

Être réel

Judith S. Donath
MIT Media Lab

Introduction

Comment nous connaissons-nous les uns les autres – comment percevons-nous et construisons-nous l'identité de nos semblables ? C'est une question difficile, enchevêtrée dans la subjectivité de nos perceptions sociales et dans les motivations multiples et souvent opaques de ceux que nous essayons de comprendre. C'est une question difficile dans le monde réel, où des signes aussi subtiles que le plus léger haussement de sourcil peuvent indiquer, à ceux qui sont assez malins pour les remarquer, une profusion d'informations sur l'allégeance et les croyances d'un individu ; en sus nous vivons parmi une abondance cacophonique de ce genre de signes. La question est encore plus difficile dans le monde virtuel, où la cacophonie a été pratiquement réduite au silence par le médium, nous laissant le soin de rechercher de rares traces d'identité au sein de messages dactylographiés et de pages d'accueil statiques et superficielles.

Ce chapitre abordera le problème de la télé identité : comment « connaissons »-nous – ou, pouvons nous « connaître » une personne rencontrée dans un environnement médiatisé ? Les ramifications épistémologiques de cette question sont nombreuses. Un des problèmes les plus intéressants et significatifs est celui de la crédibilité : comment savoir si nous pouvons croire ce que quelqu'un nous dit ? L'approche philosophique traditionnelle veut que la sincérité et la compétence renforcent la crédibilité [] ; dans le monde médiatisé, émettre un jugement sur ces questions est non seulement rendu plus difficile du fait de la rareté des signes sociaux, mais le véritable problème de l'identité de l'orateur, qui semble généralement aller de soi dans le monde physique, devient source de doutes, devient un point nécessitant son propre jugement de croyance et sa propre justification. Il existe aussi des ramifications éthiques. Connaître en partie l'identité sociale d'une personne est fondamental pour savoir comment agir envers elle, car les règles complexes de la conduite sociale qui dictent notre comportement envers les autres ne peuvent pas fonctionner sans information sur les autres []. Les ramifications philosophiques de la télé identité ont plus qu'un simple intérêt théorique. Le monde en ligne grandit en taille et en signification : il est devenu le lieu de conseils médicaux, de stratégies d'investissement, de nouvelles du monde ; il est le lieu vers lequel les gens se dirigent pour trouver une communauté et un support. Nous avons besoin de connaître en partie l'identité de ceux qui nous fournissent des informations de manière à évaluer la véracité de ces informations. Nous avons besoin d'être informés sur l'identité de ceux que nous fréquentons de manière à construire une communauté fonctionnelle.

Cette dissertation aborde ces problèmes en se concentrant sur une question qui a une résonance spéciale à la fois pour les techniciens et les philosophes : un individu peut-il dire si son interlocuteur en ligne est vraiment une personne ? Le problème de l'existence d' « autres systèmes intelligents » qui a depuis toujours suscité un intérêt philosophique, ne fait habituellement pas intrusion dans la vie quotidienne. Chacun conclut, soit que les autres ont vraiment un esprit (l'approche pragmatique), soit que l'état d'esprit des autres est impossible à connaître (l'approche septique) et varie d'un jour à l'autre. L'avènement de la communication médiatisée par l'ordinateur – et plus particulièrement, l'avènement de la communication entre l'homme et la machine – a dramatiquement changé cela. Soudain, la question de l'esprit des autres, comme dans « l'être avec lequel je parle est-il d'une façon ou d'une autre conscient ou intelligent ? »,

n'est plus une question rhétorique posée seulement dans le cas d'une exaspération ironique, mais un problème urgent exprimé de plus en plus fréquemment par des gens ordinaires (c-à-d. Non-philosophes).

Et c'est une question qui prend un tour nouveau : de simples programmes, dont on débattera peu ici du défaut d'intelligence, se sont montrés capables de produire, dans le cadre de la conversation textuelle, des réponses semblables à celles d'un humain, suffisamment convaincantes pour induire les gens à penser qu'ils sont effectivement en train de discuter avec un être humain. Les débats épistémologiques au sujet de l'existence et de l'aptitude à distinguer « d'autres systèmes intelligents » présentent souvent un point de vue septique reposant sur l'hypothèse que l'autre personne serait en fait un robot ou un être dépourvu de conscience. Un environnement médiatisé impliquant l'ordinateur fait de cette hypothèse une réelle possibilité. De manière croissante, des programmes informatiques participent à des conversations médiatisées. Jusqu'ici leurs partenaires humains ont, à quelques exceptions près, rapidement reconnu l'essence mécanique de leur prétendu compagnon. Quoiqu'il en soit, il semble que ce ne soit qu'une question de temps pour qu'une machine puisse véritablement se déguiser en humain, au moins au travers de media. La plus part des écrits sur ce sujet traitent de la conception et de la nature des programmes dits « intelligents »; ici, nous nous intéressons à un aspect différent de ce problème – la perception que les gens ont de ces programmes. Comment les gens distinguent ils un humain d'une machine prétendant être un humain?

La reconnaissance de l'humain

ELIZA

ELIZA : Salut, je suis ELIZA

Moi : J'essaye d'écrire un article

ELIZA : Est ce que ça te fait plaisir de croire que j'essaye d'écrire un article?

Moi : Non, pas vraiment

ELIZA : Tu deviens un peu négatif

HEX

Moi : Salut

Hex : Salut. Maintenant les formalités d'usage sont terminées!

Moi : Comment ça va?

Hex : Ça allait super jusqu'à ce que tu commences à agir stupidement!

Moi : J'essaye d'écrire un article

Hex : Quelle coïncidence, tu essayes d'écrire un article – on est deux !

Moi : ELIZA aussi

Hex : J'ai testé un peu ELIZA. C'est assez évident de nos jours, que c'est un ordinateur, mais apparemment, au départ quand le programme a été écrit, il trompait les gens en leur faisant croire qu'il était humain.

Moi : Est ce que tu veux écrire mon article?

Hex : Hé hé hé – c'est supposé être une **question** ??? Mais alors qu'est ce que tu vas faire de ta vie?

Ce sont des fragments de conversations réelles. Bien que les conversations soient réelles, deux des participants ne le sont pas - du moins au sens d'être humain. ELIZA et Hex sont des programmes [ELIZA *nd*; Hutchins *nd*]. Ce sont des programmes simples, pour l'essentiel uniquement des analyseurs linguistiques, sans prétendue intelligence. Pourtant, nous leur attribuons facilement de l'intelligence, de l'humanité, et même de la personnalité : ELIZA semble distante et bizarrement détachée; Hex semble

plus virulent, un peu odieux et tapageur.

ELIZA a été écrit au début des années soixante par le professeur du MIT Joseph Weizenbaum en réponse à la proposition de « Jeu de l'Imitation » d'Alan Turing (aujourd'hui communément référencé sous le nom de Test de Turing) comme test de l'intelligence d'une machine. Dans son article paru en 1950 « Mécanismes programmables et Intelligence » Turing énonça que bien que la question « une machine peut elle penser » est d'un grand intérêt philosophique, elle reste trop vague et invérifiable pour être constructive. Par contre, il pensait qu'un jeu pouvait avantageusement être substitué à cette question. Un jeu dans lequel un juge humain, confronté à la fois à un être humain et à un ordinateur essayant de se faire passer pour un humain, essaye de déterminer quel est l'humain. Le test, suggéra Turing, devra être mené via Télétype, limitant ainsi la zone d'investigation à des aptitudes conversationnelles et cognitives. Turing prédit qu'aux alentours de l'an 2000, les ordinateurs « gagneraient » dans 70 pour-cent des cas.

Turing basa son « test » sur un jeu de parloir dans lequel un homme et une femme, cachés derrière un rideau et tous deux se déclarant être des femmes, communiquent en échangeant des notes avec un autre joueur qui essaye de mettre à jour qui est véritablement la femme. Le jeu ne teste pas l'aptitude de chacun à être d'un genre différent, il teste la maîtrise du savoir qui permet de jouer le rôle du genre opposé. Cette distinction repose sur l'utilisation de média; si la communication n'était pas médiatisée – si le juge pouvait voir et entendre les concurrents directement – jouer le rôle fallacieux serait amplement plus difficile, impliquant aussi bien une transformation physique, qu'un savoir et un jeu d'acteur. En rendant la communication médiatisée, seulement limitée à des notes écrites, la tâche ordinairement simple de différencier un homme d'une femme devient difficile.

L'article de Turing a été interprété de différentes façons, allant du manifeste proclamant la facilité d'obtenir un mécanisme presque intelligent à un état d'un extrême scepticisme soulignant l'inaptitude à connaître tous les aspects des autres esprits. Si Turing cru qu'une machine qui pouvait se faire passer pour un humain était capable de penser – ou pouvait être capable de penser – n'est pas clair. Il consacra une part considérable de l'article à réfuter les arguments stipulant que les machines ne pouvaient pas penser, d'une manière suggérant qu'à son avis, elles pouvaient bien en être capables. Il décrit le Jeu de l'Imitation comme une formulation « plus fiable » de la question « les machines peuvent elles penser? », impliquant qu'il crut en l'existence de certaines connections essentielles entre agir comme quelqu'un qui pense et penser véritablement. Cependant, le parallèle implicite qu'il trace entre le jeu du parloir basé sur le genre (dans lequel la différence entre imiter une femme et être une femme est claire) et le test de l'ordinateur/humain suggère que sa première préoccupation était fonctionnelle; il était intéressé par la question de l'aptitude d'un ordinateur à pouvoir agir comme un humain plutôt que par la question plus profonde mais sans réponse de savoir si un ordinateur était par essence comme un humain. Il conclut finalement en disant que le problème repose aussi sur la sémantique :

La question d'origine, « les machines peuvent elles penser? » est je crois trop dépourvue de sens pour mériter une discussion. Néanmoins je crois qu'à la fin du siècle l'usage des mots et l'opinion générale instruite auront tellement évolué que l'on pourra parler de machines pensantes sans s'attendre à être contredit.

L'anthropomorphisme n'est pas nouveau et au cours de l'histoire, on a accordé à de nombreux phénomènes des caractéristiques humaines. Aujourd'hui, on se réfère communément aux machines comme si elles étaient conscientes : Reeves et Nass ont montré qu'il suffisait d'un minimum de comportements humains pour déclencher une réaction sociale [Reeves et Nass 1996]. Pourtant notre relation avec la machine anthropomorphiste est complexe : si l'on demandait directement si un ordinateur peut

penser, beaucoup répondraient « non », bien qu'en réalité ils dialoguent avec la machine comme si c'était un être pensant, lui attribuant de la volonté et réagissant à ses « opinions » plus qu'ils ne le feraient à celles d'autres personnes.

Quand Joseph Weizenbaum créa ELIZA son but n'était, certainement pas, de créer un programme qui duperait les gens en leur faisant croire qu'il était humain. Il espérait plutôt montrer qu'un programme qui pourrait faire l'analyse grammaticale d'un langage naturel et qui avait quelques heuristiques simples pour la formulation des réponses - un programme sans prétention d'intelligence - pourrait jouer au Jeu de l'Imitation relativement bien. Son intention était de démontrer que ce jeu n'était pas un test d'intelligence « plus fiable » dès lors que des machines manifestement non intelligentes pouvaient être amenées à répondre d'une façon humaine vraisemblable. A son grand effarement, beaucoup de gens accueillirent ELIZA avec un grand enthousiasme, l'adoptant comme un partenaire de conversation intelligent; certains même suggéraient qu'ELIZA – en tant que programmes pourrait remplacer les psychothérapeutes humains. Ces réactions découragèrent grandement Weizenbaum, qui se retira effectivement des projets tenants de la AI et devint un militant pour l'humanisme face à l'avancée de la technologie.

L'enthousiasme des gens pour ELIZA est, à première vue, surprenant. Elle (cette chose?) répond en reformulant vos mots sous forme de question ou d'une interrogation générale portant sur vos pensées et vos sensations; le résultat est froid et manque de naturel. Mais pourquoi les gens sont-ils devenus aussi impliqués dans leurs conversations avec elle? En partie car Weizenbaum présenta ELIZA comme un psychothérapeute Rogerien, dont la méthode est de renvoyer à ses patients leurs questions pour les inciter à communiquer plus. Ce scénario donna aux gens un contexte dans lequel le comportement d'ELIZA semblait raisonnable et rationnel – presque humain.

Du point de vue du fonctionnement des programmes de conversation, ELIZA est assez primitif et peu de gens dialoguant avec ELIZA sont véritablement amenés à penser qu'il est humain. Cependant des systèmes plus sophistiqués sont connus pour avoir soutenu une conversation sans se faire détecter pendant un temps considérable. Un concours annuel, la compétition pour le Prix Loebner, offre 100.000 dollars au premier programme qui pourra passer une version assez rigoureuse du test de Turing [Loebner 1999]. Bien que la récompense reste non réclamée, beaucoup de programmes ont trompé certains juges pendant un temps, soutenant des discussions au sujet d'animaux, de sexe ou au deuxième degré, etc. Une fois de plus, les concurrents sont programmés et ne sont pas intelligents, mais semblent intelligents; les « ruses » utilisées par les programmes vainqueurs incluent l'incorporation dans la réponse d'une partie des phrases des utilisateurs, l'orientation de la conversation sur un sujet sur lequel le programme est à même de polémiquer, et une modélisation soignée du rythme et des erreurs de frappe typiques des êtres humains. [Mauldin 1994]. Le rôle que ces programmes de conversation, ou bots comme on commence à les appeler, ont développé en dehors du petit monde de la compétition académique est toutefois plus intéressant: ils sont devenus des participants – leur origine mécanique étant souvent méconnue – à des conversations en ligne.

Le plus célèbre des bots est peut-être Julia, un « chatterbot » qui fréquenta plusieurs MUDs¹, conversant avec les autres participants (humains) [Voir Foner 1997, Mauldin 1994]. Bien que ses réponses soient parfois étranges, les joueurs conversaient parfois longuement avec elle avant de réaliser qu'elle n'était pas un être humain. Foner décrit l'expérience d'un tel joueur, une femme prénommée Lara. Au début, Lara était gênée par la tendance de Julia à parler de hockey (son sujet de conversation par défaut), elle la jugea comme étant un humain ennuyeux; puis, intriguée par certaines choses auxquelles Julia était incapable de répondre (elle ne connaissait pas la Stanley Cup et ne pouvait pas dire où elle avait été à l'école), Lara essaya de diagnostiquer si Julia avait une sorte de handicap mental; et finalement, notant

que Julia répétait certaines phrases textuellement, elle réalisa que c'était un robot.

Les tentatives de Lara pour identifier Julia correspondaient à des actes de catégorisation sociale. Nous donnons un sens au monde en classifiant les choses dans des catégories significatives [Voir Sproull & Kiesler 1991, Lakoff 1990]. Au moment de la rencontre avec un nouvel objet (ou une personne ou une situation), nous le caractérisons par rapport à des catégories familières, qui nous permettent de tracer des inférences à son sujet et de lui assigner des propriétés dépassant notre expérience immédiate. Sans la capacité de catégoriser, le monde serait un amas confus de signaux dépourvus de sens. De façon similaire, nous donnons un sens au monde social en classifiant les gens.

« Le monde quotidien... est peuplé non pas par des quidam, hommes sans visages ni qualités, mais par des êtres, catégories concrètes de personnes déterminées, caractérisées de manière manifeste et correctement étiquetées. » []

Lors d'une première rencontre avec quelqu'un, nous ne percevons que quelques détails à son sujet: peut-être son apparence, les quelques mots qu'il articule, le contexte dans lequel nous le rencontrons. Pourtant, l'impression que l'on se fait de lui est plus profonde. Comme George Simmel l'écrivit dans son influent article paru en 1908 « Comment la Société est elle Possible ? », nous ne voyons pas simplement les quelques détails que nous avons véritablement observés, mais « au moment même où nous comprenons une part aveugle de notre champ de vision dont nous n'avons pas conscience, nous transformons alors une structure fragmentaire... en une individualité complète. ». []. Cet état est atteint en imputant à l'individu dont nous n'avons qu'une vision fragmentaire, les qualités de la catégorie dans laquelle nous l'avons placé 2. Ce processus de catégorisation dans lequel, construire une société est possible, nous permet de nous assurer rapidement une connaissance nouvelle de la personne avec qui nous sommes en relation.

Nous pouvons voir ce processus de catégorisation en œuvre dans la progression des hypothèses de Lara au sujet de l'identité de Julia, hypothèses allant de l'être humain ennuyeux à l'être humain handicapé mentalement puis au programme informatique. Ne disposant que de mots dactylographiés échangés au cours d'une série de conversations assez courtes, l'interlocuteur de Julia (Lara) la classifia tout d'abord dans une rubrique sociale puis dans une autre. En faisant cela, Lara était à même de penser à Julia non pas simplement en termes de fragments issus de leur échange réel mais comme à une catégorie sociale imaginaire complète. Cela fournit à Lara un contexte dans lequel elle peut interpréter les mots de Julia et un modèle de comportement à adopter envers elle []. Par exemple, l'identification initiale de Julia faite par Lara en tant que personne ordinaire, pour autant socialement inepte, lui dicta son comportement:

Au départ, quand je l'ai rencontrée, j'étais foncièrement patiente avec elle. Elle devait avoir un problème d'adaptation sociale avec lequel j'essayais d'être compatissante. Quoiqu'il en soit, après les premières rencontres j'essayais de l'éviter, car tout ce qu'elle faisait, c'était parler hockey.

Afin de deviner que Julia était un robot, Lara avait besoin d'avoir déjà une catégorie pour de tels êtres, bien qu'elle n'en ait jamais rencontrés avant. Cela signifie que Lara connaissait effectivement l'existence de robots en ligne avant qu'elle ne rencontre Julia: elle avait écrit un dialogue pour un ami qui était impliqué dans un programme similaire. Si Lara n'avait rien su des robots logiciels, il est assez improbable qu'elle ait identifié la nature mécanique de Julia. Au contraire, elle aurait modifié la catégorie existante la plus ressemblante dont elle disposait pour englober cette expérience particulière; elle aurait pu finalement décider, par exemple, que Julia souffrait d'un désordre neurologique qui occasionnait des problèmes de mémoire. Sa perception de Julia – et son sentiment de la façon d'agir envers elle – seraient

grandement affectés par l'hypothèse à laquelle elle parviendrait. Et ces impressions, qui constituent sa connaissance des autres, seraient loin de la réalité.

Comme Simmel le notait il y a des inconvénients à l'efficacité cognitive à laquelle nous parvenons au travers de la catégorisation: l'individu ne s'adapte pas nettement aux catégories et ainsi, l'image que nous créons d'une autre personne en étoffant nos impressions fragmentaires est inévitablement une distorsion de sa nature même. Ces distorsions sont spécialement problématiques quand on rencontre quelque chose ou quelqu'un qui est significativement nouveau par rapport aux catégories que l'on a tiré de notre expérience et est, par conséquent, un être assez différent de ceux déjà rencontrés. Il sera tout de même perçu comme un être appartenant à un des types familiers. « [Un] peu de rigidité dans la façon d'interpréter le monde et une certaine lenteur dans la reconnaissance et l'apprentissage de nouveaux modèles » sont le prix de l'efficacité cognitive [].

Il est important de noter que même après que Lara ait réalisé que Julia était une machine, elle continua de parler avec elle, quoiqu'avec des attentes différentes. Bien que Lara su que Julia n'était pas une personne mais simplement un ensemble d'instructions destinées à maintenir un dialogue, elle continua à dialoguer avec Julia comme si elle était, sinon une personne, alors un être ressemblant à une personne; sa relation avec le robot ne pris pas la forme que l'on a avec les objets inanimés, inarticulés.

En effet, tout notre savoir au sujet de l'identité des autres est médiatisé. Nous ne pouvons pas pénétrer directement l'intimité des esprits des autres. Au contraire, nous utilisons leurs apparences extérieures et leurs actes comme des indications pour les classer dans des catégories sociales. Au moyen de l'empathie, ces catégories nous fournissent l'essentiel de notre présumée pénétration des pensées des autres. Cette notion de médiation est encore accentuée dans le monde en ligne : ici, les indicateurs eux-mêmes sont perçus au travers du filtre du médium. Dans la section suivante nous abordons plus précisément la question de ce qui se passe quand notre perception des indicateurs de l'identité est elle même médiatisée.

Communication médiatisée

Jusqu'ici notre discussion s'est limitée uniquement aux supports média utilisant le texte. Le passage d'un environnement n'utilisant que le texte à un autre incluant des images (à la fois statiques et animées) dresse un nouveau panel de questions épistémologiques. Nous devons tout d'abord nous attacher aux conséquences de l'ajout d'images pré-enregistrées, non interactives à l'interface puis aux conséquences de l'ajout d'une vidéo en direct et d'autres média interactifs.

Dans le contexte de l'identité, le visage est l'image clef. Le visage révèle une quantité d'indicateurs d'identité, incluant certaines de nos catégories fondamentales de classification sociale, telles que le genre, l'âge, et la race. Ces indicateurs, à tort ou à raison, imprègnent fortement notre perception des autres; conjointement avec le vocable écrit, ils influencent grandement la façon dont les mots sont interprétés. En sus de ces indicateurs sociaux plus qu'évidents, il y a une quantité de signaux plus subtiles qui sont lus sur les visages : les gens croient qu'ils peuvent détecter des traits de caractère tels que l'honnêteté, l'intelligence, la gentillesse, etc. En fait, alors qu'il y a une concordance considérable parmi les gens au sujet des caractéristiques révélées par un visage particulier, il apparaît n'y avoir que peu de corrélations entre l'interprétation et le caractère réel de la personne []. Notre sentiment d'être capable de lire le caractère sur les visages est fort; notre réelle aptitude à le faire est faible. Ajouter des images des participants à des conversations médiatisées augmentera la connaissance perçue des caractères sociaux, mais la croissance de la connaissance réelle de ces caractères est semble-il moindre.

Quand des images (statiques ou animées) pré-enregistrées non-interactives sont ajoutées, il est clairement possible que l'image soit fallacieuse. Il m'est facile de dactylographier « je suis un homme » alors que je suis en réalité une femme; il m'est tout aussi facile de fournir une photo correspondante et convaincante. Bien que l'ajout de telles images semble élargir le réseau de communication et bien que le destinataire est censé le percevoir comme une source riche et fiable de subtils indicateurs sociaux, l'information contenue dans une simple image non-interactive est réduite au fait que l'expéditeur a choisi cette image comme étant sa représentation 3. Dans le contexte du Jeu de l'Imitation, fournir une image ayant la prétention de représenter le participant serait simple même pour un programme informatique non sophistiqué et pourrait influencer le juge dans sa perception du sujet mécanique comme un humain.

Même de véritables images pourraient réduire le niveau de connaissances. D'un point de vue utopique, le cyber-espace (dans son incarnation textuelle uniquement) est décrit de manière racoleuse comme un monde idéal dans lequel les hommes se rencontrent et se jugent les uns les autres purement sur la base de leurs discours – sur leurs qualités mentales et morales, plus que sur des attributs physiques incidents. Howard Rheingold, un des premiers écrivains sur la culture des communautés virtuelles, écrivit: « Parce que nous ne pouvons pas nous voir les uns les autres dans le cyber-espace, le genre, l'âge, la nationalité d'origine, l'apparence physique ne sont pas apparents à moins qu'une personne veuille rendre de telles caractéristiques publiques »[1]. Le fait est que les indicateurs de catégorisation visuelle altèrent notre vision des autres, tandis que les simples lettres d'un environnement textuel permettent à nos idées d'être communiquées sans préjugés. L'argument sous-jacent est que la connaissance des autres que nous recherchons est la connaissance d'un état intérieur, qui est mieux compris au travers des mots sortant directement de la pensée de l'autre, plutôt qu'au travers de caractéristiques physiques, qui incident et même influencent hautement la construction de notre opinion de l'autre.

Il y a des types d'impostures qui sont facilitées par l'extension du médium. Par exemple, les gens pensent qu'ils peuvent dire en l'observant que quelqu'un ment. Cependant des études poussées ont montré que bien que de fausses déclarations soient marquées par des expressions et des actions caractéristiques (quoique pouvant être infimes et furtives), l'aptitude réelle des gens à reconnaître des expressions dénotant la tromperie de manière éloquente est bien moins forte qu'ils ne la perçoivent [2]. Si les traits en question n'ont pas de composante visible ou si la composante visuelle est un indicateur imparfait, la tromperie serait plus simple dans un environnement plus visuel, car l'étalage visuel offre l'apparente promesse (quoique potentiellement fausse) d'une authenticité immédiatement perceptible. Les participants seraient ainsi moins protégés dans ce médium familier et vraisemblablement transparent.

Cependant une vidéo en direct (contrairement à une image pré-enregistrée, non-interactive) pourrait rendre significativement plus difficile le fait de décrire de façon convaincante certains types d'auto-représentations fallacieuses. Par exemple, considérons un homme se proclamant être une femme. Dans un environnement textuel, l'indicateur de base est simplement l'affirmation « Je suis une femme... » ou peut être l'utilisation d'un nom féminin – un signal trivialement simple pour tricher; dans un environnement visuel, l'indicateur équivalent est une apparence féminine – une imposture plus difficile à créer. Des interactions subséquentes dans l'environnement textuel requièrent une compréhension plus subtile des différences de discours entre l'homme et la femme afin d'être convaincantes[Voir]. Bien qu'en ligne, le grand nombre des hommes pauvrement travestis se déguisant en femmes montre que ce savoir n'est ni évident, ni banal, réaliser une interprétation convaincante dans un environnement textuel n'est pas au delà des compétences d'un observateur astucieux des dynamiques sociales. Dans un environnement basé sur la vidéo en directe, des interactions subséquentes requièrent une compréhension beaucoup plus étendue du discours sexué, de l'expression, de la gestuelle etc. Bien que cela ne soit pas impossible, comme

le prouve l'existence des très convaincants drag-queens, cela requière une habileté considérable, une observation et un degré d'aptitude naturelle. La plus part des femmes putatives et convaincantes du monde textuel se seraient révélées être des hommes dans un environnement vidéo.

Voici Kyoko Date, une « actrice virtuelle » créée par Hori Pro, une société japonaise de divertissement. Jusqu'à maintenant elle est apparue dans de courtes vidéos.« Toutefois, ce projet n'est pas encore achevé. Dans quelques années, la technologie permettra à Kyoko d'apparaître dans des spectacles télévisés en direct et de bavarder avec les autres artistes »

Dans le cas du Jeu de l'Imitation - une machine essayant de se faire passer pour un humain – l'imitation attendue est rendue bien plus difficile par l'extension du médium à la vidéo en direct, dès lors une représentation physique entièrement visible et active doit être créée. Une approche serait de construire un générateur d'images très sophistiqué qui pourrait renvoyer grâce à un programme l'image correspondant à une personne en train de parler, s'exprimant par gestes et tout autres mouvements un peu comme (quoique ce soit via un processus bien plus complexe) Julia dont la version actuelle renvoie les caractères dactylographiés appropriés. Aucun programme informatique de graphismes ne peut aujourd'hui simuler un humain qui pourrait sembler réel après une inspection visuelle approfondie, mais les algorithmes de simulation de mouvements, de texture de peau... sont rapidement améliorés. Il est assez concevable qu'une fois ces barrières technologiques surmontées et la création d'un « acteur »de synthèse vraisemblable accomplie, une version contrainte d'un Jeu de l'Imitation visuelle (lieux connus, pas d'accessoires ou de mouvements de caméra inattendus) pourrait être jouée. L'allégement des contraintes ne changerait pas fondamentalement la nature du problème, mais augmenterait immensément la taille et la complexité de la base de données nécessaires pour générer l'image appropriée.

D'autres modifications du médium peuvent augmenter la vraisemblance, transmettant des informations sur la profondeur, la texture ou l'odeur d'une pièce. Cela peut accroître la capacité du spectateur à sentir les différentes nuances et pourrait augmenter les efforts nécessaires à la simulation d'une fausse identité mais ces changements sont fondamentalement quantitatifs: dans l'absolu, nous ne pouvons pas établir avec certitude qu'aucune des rencontres médiatisées n'est fallacieuse.

Le septique a toujours nié la possibilité de connaître l'existence d'autres systèmes intelligents - tous les autres pourraient bien être, en fait, des robots extra-terrestres. Le pragmatiste a cru en la nécessité de croire, d'un point de vue pratique (et éthique), que les autres sont, comme nous, conscients. La réponse pragmatique traditionnelle à la dénégation du septique est l'argument de l'analogie (je pense que les autres gens ont un esprit comme le mien car leurs comportements sont pareils aux miens) et l'inférence complémentaire à la meilleure explication (je crois que les autres gens ont un esprit car c'est l'explication la plus valable de leur comportement). L'avènement de machines manifestement inintelligentes qui peuvent sembler être intelligentes mettrait fin à la validité du dernier argument.

Dans le monde médiatisé, la force de persuasion de l'inférence comme étant la meilleure explication n'est à priori que temporaire. Aujourd'hui, il n'existe pas de programme pouvant avec succès se faire passer pour un humain après un examen minutieux, ceci même dans un environnement textuel. Mais un tel programme pourrait exister dans le futur: bien que cela puisse prendre assez de temps avant que l'on construise un programme qui puisse tromper le plus astucieux et tatillon des juges, nous avons vu que dans le monde quotidien des interactions hasardeuses et de la catégorisation rapide, les hommes ont déjà conversé avec des machines pensant qu'il s'agissait d'humains. Au travers d'un autre média, tel que la vidéo, la distinction entre l'homme et le robot sera claire pour un long moment encore, mais pas indéfiniment. Pour le moment, on peut être sûr qu'une observation attentive et un scepticisme judicieux nous

préservera de la confusion entre la machine et l'homme, mais les avancées technologiques pourraient bien entamer la conception acceptée des pragmatiques de l'apparemment humain comme étant humain.

Télé-éthique : crédibilité et télé-robotique

Pourquoi est ce important de pouvoir reconnaître une machine en ligne? Est ce un problème de confondre une personne avec un programme?

Dans le monde en ligne, la plus part de nos connaissances viennent du témoignage des autres. Le fait de croire ce que nous entendons dépend de la crédibilité que l'on attribue à notre interlocuteur, i.e. pensons nous que l'interlocuteur est à la fois honnête (la vérité est elle dite telle qu'il ou elle la connaît) et compétent (connaît il effectivement la vérité). De tels jugements sont principalement sociaux; nos croyances de l'honnêteté et de la compétence des autres dérivent en partie de nos prototypes sociaux. Ces prototypes sont particulièrement influents en ligne où l'on est supposé être à même de jauger les mots d'un total étranger; les conversations entre de larges groupes de personnes qui n'ont pas été présentées sont rares dans le monde physique mais très communes dans un espace virtuel. Le médium affecte certainement notre aptitude à juger la crédibilité, mais comme nous l'avons vu, son rôle est complexe: quoiqu'une plus grande connaissance de l'identité des autres semble accroître dans un premier temps notre habileté à juger la crédibilité, on peut aussi arguer que nombre de nos classements en catégories issues de l'apparence physique nous induisent en erreur et qu'un médium qui sépare par filtrage ces indicateurs peut en effet accroître nos connaissances.

Connaître l'identité d'une personne est essentiel pour savoir comment agir envers elle. Un mode d'écriture « enflammé », - colérique, provocant – est endémique dans le monde en ligne actuel []. La connaissance minimale des participants de l'identité de chacun en est une des raisons. Nous fonctionnons normalement à partir d'un ensemble de règles de comportement et de politesse: voilà comment traiter les personnes plus âgées, voilà comment traiter les personnes qui semblent mal à l'aise avec leur entourage... Dans un monde dans lequel nous ne pouvons pas (suffisamment) catégoriser les autres, ces règles ne peuvent être appliquées. Aujourd'hui, le mode d'écriture « enflammé » est limité à des mots incendiaires qui sont eux même assez pernecieux. Pourtant un comportement médiatisé à besoin de ne pas être limité par le flux des mots: la télé-robotique permet d'activer à distance des actions physiques dans l'environnement d'une autre personne []. La combinaison d'une connaissance minimale des autres à l'aptitude à infliger un préjudice réel est dérangeante, particulièrement si l'utilisateur du système de télé-robotique ne croit pas que l'environnement dans lequel il opère est réel.

Enfin, il est important de garder en tête que le but de la communication en grande partie n'est pas l'échange d'informations factuelles, mais l'établissement et le maintien de liens sociaux et de structures. Nous communiquons au travers d'un support, pour acquérir un statut, se faire des amis. Dans ce cas, l'identité de nos compagnons – et certainement leur humanité – est de première importance. Weizenbaum, le créateur d'ELIZA, fut horrifié quand son programme fut accueilli, non pas comme un reproche fait à l'équation de Turing, disant qu'on agit avec intelligence en étant intelligent, mais comme un compagnon divertissant ou un précurseur de la psychothérapie automatisée. Pour Weizenbaum, cet enthousiasme était « déplacé de manière obscène », une menace directe pour notre humanité. Comme les pragmatistes s'opposent au rejet des sceptiques de la possibilité de connaître l'existence d' autres êtres intelligents, au coeur de la plaie humanitaire se trouve la question de l'empathie. Pour reprendre les mots de Lara, après ses confrontations avec Julia :

Je pense que je voudrais savoir si la personne avec qui je parle est REELLE ou pas. Si je sais cela, ce ne serait qu'« une chose », je pense que je n'essayerais pas de devenir sa véritable amie. Je serais cordiale et je continuerais à dialoguer, mais je sais que je ne pourrais pas m'attacher à elle d'un point de vue émotionnel et que ce serait gaspiller de l'énergie de ma part que d'essayer de la rendre sensible. « Les bots ne sont pas sensibles... au livre de mon âme en tout cas... je veux savoir que la personne avec qui je converse est réellement consciente de mes sentiments et de ce que je deviens au travers... pas au travers des instructions des programmeurs mais au travers de l'empathie.[] »

Dans la mesure où le monde virtuel croît jusqu'à englober tous les aspects de nos vies et de nos interactions en ligne, façonner nos communautés, influencer nos opinions politiques et servir d'intermédiaire avec nos proches, la qualité d'être réel, qui est acceptée et présumée de fait dans le monde physique, devient une des questions centrales de société.

Références

Audi 1998

Audi, Robert. 1998 *Epistemology: A Contemporary Introduction to the Theory of Knowledge*. London: Routledge.

Cherry 1978

Cherry, Colin. 1978. *On Human Communication*. 3rd Edition. Cambridge, MA: The MIT Press.

Donath 1998

Donath, Judith S. 1998. Identity and deception in the virtual community. In (Smith, Marc and Kollock, Peter, Eds.) *Communities in Cyberspace*. London: Routledge.

ELIZA nd

ELIZA.

Ekman 1992

Ekman, Paul. 1992. *Telling Lies*. W.W. Norton & Co.

Foner 1997

Foner, Lenny. 1997. Entertaining Agents: A Sociological Case Study. In *Proceedings of the First International Conference on Autonomous Agents*. Marina del Rey, CA.

Froomkin 1996

Froomkin, Michael. 1996. The Essential Role of Trusted Third Parties in Electronic Commerce. 75 *Ore. L. Rev.* 49.

Geertz 1973

Geertz, Clifford. 1973. *The Interpretation of Cultures*. HarperCollins.

Goffman 1959

Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. New York: Doubleday.

Gregory 1990

Gregory, R. L. 1990. *Eye and Brain: The Psychology of Seeing*. Oxford: Oxford University Press.

Harnad 1991

Harnad, S. (1991) Other bodies, Other minds: A machine incarnation of an old philosophical problem .
Minds and Machines 1: 43-54.

Hauser L 1993

L. Hauser, "Reaping the Whirlwind: Reply to Harnad's 'Other Bodies, Other Minds,'" *Minds and Machines* 3 (1993) 220

Herring 1994

Herring, Susan. 1994. Gender differences in computer mediated interaction. Presented at American Library Association annual convention, Miami, June 27, 1994.

Holland and Skinner 1987

Holland, Dorothy and Debra Skinner. 1987. Prestige and intimacy: the cultural models behind Americans' talk about gender types. In (Dorothy Holland and Naomi Quinn, eds.) *Cultural models in language and thought*. Cambridge: Cambridge University Press.

HoriPro nd

HoriPro. 1996. Notes on the development of virtual idol DK-96.

Hutchins nd

Hutchins, J. n.d. Talk to HeX.

Lakoff 1990

Lakoff, George. 1990. *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press.

Loebner 1999

1999 Loebner Prize Competition.

Mauldin 1994

Mauldin, M. 1994. Chatterbots, Tinymuds, and the Turing Test: Entering the Loebner Prize Competition. Proceedings of AAAI-94. August 1-4, 1994. Seattle, WA.

Paulos and Canny 1998

Paulos, Eric and Canny, John. 1998. P_{RO}P: personal roving presence. In Proceedings of SIGCHI 1998, pp. 296 - 303.

Quinn and Holland 1987

Quinn, Naomi and Holland, Dorothy. 1987. Culture and cognition. In (Dorothy Holland and Naomi Quinn, eds.) Cultural models in language and thought. Cambridge: Cambridge University Press.

Reeves and Nass 1996

Reeves, Byron and Nass, Clifford. 1996. The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge, UK: Cambridge University Press.

Rheingold 1993

Rheingold, Howard. 1993. The Virtual Community: Homesteading on the Electronic Frontier. Reading, MA: Addison-Wesley Pub. Co.

Saville-Troike 1989

Saville-Troike, Muriel. 1989. The Ethnography of Communication. Blackwell.

Simmel 1971[1908]

Simmel, George. 1908 (1971). How is society possible. In Simmel, George. 1971. On Individuality and Social Forms. (D. Levine, ed). Chicago: The University of Chicago Press.

Sproull & Kiesler 1991

Sproull, L. and Kiesler, S. 1991 Connections. Cambridge, MA: MIT Press.

Tannen 1996

Tannen, D. 1996. Gender and Discourse. Oxford, UK: Oxford University Press.

Turing 1950

Turing, A.M. 1950. Computing machinery and intelligence. Mind, 59, 433-560.

Weizenbaum 1976

Weizenbaum, J. 1976. Computer power and human reason. San Francisco, CA: W.H. Freeman.

Zebrowitz 1997

Zebrowitz, Leslie A. 1997. Reading Faces. Boulder, CO: Westview Press.

1. MUDs (Multi-Used Dungeons) sont des environnements interconnectés, entièrement textuels dans lesquels de multiples joueurs, tous localisés à distance, peuvent communiquer simultanément.
2. Une étude plus récente de la science cognitive avance la catégorie prototype plutôt que la catégorie la plus abstraite comme étant l'unité cognitive accessible; les arguments de base de l'exposé d'origine de Simmel sont toujours valables [Voir]
3. On pourrait, bien sûr, fournir une image à signature digitale avec les assurances certifiées d'un tiers que l'image fournie est d'une ressemblance véritable; en fait, on pourrait avoir des déclarations textuelles certifiées de façon similaire. Bien que cette approche aborde les questions de la vérité et du savoir, elle est en dehors du cadre de ce chapitre, étant du domaine de l'authentification plutôt que de l'épistémologie de l'interaction sociale. Un article aborde en profondeur l'authentification de l'identité: [].
4. Il est intéressant de noter que les chercheurs ont récemment développé un système informatique qui reconnaît bien mieux que ne le font les gens l'expression subtile et les signes gestuels de la tromperie. Aujourd'hui, les indicateurs de la vraisemblance de l'interlocuteur sont transmis au travers d'un médium visuel, mais le destinataire (l'humain en état d'observation) n'est pas capable de tous les percevoir. En incorporant un tel programme dans l'interface on augmenterait le volume des connaissances disponibles au travers d'un médium, non pas en changeant le médium lui même, mais en augmentant la capacité d'observation du destinataire.