'âge, du genre ou de la culture de l'utilisateur ?

À l'heure actuelle, pour lutter contre les stéréotypes à travers les agents virtuels, la plupart des systèmes étudiés proposent à l'utilisateur d'incarner *la victime de discrimination* dans un environnement virtuel. Former les témoins des scènes de discrimination à réagir pourrait être une autre approche pour lutter contre les discriminations sociétales. Elle permettrait à la fois de sensibiliser les individus en les formant à identifier des situations de discrimination et de s'entraîner à réagir dans un environnement virtuel, avant d'y être potentiellement confrontés dans la vie réelle. De plus, former les témoins plutôt que les auteurs de discriminations permettrait de toucher un plus grand nombre d'individus.

V. BIBLIOGRAPHIE

- Abroms, L. C., & Maibach, E. W. (2008). The Effectiveness of Mass Communication to Change Public Behavior. *Annual Review of Public Health, 29*(1), 219- 234.
- Albrecht, K. (2006). *Social intelligence: The new science of success*. John Wiley & Sons.
- Aitamurto, T., Zhou, S., Sakshuwong, S., Saldivar, J., Sadeghi, Y., & Tran, A. (2018). Sense of Presence, Attitude Change, Perspective-Taking and Usability in First-Person Split-Sphere 360° Video. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 1- 12.
- Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley.
- Aylett, R., Vala, M., Sequeira, P., & Paiva, A. (2007). Fearnot!—an emergent narrative approach to virtual dramas for anti-bullying education. In International Conference on Virtual Storytelling (pp. 202-205). Springer, Berlin, Heidelberg.
- Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2003). Interpersonal Distance in Immersive Virtual Environments. *Personality and Social Psychology Bulletin, 29*(7), 819- 833.
- Bailenson, J. N., Swinth, K., Hoyt, C., Persky, S., Dimov, A., and Blascovich, J. (2005). The independent and interactivity effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of co presence in immersive virtual environments. *Presence: Teleoperators and Virtual Environments, 14*(4):379–393.
- Banakou, D., Hanumanthu, P. D., & Slater, M. (2016). Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Frontiers in human neuroscience, 10*, 601.
- Baron, R. A., & Branscombe, N. R. (2011). *Social Psychology (13th Edition)* (13e éd.). Kansas, USA : Pearson.
- Basdogan, C., Ho, C.-H., Srinivasan, M. A., & Slater, M. (2000). An experimental study on the role of touch in shared virtual environments. *ACM Transactions on Computer-Human Interaction, 7*(4), 443- 460.
- Batson, C. D., & Ahmad, N. Y. (2009). Using Empathy to Improve Intergroup Attitudes and Relations. *Social Issues and Policy Review, 3*(1), 141- 177.
- Belk R.W. (1974), An Exploratory Assessment of Situational Effects in Buyer Behavior, *Journal of Marketing Research, 11*, 156-163.
- Bertrand, P., Guegan, J., Robieux, L., McCall, C. A., & Zenasni, F. (2018). Learning empathy through virtual reality: multiple strategies for training empathy-related abilities using body ownership illusions in embodied virtual reality. *Frontiers in Robotics and AI, 5*, 26.
- Boven, L. V., Loewenstein, G., Dunning, D., & Nordgren, L. F. (2013). Changing Places. *Advances in Experimental Social Psychology, 117- 171*.
- Bruijnes, M., Linssen, J., & Heylen, D. (2019). Special issue editorial: Virtual Agents for Social Skills Training. *Journal on Multimodal User Interfaces, Volume 13, Issue 1*.
- Burnham, T. C., & Hare, B. (2007). Engineering Human Cooperation. *Human Nature, 18*(2), 88- 108.
- Butler, R. N. (1969). Age-Isms : Another Form of Bigotry. *The Gerontologist, 9*(4 Part 1), 243- 246.
- Campbell, R. G., & Babrow, A. S. (2004). The role of empathy in responses to persuasive risk communication: Overcoming resistance to HIV prevention messages. *Health Communication, 16*, 159-182.
- Carey, K., Saltz, E., Rosenbloom, J., Micheli, M., Choi, J. O., & Hammer, J. (2017). Toward Measuring Empathy in Virtual Reality. *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play, 551- 559*.
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog Select. Doc. Psychol. 10:85*.
- Decety, J., & Jackson, P. L. (2004). The Functional Architecture of Human Empathy. *Behavioral and Cognitive Neuroscience Reviews, 3*(2), 71- 100.
- Fiske, S. (1998), Stereotyping, prejudice, and discrimination, in D. Gilbert, S. Fiske & G. Lindzey, 775 eds, 'Handbook of Social Psychology', Boston: McGraw-Hill, pp. 357-411.
- Galinsky, A. D., Ku, G., & Wang, C. S. (2005). Perspective-Taking and Self-Other Overlap : Fostering Social Bonds and Facilitating Social Coordination. *Group Processes & Intergroup Relations, 8*(2), 109- 124.
- Gasquet-Cyrus, M. (2016). Philippe Blanchet Discriminations : combattre la glottophobie 2016, Paris, Textuel, coll. Petite Encyclopédie critique, 192 p. *Langage et société, 2*(2), 133-136.
- Goffman, E. (1988). L'ordre social et l'interaction, Seuil/Minuit, Paris (textes recueillis et présentés par Yves Winkin), p. 95-103
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition : The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464- 1480.
- Groom, V., Bailenson, J. N., & Nass, C. (2009). The influence of racial embodiment on racial bias in immersive virtual environments. *Social Influence, 4*(3), 231- 248.
- Guimond, S. (2000). Group socialization and prejudice: The social transmission of intergroup attitudes and beliefs. *European Journal of Social Psychology, 30*(3), 335–354.

- Hall, Lynne, et al. "Fostering empathic behaviour in children and young people: interaction with intelligent characters embodying culturally specific behaviour in virtual world simulations." INTED2011 Proceedings (2011): 2804-2814.
- Heitland, K., Bohner, G. An Indirect Paper-and-Pencil Measure of Prejudice: A German Version of the Racial Argument Scale. *Curr Psychol* 30, 1–21 (2011).
- Herrera F, Bailenson J, Weisz E, Ogle E, Zaki J (2018) Building long-term empathy: A large scale comparison of traditional and virtual reality perspective-taking. *PLoS ONE* 13(10): e0204494.
- Hoffmann, L., Krämer, N. C., Lam-chi, A., & Kopp, S. (2009). Media Equation Revisited : Do Users Show Polite Reactions towards an Embodied Agent ? *Intelligent Virtual Agents*, 159- 165.
- Janezic, I. (2015). *L'empathie comme variable médiatrice de persuasion* (Vol.18, n.1). Consulté à <http://www.composite.org/index.php/revue/article/view/213/173>
- Jones, E. E., & Nisbett, R. E. (1987). *The actor and the observer: Divergent perceptions of the causes of behavior*. In E. E.
- Kopp, S., Gesellensetter, L., Krämer, N. C., Wachsmuth, I.: A conversational agent as museum guide - design and evaluation of a real-world application, in: Panayiotopoulos, T. et al. (Eds.), *Intelligent Virtual Agents 2005*. Springer, Hamburg (2005) 329-343
- Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy : reliability and validity of the Empathy Quotient. *Psychological Medicine*, 34(5), 911- 920.
- Lopez, S., Yang, Y., Beltran, K., Kim, S. J., Cruz Hernandez, J., Simran, C. Yuksel, B. F. (2019). Investigating Implicit Gender Bias and Embodiment of White Males in Virtual Reality with Full Body Visuomotor Synchrony. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1- 12.
- McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91–125). London: Academic Press.
- McGuire W. J. (1964). Inducing resistance to persuasion: some contemporary approaches, in *Advances in Experimental Social Psychology, Vol. 1*, ed Berkowitz L. (New York, NY: Academic Press;), 191–229.
- Moscovici, S. (1972). Theory and society in social psychology. In J. Isreal & H. Tajfel (Eds.), *The context of social psychology: A critical assessment*. London: Academic.
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81–103.
- Ochs, M., Pelachaud, C., & Sadek, D. (2008, May). An empathic virtual dialog agent to improve human-machine interaction. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1* (pp. 89-96). International Foundation for Autonomous Agents and Multiagent Systems.
- Oh, Soo Youn; Bailenson, Jeremy; Weisz, Erika; Zaki, Jamil (2016). Virtually old: Embodied perspective taking and the reduction of ageism under threat. *Computers in Human Behavior*, 60(), 398–410.
- Paiva, A., & Santos, F. (2018). Engineering Pro-Sociality With Autonomous Agents. *AAAI*.
- Paiva, A., Leite, I., Boukricha, H., & Wachsmuth, I. (2017). Empathy in virtual agents and robots: A survey. *ACM Transactions on Interactive Intelligent Systems (TiIS)*, 7(3), 11.
- Peck, T. C., Seinfeld, S., Aglioti, S. M., & Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and Cognition*, 22(3), 779- 787.
- Poeschl, S., Doering N., 2015. Measuring co-presence and social presence in virtual environments - Psychometric construction of a german scale for a fear of public speaking scenario. *Annual Review of CyberTherapy and Telemedicine* 13, February 2016 (2015), 58–63.
- Reny Delisle, M. (2013). *L'empathie virtuelle comme facteur d'influence de la réception de publicités sociales : développement d'un outil de mesure* (Thèse). Consulté à <https://corpus.ulaval.ca/jspui/handle/20.500.11794/24468>
- Ris, C., Renoult, P., & Ghorra, J. (2018). *De la présence à l'immersion dans les environnements virtuels*. Consulté à https://idmc.univ-lorraine.fr/wp-content/uploads/2018/06/IDMC-2017_2018-ProjetTutore-RAPPORT-_Ris_Renoult_Ghorra.pdf
- Rosenberg, R. S., Baughman, S. L., & Bailenson, J. N. (2013). Virtual Superheroes : Using Superpowers in Virtual Reality to Encourage Prosocial Behavior. *PLoS ONE*, 8(1), e55003.
- Salmanowitz, N. (2018). The impact of virtual reality on implicit racial bias and mock legal decisions. *Journal of Law and the Biosciences*, 5(1), 174- 203.
- Saucier, D. A., & Miller, C. T. (2003). The Persuasiveness of Racial Arguments as a Subtle Measure of Racism. *Personality and Social Psychology Bulletin*, 29(10), 1303- 1315.
- Schubert, T. W. (2003). The sense of presence in virtual environments : *Zeitschrift für Medienpsychologie*, 15(2), 69- 71.
- Schutte, N. S., & Stilinović, E. J. (2017). Facilitating empathy through virtual reality. *Motivation and Emotion*, 41(6), 708- 712.
- The structure of the Life-World* (édition-traduction posthume de *Die Strukturen der Lebenswelt*, à partir du manuscrit transmis par Ilse Schütz), traduction de Richard M. Zaner et H. Tristram Engelhardt Jr., Londres, Heinemann, 1974.
- Shamay-Tsoory, S. G. (2010). The Neural Bases for Empathy. *The Neuroscientist*, 17(1), 18- 24.
- Shin, D. (2018). Empathy and embodied experience in virtual environment : To what extent can virtual reality stimulate empathy and embodied experience ? *Computers in Human Behavior*, 78, 64- 73.
- Schubert, T., Friedmann, F., and Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators and virtual environments*, 10(3):266–281
- Stanney, K. M., Kennedy, R. S., Drexler, J. M., & Harm, D. L. (1999). Motion sickness and proprioceptive aftereffects following virtual environment exposure. *Applied Ergonomics*, 30(1), 27- 38.
- Stephan, W. G., & Finlay, K. (1999). The Role of Empathy in Improving Intergroup Relations. *Journal of Social Issues*, 55(4), 729- 743.
- Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism : Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, 68(2), 199- 214.
- Upshaw, M. B., Kaiser, C. R., & Sommerville, J. A. (2015). Parents' empathic perspective taking and altruistic behavior predicts infants' arousal to others' emotions. *Frontiers in Psychology*, 06, 0.
- Weech, S., Kenny, S., & Barnett-Cowan, M. (2019). Presence and Cybersickness in Virtual Reality Are Negatively Related: A Review. *Frontiers in psychology*, 10, 158.
- Yee, N., & Bailenson, J. (2007). The Proteus Effect : The Effect of Transformed Self-Representation on Behavior. *Human Communication Research*, 33(3), 271- 290.
- Youngblut, C. (2003). *Experience of Presence in Virtual Environments*. Institute for Defense Analyses, Alexandria, Virginia D-2960.