Informatique affective : L'utilisation des systèmes de reconnaissance des émotions est-elle en cohérence avec la justice sociale ?

Mémoire

Alexandra Prégent

Maîtrise en philosophie - avec mémoire
Maître ès arts (M.A.)

Québec, Canada

Informatique affective :
L'utilisation des systèmes de reconnaissance des émotions est-elle en cohérence avec la justice sociale ?

Mémoire

Alexandra Prégent

Sous la direction de :

Jocelyn Maclure, directeur de recherche

Résumé

Identifier correctement, indistinctement de la culture, de l'ethnicité, du contexte, du genre et de la classe sociale, les émotions d'autrui à partir d'une analyse de leurs expressions faciales, c'est ce qu'offrent, en principe, les systèmes de reconnaissance des émotions (SRÉ). En prétendant à un universalisme dans l'expression ainsi que dans la reconnaissance des émotions, nous tenterons de démontrer que les SRÉ présentent des risques non-négligeables de causer des torts importants à certains individus, en plus de viser, dans certains contextes, des groupes sociaux spécifiques. S'appuyant sur un vaste champ de connaissances multidisciplinaires – qui inclut la philosophie, la psychologie, l'informatique et l'anthropologie – ce projet de recherche vise à identifier les limites actuelles des SRÉ ainsi que les principaux risques que leur utilisation engendre, dans l'objectif de produire une analyse claire et rigoureuse de l'utilisation des SRÉ ainsi que de leur participation à une plus grande justice sociale. Mettant de l'avant les limites techniques, nous réfutons, d'une part, l'idée selon laquelle les SRÉ sont en mesure de prouver le lien de causalité entre des émotions spécifiques et des expressions faciales spécifiques. Nous appuyons notre argument par des preuves prouvant l'incapacité des SRÉ à distinguer les expressions faciales d'émotions des expressions faciales en tant que signaux de communication. D'autre part, en raison des limites contextuelles et culturelles des SRÉ actuels, nous réfutons l'idée selon laquelle les SRÉ sont en mesure de reconnaître, à performance égale, les émotions des individus, indistinctement de leur culture, ethnicité, genre et classe sociale. Notre analyse éthique démontre que les risques sont considérablement plus nombreux et plus importants que les bénéfices que l'on pourrait tirer d'une utilisation des SRÉ. Toutefois, nous avons séparé un type précis de SRÉ, dont l'utilisation se limite au domaine du *care*, et qui démontre un potentiel remarquable pour participer activement à la justice sociale, non seulement en se conformant aux exigences minimales, mais en répondant aussi au critère de bienfaisance. Si, actuellement, les SRÉ posent des risques importants, il est toutefois possible de considérer la possibilité que certains types spécifiques participent à la justice sociale et apportent une aide ainsi qu'un support émotionnel et psychologique à certains membres de la « société »

Abstract

Emotion recognition systems (ERS) offer the ability to identify the emotions of others, based on an analysis of their facial expressions and regardless of culture, ethnicity, context, gender or social class. By claiming universalism in the expression as well as in the recognition of emotions, we believe that ERS present significant risks of causing great harm to some individuals, in addition to targeting, in some contexts, specific social groups. Drawing on a wide range of multidisciplinary knowledge - including philosophy, psychology, computer science and anthropology - this research project aims to identify the current limitations of ERS and the main risks that their use brings, with the goal of providing a clear and robust analysis of the use of ERS and their contribution to greater social justice. Pointing to technical limitations, we refute, on the one hand, the idea that ERS are able to prove the causal link between specific emotions and specific facial expressions. We support our argument with evidence of the inability of ERS to distinguish facial expressions of emotions from facial expressions as communication signals. On the other hand, due to the contextual and cultural limitations of current ERS, we refute the idea that ERS are able to recognise, with equal performance, the emotions of individuals, regardless of their culture, ethnicity, gender and social class. Our ethical analysis shows that the risks are considerably more numerous and important than the benefits that could be derived from using ERS. However, we have separated out a specific type of ERS, whose use is limited to the field of care, and which shows a remarkable potential to actively participate in social justice, not only by complying with the minimum requirements, but also by meeting the criterion of beneficence. While ERS currently pose significant risks, it is possible to consider the potential for specific types to participate in social justice and provide emotional and psychological support and assistance to certain members of society.

Table des matières

Liste des illustrations

Liste des abréviations, sigles, acronymes

GAFA	Google, Amazon, Facebook, Apple
IA	Intelligence Artificielle
LFW	Labeled Faces in the Wild
RPC	République populaire de Chine
SIA	Système d'intelligence artificielle
SRÉ	Système de reconnaissance d'émotions
SRF	Système de reconnaissance Faciale
TSA	Trouble du spectre de l'autisme
UÉ	Universalité des émotions

Introduction

Un outil technologique qui permettrait d'avoir accès à l'intériorité d'autrui; à ses sentiments, ses appréhensions, ses joies et ses passions : voilà à la fois la promesse et l'ambition des systèmes de reconnaissance des émotions. Depuis plusieurs années, la popularité des systèmes de reconnaissance des émotions (SRÉ) connaît une augmentation drastique dans de nombreux pays[1] (Chine[2], États-Unis[3], les États membres de l'Union Européenne[4], etc.) et dans de nombreuses compagnies privées[5]. Par ailleurs, la demande pour ces types de systèmes semble venir de presque toutes les sphères de la société ; sécurité, marketing, éducation, finance, *care*, politique, ressources humaines ; il semblerait que ces systèmes pourraient être très utiles en augmentant la sécurité ou en améliorant l'approche-client par exemple[6]. Leur utilisation semble toutefois créer une tension avec la volonté de la société de protéger et respecter les droits et libertés fondamentaux (puisque, comme nous le verrons, ces technologies ne semblent pas en mesure de protéger équitablement les droits et libertés de tous les individus) ainsi que sa capacité à octroyer un traitement juste (c'est-à-dire un traitement constant et impartial), soit deux exigences fondamentales sur lesquelles repose la justice sociale.

Ainsi, il appert que l'état actuel de la réflexion concernant la légitimité des SRÉ est très peu développé, alors que la demande à l'égard de tels systèmes est mondialement en constante augmentation. Réfléchir aux risques et aux enjeux éthiques et sociaux en amont d'un déploiement massif de ces systèmes confère la possibilité de prévenir les torts probables plutôt que de réagir à ces derniers, une fois commis.

Ainsi, considérant la situation actuelle, il semble que l'étude des systèmes de reconnaissance des émotions soit plus que nécessaire. Les premières apparitions de ce type de systèmes ont déjà été accompagnées de critiques

---

[1] Article 19. « Emotional Entanglement: China's emotion recognition market and its implications for human rights », *Article 19*, 2021, https://www.article19.org/wp-content/uploads/2021/01/ER-Tech-China-Report.pdf.

[2] Mou, X. « Artificial Intelligence: Investment Trends and Selected Industry Uses », *International Finance Corporation*, 2019, https://www.ifc.org/wps/wcm/connect/7898d957-69b5-4727-9226-277e8ae28711/EM Compass-Note-71-AI-Investment Trends.pdf?MOD=AJPERES&CVID=mR5Jvd6.

[3] NITRD. « Supplement to the President's FY2020 Budget », *The Networking and Information Technology Research and Development Program*, 2019, https://www.nitrd.gov/pubs/FY2020-NITRD-Supplement.pdf#page=17. Et Cornillie, C. « Finding Artificial Intelligence Money in the Fiscal 2020 Budget », *Bloomberg government*, 2019, https://about.bgov.com/news/finding-artificial-intelligence-money-fiscal-2020-budget/ (Page consultée le 19 mars 2021)

[4] Commission européenne. « Façonner l'avenir numérique de l'Europe: la Commission présente des stratégies en matière de données et d'intelligence artificielle », *Commission Européenne*, 2020, https://ec.europa.eu/commission/presscorner/detail/fr/ip_20_273 (Page consultée le 15 mars 2021)

[5] Lewis, T. « AI can read your emotions. Should it? », *The Gardian*, 2019, https://www.theguardian.com/technology/2019/aug/17/emotion-ai-artificial-intelligence-mood-realeyes-amazon-facebook-emotient (Page consultée le 17 mars 2021)

[6] Crawford, K. *The Atlas of AI : Power, Politics and the planetary cost of Artificial Intelligence*, Yale University Press, New Haven and London, 2021, p. 152.

sévères quant à ce que leur utilisation représente pour la société et ses citoyens[7]. En continuation avec ces critiques, ce mémoire propose d'explorer les principaux lieux de tension entre l'utilisation des systèmes de reconnaissance des émotions et la justice sociale afin (1) d'identifier les risques concrets et (2) de proposer des pistes de solution. Pour ce faire, nous proposons de relever les différents points de tension que sont : la complexité du phénomène émotionnel, les désaccords historiques quant au rôle des émotions dans la rationalité, le manque d'appui scientifique quant à une corrélation significative entre les expressions faciales et les émotions, ainsi que les limites contextuelles et culturelles des SRÉ. Ces différents points de tension seront par la suite utilisés afin d'élaborer une analyse éthique des risques et des bénéfices de l'utilisation de ces systèmes au regard de leur compatibilité à la justice sociale. Plus largement, notre objectif est de fournir une analyse qui se trouve en amont des déploiements de masse afin d'apporter une réflexion pertinente sur les conditions normatives nécessaires à la légitimation de l'utilisation des SRÉ, pour la société canadienne et ailleurs dans le monde.

L'aboutissement de ce mémoire a dû relever trois grandes difficultés que nous tenons à souligner en amont de la lecture du projet de recherche puisqu'elles ont toutes les trois modifié notre approche vis-à-vis celui-ci. La première difficulté identifiée concerne l'état des connaissances actuelles sur les émotions. En effet, les « émotions » – aussi nommés « affects » ou « états émotionnels » – n'ont pas encore reçu de définition qui soient collectivement acceptée au sein des scientifiques[8]. Les définitions permettent d'encadrer le concept en posant les limites sur ce qui est inclus dans le concept et, à l'inverse, sur ce qui est exclu. Une absence de définition crée un flou quant à la position de ces limites, ce qui, en un effet domino, nous oblige à mener notre recherche sans une définition du phénomène observé par les systèmes que nous analysons.

Cette première difficulté nous mène directement à la deuxième, à savoir que les SRÉ sont des technologies qui proposent de reconnaître un phénomène que les scientifiques ne sont pas capables d'identifier, de définir ou même de cerner de manière consensuelle. Ainsi, avec les systèmes de reconnaissance des émotions, le système doit identifier ou reconnaître une émotion alors que le terme « émotion », de même que le concept

[7] Lighthill, M. J. *Artificial Intelligence : A General Survey*. In Science Research Council, *Artificial Intelligence: A Paper Symposium*, London, SRC, 1-21, 1973 ; Dreyfus, H. *What Computers Can't Do : A Critique of Artificial Reason*, Harper & Row, New York, 1972.

[8] Fehr et Russell. « Concept of Emotion Viewed from a Prototype Perspective », Journal of experimental psychology, 113(3), 464-486. 1984, https://psycnet.apa.org/doi/10.1037/0096-3445.113.3.464. "Many have sought but no one has found a commonly acceptable definition for the concept of emotion." p. 464. ; Russell, J.A. et Barrett, L.F. « Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant », Journal of Personality and Social Psychology, 76(5), 805-819, 1999, https://psycnet.apa.org/doi/10.1037/0022-3514.76.5.805. p. 805 ; Crawford, K. *The Atlas of AI : Power, Politics and the planetary cost of Artificial Intelligence*, Yale University Press, New Haven and London, 2021, p. 17.

auquel il appartient, ne renvoie à aucune définition précise. Cette difficulté a pour effet que les SRÉ doivent reconnaître un phénomène au sujet duquel nous avons des connaissances limitées.

Enfin, si l'émotion, en tant que concept et en tant que phénomène, est difficile à circonscrire, il est faux de dire que les SRÉ identifient les émotions. Les SRÉ ne s'intéressent pas concrètement à l'émotion puisqu'ils n'y ont pas directement accès ; ils analysent plutôt les symptômes de l'émotion. Ils sont programmés pour identifier des effets physiologiques, qui sont supposés être ceux des émotions, et analysent ceux-ci dans l'objectif d'émettre des hypothèses sur l'émotion vécue par la personne. Par ailleurs, puisque les émotions, en tant qu'expériences subjectives, ne sont pas des phénomènes auxquels autrui est en mesure d'avoir un accès direct, les SRÉ ne cherchent pas tous les mêmes « symptômes » de l'émotion. Cette variation dans les symptômes recherchés sur la base des traits ou expressions du visage pose une difficulté à notre recherche puisque la comparaison des SRÉ entre eux tout comme leur regroupement dans des catégories spécifiques est une tâche ardue qui devient rapidement problématique. Ainsi, notre analyse éthique des SRÉ relève des enjeux éthiques communs qui sont présents dans la majorité des SRÉ utilisés mais non dans tous les SRÉ présents sur le marché. Par ailleurs, cette pluralité dans les méthodes utilisées par les SRÉ nous a contraint à nous concentrer plus particulièrement sur les SRÉ les plus utilisés, soit ceux qui utilisent l'apprentissage machine et l'apprentissage profond dans l'objectif d'établir une corrélation significative entre certaines expressions faciales et des émotions spécifiques, ressenties et vécues par une personne.

Le premier chapitre porte son attention sur la justice sociale, le rôle des émotions dans les processus de décision ainsi que l'utilisation malveillante des SRÉ. Cette succession, en trois temps, des grands thèmes de notre projet de recherche nous permet de faire ressortir dès le début les liens pertinents entre la justice sociale, le rôle des émotions et l'utilisation des SRÉ. Ainsi, dans un premier temps, nous aborderons la justice sociale et verrons qu'il existe diverses manières de la mettre en pratique. Nous ferons ressortir deux exigences minimales de la justice sociale qui se trouvent être des constantes à l'intérieur des diverses pratiques. Ces exigences seront nos critères de base tout au long de notre projet de recherche, qui nous permettront d'évaluer la participation des SRÉ à une plus grande justice sociale. Dans un deuxième temps, dans l'objectif de bien saisir à la fois l'importance de l'émotion dans notre vie quotidienne et les conséquences d'une analyse extérieure de celle-ci par un SRÉ, nous nous tournerons vers la psychologie et les sciences cognitives et observerons les liens établis entre les émotions et le processus de décision chez l'humain. L'explicitation des liens entre les émotions et le processus décisionnel se fera à travers notamment les recherches du neuroscientifique Antonio R. Damasio et du chercheur en science cognitive Daniel Kahneman. Dans un troisième temps, nous observerons l'intégration de la modélisation informatique des émotions et son impact dans la manière d'étudier le comportement des individus à travers le marketing et la politique. Finalement, dans un quatrième temps, nous analyserons l'un des

risques possibles de l'utilisation de SRÉ par le marketing et la politique, qui est l'utilisation malveillante à des fins de manipulation des décisions de l'individu.

Le deuxième chapitre porte son attention sur l'inférence causale sur laquelle reposent les SRÉ ainsi que sur les limitations de cette inférence. Les SRÉ qui identifient et reconnaissent les émotions des individus par une analyse des expressions faciales établissent une inférence causale entre les émotions vécues et des séries spécifiques de contractions musculaires. Ainsi, dans un premier temps, nous verrons les recherches de Charles Darwin et de Paul Ekman qui menèrent à populariser l'hypothèse de l'existence de cette inférence pour ensuite démontrer les différentes couches d'inférences causales qui sous-tendent l'hypothèse qu'il est possible d'avoir un accès universel aux états émotionnels intérieurs des individus en analysant leurs expressions faciales. Dans un deuxième temps, nous questionnons cette prétention à l'universalité, c'est-à-dire à reconnaître « universellement » les états émotionnels intérieurs en analysant les expressions faciales.

Le troisième chapitre abordera en profondeur deux des principaux risques qui accompagnent une utilisation des SRÉ. L'identification de ces deux importantes limites a deux objectifs : infirmer la thèse de l'universalisme des SRÉ et identifier les facteurs pertinents sur lesquels se concentrer pour améliorer les prochains SRÉ. Ainsi, d'une part, nous avancerons que les SRÉ ne sont pas en mesure, à l'heure actuelle, de tenir compte du contexte, qui influence pourtant les expressions faciales des individus selon plusieurs facteurs tel que la classe sociale et le genre. Nous observerons les impacts de ces deux facteurs sur les expressions faciales émotionnelles des individus. Nous verrons que ces facteurs motivent les individus à accentuer, atténuer ou neutraliser l'expression de certaines émotions dans des contextes donnés. D'autre part, nous verrons que les normes sociales varient selon les cultures, ce qui signifie que des individus issus de deux cultures différentes n'afficheront pas nécessairement les mêmes expressions faciales (d'une émotion donnée) lorsque mis dans une même situation. De plus, les SRÉ n'ont pas démontré une efficacité équivalente pour tous les individus issus de toutes les cultures. Ainsi, il ne peut être prétendu que l'utilisation de SRÉ sera aussi efficace dans toutes les cultures. Par ailleurs, dans nos sociétés pluriculturelles, il n'est pas acceptable qu'une technologie n'ait pas les mêmes taux d'efficacité pour tous et chacun en raison de la culture d'appartenance.

Enfin, le quatrième chapitre abordera directement la question de la participation des SRÉ à une plus grande justice sociale à travers une analyse éthique de leur utilisation. En ce sens, nous identifions trois critères avec lesquels nous analysons l'utilisation de plusieurs types de SRÉ selon divers objectifs (sécurité, marketing, finance, etc.). Par la suite, nous relevons les bénéfices ainsi que les risques que posent leur utilisation ainsi que les groupes désavantagés par cette utilisation. Nos critères d'évaluation sont la non-discrimination, l'efficacité et la fiabilité. La satisfaction de ces critères est primordiale afin de respecter les deux exigences de la justice sociale. Enfin, après l'analyse de l'utilisation des SRÉ en général, nous porterons notre attention sur une

utilisation plus spécifique des SRÉ, qui se situe dans la catégorie du *care*. Ces types de SRÉ, qui présentent des risques et des désavantages similaires aux autres SRÉ, se distinguent par leur objectif, qui en est un de bienfaisance. Cet objectif apporte des avantages considérables qui rebalancent l'analyse éthique pour ces types de SRÉ.

L'intelligence artificielle : quelques bases historiques

L'intelligence artificielle (IA) connaît ses débuts au milieu des années 1950, en parallèle avec les débuts de l'informatique. Principalement liées aux mathématiques, les premières applications de l'IA prouvent des théorèmes mathématiques de manière autonome[9]. Peu de temps après, l'apprentissage machine se tournera vers la traduction de langues et aura notamment un rôle prédominant dans la Guerre Froide où elle sera en mesure de traduire automatiquement des phrases du russe à l'anglais[10]. À la fin des années 1960, l'IA connaîtra une période de crise où les recherches stagnent lorsque les promesses et prédictions au sujet de l'IA se révèlent être des fantasmes scientifiques – qui incluaient entre autres choses les voitures volantes et les robots munis d'une intelligence comparable à celle des êtres humains. Cet ajustement déflationniste quant aux avancements probables de l'IA fait chuter les investissements et plonge les recherches dans un premier « hiver de l'IA ».

Les années 1970 feront place à l'émergence des grandes questions éthiques à l'égard de l'IA et déjà certaines des premières grandes critiques se penchent sur la question des émotions; notamment sur les différences incontestables entre le « vécu » humain et la simple capacité de calcul. Certains chercheurs, dont Hubert Dreyfus, mettront l'accent sur l'importance du ressenti et des émotions dans l'apprentissage humain ainsi que leur rôle dans l'intelligence[11]. Les années 1980 marqueront une modeste renaissance de l'IA avec les systèmes experts, qui ont la capacité d'effectuer les mêmes analyses que les humains dans des domaines précis, dont certaines analyses médicales. Cette renaissance relancera les investissements qui permettront un développement important des algorithmes d'apprentissage même si, toutefois, cette relance sera momentanément freinée par la commercialisation des ordinateurs au grand public qui détournera l'attention de l'IA à l'informatique classique. La renaissance de l'IA est en partie dû à la commercialisation des ordinateurs, maintenant conçus et développés pour le grand public. Cette alliance permettra, une décennie plus tard, l'augmentation drastique de la puissance de calcul, de la capacité de stockage et de l'accumulation de

[9] The History of Computing. « Logic Theorist – Complete History of the Logic Theorist Program », The History of Computing, 2021, https://history-computer.com/logic-theorist-complete-history-of-the-logic-theorist-program/ (Page consultée le 11 juin 2021)
[10] Chomsky, N. *Syntactic Structures*, Mouton, The Hague, 1957.
[11] Dreyfus, H. *Alchemy and Artificial Intelligence*, Rand Corporation, New York, 1965 ; Dreyfus, H. *What Computers Can't Do : A Critique of Artificial Reason*, Harper & Row, New York, 1972 ; Lighthill, M. J. *Artificial Intelligence : A General Survey*. In Science Research Council, *Artificial Intelligence: A Paper Symposium*, London, SRC, 1-21, 1973.

données qui permettront à leur tour l'apparition des données massives (Big Data) et de certaines IA très performantes, comme celles intégrées aux véhicules autonomes, les suggestions dans les moteurs de recherche, les détecteurs de fraude, les systèmes de reconnaissance faciale (SRF) et les systèmes de reconnaissance des émotions (SRÉ).

De leur côté, bien que les SRF furent critiquées à leur début par le monde scientifique (en raison des graves échecs de reconnaissance sur certains groupes ethniques spécifiques (accusant ainsi de lourds biais discriminatoires[12]), ils devinrent tout de même très vite la technologie biométrique la plus financée au monde. Le marché mondial de la reconnaissance faciale générait, en 2019, 3,4 milliards de dollars de revenus avec un taux d'augmentation annuel prévu de 14,5% sur 8 ans[13]. La reconnaissance faciale est une technologie informatique qui fonctionne de manière probabiliste en ce sens que les déterminations sont basées sur une prédiction. Les SRF utilisent des algorithmes d'IA afin de reconnaître automatiquement une personne sur la base de son visage. La reconnaissance faciale a ainsi deux fonctions, celle de l'authentification, qui sert à vérifier l'identité de la personne, et celle de l'identification, qui vise à reconnaître ou identifier des personnes simplement en analysant leur visage. Ce procédé se fait généralement en trois grandes étapes : (1) détecter le visage par une image ou une vidéo, (2) créer un code numérique unique au visage et (3) comparer celui-ci avec des codes numériques de visages identifiés.

La reconnaissance faciale est convoitée dans plusieurs domaines et c'est pour cette raison que plusieurs entreprises multimilliardaires ont participé à sa création et à son évolution, dont les GAFA. Les GAFA ont financé et publié de nombreuses recherches en reconnaissance faciale et ont participé activement à ses récents succès. En mettant à profit leur monopole sur les réseaux sociaux et les applications de téléphone et d'ordinateur, en exposant leurs recherches et en en finançant d'autres, les GAFA ont permis à la reconnaissance faciale de connaître des progrès significatifs, alors qu'ils avoisinent maintenant le 100% d'efficacité, du moins dans des contextes d'application spécifiques.

[12] Crawford, K. et al. « AI Now 2019 Report », *AI Now Institute*, New York, 2019, https://ainowinstitute.org/AI_Now_2019_Report.html.
[13] Grand View Research. « Facial Recognition Market Size, Share & Trends Analysis Report By Technology, 2021 – 2028 », *Grand View Research*, 2021, https://www.grandviewresearch.com/industry-analysis/facial-recognition-market.

En 2014, l'algorithme GaussianFace, développé par des chercheurs de l'université de Hong Kong, reconnaissait les visages à 98,52 %[14] sur la base de données « *Labeled Faces in the Wild* (LFW) [15]». De son côté, l'algorithme de Facebook, DeepFace, un nouveau logiciel en mesure de déterminer si deux visages appartiennent ou non à la même personne a eu un taux de succès de 97,35%[16] sur la même base de données. En 2015, Google entrait dans la compétition avec FaceNet, qui réussissait presque à reconnaître un visage et à identifier la personne à 100% (99,63%)[17] sur LFW. Un peu plus tard, en 2018, c'est Amazon qui fait la promotion du logiciel Rekognition[18] qui est en mesure d'identifier 100 personnes sur une image en plus d'intégrer la reconnaissance de certaines émotions.

Dans des domaines plus spécifiques tel que la santé, ce qui est généralement appelé l' « analyse faciale » permettrait entre autres choses de détecter certaines maladies génétiques[19]. En sécurité, elle attire à la fois les gouvernements et les individus puisque ses possibilités vont de la lutte contre la criminalité à la protection de documents privés. À l'heure actuelle, la Chine et l'Inde dominent le marché mondial de la reconnaissance faciale. Pour sa part, la Chine s'est servie de cette technologie pour installer un réseau de vidéo-surveillance sur (presque) tout le territoire, qui comprenait 200 millions de caméras en 2018[20]. En Inde, c'est le projet Aadhaar qui se démarque par sa base de données biométriques, comprenant 1, 27 milliards de résidents[21], la plus grande base au monde. Suivant l'Inde, le Brésil a déclaré à son tour vouloir initier une « collecte » de données biométriques pour ses 200

---

[14] Hu, J., Lu, J. et Tan, Y. « Discriminative Deep Metric Learning for Face Verification in the Wild », *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR),*1875-1882, 2014, http://openaccess.thecvf.com/content_cvpr_2014/papers/Hu_Discriminative_Deep_Metric_2014_CVPR_paper.pdf

[15] *Labeled Faces in the Wild* est une référence publique pour la vérification des visages, également connue sous le nom de "pair matching". Voir : http://vis-www.cs.umass.edu/lfw/.

[16] Taigman, Y., Yang, M., Ranzato, M. et Wolf, L. « DeepFace: Closing the Gap to Human-Level Performance in Face Verification », *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR),*1701-1708, 2014, https://www.cv-foundation.org/openaccess/content_cvpr_2014/papers/Taigman_DeepFace_Closing_the_2014_CVPR_paper.pdf.

[17] Schroff, F., Kalenichenko, D. et Philbin, J. « FaceNet: A Unified Embedding for Face Recognition and Clustering », *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 815-823, 2015, https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Schroff_FaceNet_A_Unified_2015_CVPR_paper.html.

[18] Amazon. « Amazon Rekognition Improves Face Analysis », *Amazon Web Services*, 2019, https://aws.amazon.com/fr/about-aws/whats-new/2019/08/amazon-rekognition-improves-face-analysis/. (Page consultee le 4 mai 2021)

[19] Briganti, G. « L'IA pour la détection des maladies génétiques grâce à la reconnaissance faciale », *NumeriCare*, 2019, https://www.numerikare.be/fr/actualites/e-health/l-rsquo-ia-pour-la-detection-des-maladies-genetiques-grace-a-la-reconnaissance-faciale.html (Page consultée le 3 février 2021)

[20] Article 19. « Emotional Entanglement: China's emotion recognition market and its implications for human rights », *Article 19*, 2021, https://www.article19.org/wp-content/uploads/2021/01/ER-Tech-China-Report.pdf

[21] Perrigo, B. « India Has Been Collecting Eye Scans and Fingerprint Records From Every Citizen. Here's What to Know », *Time*, 2019, https://time.com/5409604/india-aadhaar-supreme-court/.

millions de citoyens[22]. Enfin, parmi les principaux pays impliqués, la Russie récolte depuis 2017 les données biométriques de ses citoyens incluant non-seulement la reconnaissance faciale, mais la voix, l'iris, les tatouages et les empreintes digitales[23].

Quant à eux, les systèmes de reconnaissance des émotions ont emboité le pas des systèmes de reconnaissance faciale et représentent désormais une industrie de 20 milliards de dollars[24]. Aujourd'hui, les SRÉ sont présents dans les systèmes de sécurité nationale, dans les aéroports, dans les systèmes d'éducation, sur les réseaux sociaux, dans les processus d'embauche, dans les systèmes de surveillance de la police, dans les hôpitaux, etc[25]. Toutefois, les systèmes de reconnaissance des émotions ne se basent pas sur des preuves et méthodes scientifiques équivalentes aux systèmes de reconnaissance faciale[26]. La légitimité des SRÉ repose sur l'hypothèse (non-prouvée) de l'existence d'une inférence causale entre les expressions faciales et les émotions vécues, soit l'idée qu'il serait possible d'identifier les émotions d'une personne en analysant ses expressions faciales. Par ailleurs, malgré plusieurs décennies d'études anthropologiques quant à la non-universalité des émotions ainsi que des expressions faciales[27] et les études, plus récentes, quant à l'absence de preuves scientifiques vis-à-vis du lien entre les expressions faciales et les émotions, la demande en systèmes de reconnaissance des émotions ne cesse de croître[28].

De la même manière que les systèmes de reconnaissance faciale ont été entraînés à reconnaître, identifier et authentifier des visages à partir des bases de données des plus importants réseaux sociaux tel Facebook, Instagram et Snapchat, les SRÉ ont été entraînés à analyser les expressions faciales

---

[22] Burts, C. « Brazil plans massive centralized biometric database of all citizens to improve agency data sharing », *Biometricupdate.com*, 2019, https://www.biometricupdate.com/201910/brazil-plans-massive-centralized-biometric-database-of-all-citizens-to-improve-agency-data-sharing (Page consultée le 6 mars 2021)

[23] Light, F. « Russia Is Building One of the World's Largest Facial Recognition Networks, *The Moscow Times*, 2019, https://www.themoscowtimes.com/2019/11/12/russia-building-one-of-worlds-largest-facial-recognition-networks-a68139 (Page consultée le 19 mars 2021)

[24] Schwartz, O. « Don't Look Now: Why You Should Be Worried about Machines Reading Your Emotions » The Guardian, 2019, https://www.theguardian.com/technology/2019/mar/06/facial-recognition-software-emotional-science (Page consultée le 24 février 2021)

[25] Crawford, K. *The Atlas of AI : Power, Politics and the planetary cost of Artificial Intelligence*, Yale University Press, New Haven and London, 2021, p. 152-153.

[26] Heaven, D. « Why Faces Don't Always Tell the Truth about Feelings » *Nature*, 2020, https://www.nature.com/articles/d41586-020-00507-5.

[27] Crawford, K. *The Atlas of AI : Power, Politics and the planetary cost of Artificial Intelligence*, Yale University Press, New Haven and London, 2021, p. 152-153.

[28] Crawford, K. et al. « AI Now 2019 Report », *AI Now Institutes*, New York, 2019, https://ainowinstitute.org/AI_Now_2019_Report.html, p. 6.

sur des réseaux similaires comprenant entre autres, Instagram, Pinterest, Tik Tok et Flickr[29]. Ayant accès à des milliards d'images, l'objectif était d'utiliser le *machine learning* – dont le *deep learning* – afin de découvrir les « indices » permettant de lier des expressions faciales spécifiques (ex : froncer les sourcils) avec des émotions spécifiques (ex : colère).

Après avoir créé leur propre système de reconnaissance faciale, les BigTech tout comme les startups développent et déploient des systèmes de reconnaissance des émotions qui sont utilisés dans des domaines variés. Le système d'Amazon, Rekognition, était l'un des premiers a affirmer être en mesure de reconnaître sept émotions comprenant la joie, la peur, la colère, la tristesse, le dégoût, la confusion et la surprise[30]. La compagnie HireVue, qui compte parmi ses clients Intel, Goldman Sachs et Unilever vend un SRÉ qui utilise l'apprentissage automatique pour évaluer les traits faciaux des futurs employés afin de déterminer leur aptitude à occuper l'emploi pour lequel ils postulent[31]. En 2016, la compagnie Apple venait d'acquérir la startup Emotient, qui affirmait développer des systèmes capables de reconnaître les émotions des individus sur une image fixe. Aux côtés d'Apple, Amazon, Microsoft et IBM ont eux aussi développés leur propre système de reconnaissance des émotions. Microsoft a développé le système API Face, qui affirme quant à lui être en mesure d'identifier les émotions d'une personne lorsqu'elles correspondent à la peur, la colère, la joie, la tristesse et la surprise. Par ailleurs, Microsoft affirme que les émotions détectées par son système sont des émotions universellement exprimées à travers les cultures[32]. Tous ces systèmes reposent donc sur la supposition qu'il existe un certain nombre d'émotions qui seraient universelles et que les individus exprimeraient inconsciemment sur leur visage : ces émotions pourraient donc être identifiées et reconnues par les SRÉ. En des termes généraux – puisque chaque SRÉ dispose de ses propres méthodes d'analyse – les SRÉ collectent, à partir d'une analyse des visages, des données qui sont par la suite corrélées afin de déterminer l'expression faciale manifestée. Par la suite, le schéma des corrélations est associé à une émotion.

---

[29] Crawford, K. *The Atlas of AI : Power, Politics and the planetary cost of Artificial Intelligence*, Yale University Press, New Haven and London, 2021, p. 152-153.

[30] Amazon. « Amazon Rekognition Improves Face Analysis », *Amazon Web Services*, 2019, https://aws.amazon.com/fr/about-aws/whats-new/2019/08/amazon-rekognition-improves-face-analysis/. (Page consultee le 4 mai 2021)

[31] Harwell, D. « A Face-Scanning Algorithm Increasingly Decides Whether You Deserve the Job », *Washington Post*, 2019, https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning -algorithm-increasingly-decides-whether-you-deserve-job/. (Page consultée le 4 mai 2021)

[32] Microsoft Azure. « Face: An AI Service That Analyzes Faces in Images », *Microsoft Azure*, https://azure.microsoft.com/en-us/services/cognitive-services/face/. Dans Crawford, K. *The Atlas of AI : Power, Politics and the planetary cost of Artificial Intelligence*, Yale University Press, New Haven and London, 2021, p. 155.

L’analyse des expressions faciales émotionnelles a pour objectif d’établir des corrélations significatives entre des schémas précis d’expressions faciales et des émotions, toutefois, ces corrélations s’appuient sur l’inférence causale que les émotions ressenties (l’expérience subjective) s’affichent sur nos visages sous la forme d’expressions faciales.

En raison donc de l’engouement grandissant pour les systèmes de reconnaissance des émotions, des lacunes importantes au niveau de leurs fondements scientifiques ainsi que de leur grandes promesses – soit avoir accès à l’intériorité d’autrui en identifiant les émotions – il est possible d’affirmer que l’étude des systèmes de reconnaissance des émotions est plus que pertinente, elle est nécessaire. Ainsi, nous poursuivrons dans ce mémoire cette étude selon un axe précis : celui de déterminer si l’utilisation des SRÉ pourrait participer à une plus grande justice sociale.

Chapitre 1 La justice sociale et l'utilisation des SRÉ

Dans ce premier chapitre, nous nous intéresserons à la justice sociale et à l'utilisation diversifiée des SRÉ afin de prendre en compte à la fois la disparité des champs dans lesquels les SRÉ sont utilisés et la pertinence d'étudier cette utilisation, qui se retrouve dans presque toutes les sphères de la société. Dans un premier temps, nous nous attarderons au concept de justice sociale dans l'objectif de déterminer les exigences que les SRÉ doivent satisfaire afin d'être évaluées comme étant en cohérence avec la justice sociale. Dans un deuxième temps, dans l'objectif de bien saisir à la fois l'importance de l'émotion dans notre vie quotidienne et les conséquences d'une analyse extérieure de celle-ci par un SRÉ, nous nous tournerons vers la psychologie et les sciences cognitives et observerons les liens établis entre les émotions et le processus de décision chez l'humain. L'explicitation des liens entre les émotions et le processus décisionnel se fera à travers notamment les recherches du psychologue Antonio R. Damasio. Dans un troisième temps, nous observerons l'intégration de la modélisation informatique des émotions et son impact dans la manière d'étudier le comportement des individus à travers le marketing et la politique. Finalement, dans un quatrième temps, nous analyserons l'un des risques possibles de l'utilisation de SRÉ par le marketing et la politique qui est l'utilisation malveillante à des fins de manipulation des décisions de l'individu.

1.1. Sur la variabilité probable dans l'application du concept de justice sociale

Quoique ce ne soit pas toujours le cas, la justice est souvent considérée comme le principe universel qui permet de juger si une action est éthiquement acceptable ou souhaitable. Toutefois, l'application de ce principe « universel » peut prendre plusieurs formes qui varient grandement selon les époques et les cultures. Ainsi, il est inévitable de constater qu'il existe plusieurs conceptions de la justice[33].

La justice sociale peut par exemple reposer sur des accords collectivement acceptés. La Déclaration universelle des droits de l'homme[34] et la Convention Européenne des droits de l'homme[35] en sont des

[33] Miller, D. « Justice », *Stanford Encyclopedia of Philosophy*, 2017, https://plato.stanford.edu/entries/justice/ (Page consultée le 4 avril 2021)
[34] Nations Unies. « Déclaration universelle des droits de l'homme », *Nations Unies*, https://www.un.org/fr/universal-declaration-human-rights/index.html (Page consultée le 4 avril 2021)
[35] Cour européenne des droits de l'homme. « Convention européenne des droits de l'homme », Cour européenne des droits de l'homme, https://www.echr.coe.int/Pages/home.aspx?p=basictexts&c=fre (Page consultée le 4 avril 2021)

exemples. Ces accords reconnaissent certains droits et libertés fondamentaux comme les piliers sur lesquels se fonde la justice. Ainsi, pour ces accords, il est nécessaire que tous les êtres humains soient reconnus comme des êtres humains qui « naissent et demeurent libres et égaux en droits[36] ». Cette reconnaissance est primordiale pour poser les bases nécessaires en matière de justice dans le monde. La Déclaration universelle des droits de l'homme[37] démontre clairement cette idée que tout être humain, de par sa position d'humain, a la responsabilité de respecter certains droits fondamentaux chez autrui – droits qui précèdent et transcendent le droit institué :

> « La reconnaissance de la dignité inhérente à tous les membres de la famille humaine et de leurs droits égaux et inaliénables *constitue le fondement* de la liberté, de la justice et de la paix dans le monde[38] » (l'italique est de nous).

Similairement au début de la Convention européenne des droits de l'homme nous pouvons y lire:

> « Ces libertés fondamentales […] *constituent les assises* mêmes de la justice et de la paix dans le monde et dont le maintien repose essentiellement sur un régime politique véritablement démocratique, d'une part, et, d'autre part, sur une conception commune et un commun respect des droits de l'homme dont ils se réclament[39] » (l'italique est de nous).

De ces deux accords, qui explicitent les exigences de base nécessaires à la justice sociale, il est possible d'en dégager une constance, soit celle voulant que chaque individu possède, du fait de son appartenance à l'humanité, un certain nombre de droits et libertés comprenant notamment le droit à la vie, à la sécurité et à l'égalité pour tous.

Par ailleurs, il nous est possible de dégager une deuxième constance. La présence de la justice est aussi demandée dans la distribution adéquate des biens et des ressources de la société aux individus. Elle a donc pour objectif de s'assurer que chacun reçoit ce qui lui est dû en termes de distribution et de redistribution[40] et, parallèlement, elle s'efforce de résoudre les conflits qui occurrent dans cette phase de répartition des biens entre individus. Toutefois, la justice peut se fonder dans plusieurs sources diverses telle l'égalité ou l'équité ou encore dans la reconnaissance de besoins universels

---

[36] Conseil constitutionnel de France. « Déclaration des droits de l'homme et du citoyen (1789) », Article 1, *Conseil constitutionnel de France*, https://www.conseil-constitutionnel.fr/le-bloc-de-constitutionnalite/declaration-des-droits-de-l-homme-et-du-citoyen-de-1789 (Page consultée le 4 avril 2021)

[37] Mentionnons au passage que le terme « homme », qui a été préféré aux termes « humains », « êtres humains », « personnes », etc., est ici plus que révélateur de l'évolution constante des mœurs collectives.

[38] Nations Unies. « Déclaration universelle des droits de l'homme », *op. cit.*

[39] Ibid.

[40] Forsé, M. & Parodi, M. « Justice distributive : La hiérarchie des principes selon les Européens ». *Revue de l'OFCE*, 2006, 3(3), 213-244. https://doi.org/10.3917/reof.098.0213.

chez les êtres humains. Cette caractéristique de variabilité dans la conception du principe de justice explique, d'une part, et seulement en partie, pourquoi il est si difficile d'arriver à des décisions qui sont consensuelles et, d'autre part, permet de voir que la justice, peu importe la manière dont elle est appliquée, se préoccupe de la manière dont les individus sont traités[41]. Déjà selon la notion aristotélicienne de la justice, il est demandé que les cas semblables soient traités de façons semblables (alors que les cas différents exigent parfois un traitement différencié)[42]. Dans des circonstances semblables, donc, les jugements devraient être similaires et ainsi faire preuve de constance et d'impartialité. Ainsi, il est aussi possible d'affirmer que la justice, dans son application, se préoccupe d'octroyer un traitement juste, c'est-à-dire un traitement impartial et constant, peu importe la manière dont il est effectué.

1.1.1. Conclusion partielle

Dans cette section, nous avons dégagé deux constances à travers les diverses possibilités de méthodes d'application possibles du principe de justice qui alimenteront notre réflexion : (1) celle que chaque individu possède, du fait de son appartenance à l'humanité, un certain nombre de droits et libertés comprenant notamment le droit à la vie, la sécurité et l'égalité pour tous et (2) qu'un traitement juste est un traitement qui inclus l'impartialité et la constance dans son application, peu importe le processus d'application. Ces constances nous serviront de guides à partir desquels nous dégagerons des critères d'exigence pour déterminer si l'utilisation des SRÉ participent à la justice sociale.

1.2. Déconstruction de l'idéal rationaliste classique

L'idéal rationaliste effectue une scission drastique entre les passions (émotions) et la raison. Cette idée, d'une pensée isolée et indépendante du corps tout comme des passions, a grandement influencé les recherches en sciences cognitives et en psychologie[43], qui ont tenté de prouver empiriquement les

[41] Miller, D. « Justice », *Stanford Encyclopedia of Philosophy*, 2017, https://plato.stanford.edu/entries/justice/.
[42] Aristote. *Éthique à Nicomaque* (trad. J. Tricot), Vrin, Paris, 2012, Livre 5, 1129a à 1131b.
[43] Schenk, F. « Les émotions de la raison », *Revue européenne des sciences sociales*, XLVII-144, 2009, 151-162, https://journals.openedition.org/ress/75#citedby. p. 160.

idées philosophiques des séparations esprit-corps et raison-passion[44]. Dû à la force de la tradition philosophique, qui voyait les émotions comme étant un « à-côté de la pensée rationnelle, voulu par la nature, mais non par le sujet pensant[45] », les recherches en psychologie avaient (intuitivement) tendance à vouloir préserver cette dualité profondément ancrée dans les esprits et ainsi à concevoir le cerveau comme étant séparé en des parties bien distinctes les unes des autres, voire, indépendantes[46]. Selon cette perspective, la faculté de raisonner était localisée dans ce que l'on appelait le « haut niveau » tandis que la faculté de ressentir des émotions se situait dans le « bas niveau ». Cependant, à la fin du XXe siècle, un questionnement quant à la possibilité de l'existence d'une relation beaucoup plus complexe qu'une simple dualité entre la raison et les émotions s'est levé dans la communauté scientifique. Antonio R. Damasio fut l'un des premiers neuroscientifiques en Occident à apporter des preuves à cet effet.

En 1848, en Nouvelle-Angleterre, un jeune homme âgé de 25 ans fort, intelligent et sensible aux autres, voit son crâne transpercé par une barre de fer. Un accident à son travail. Il s'appelle Phinéas Gage. La chance lui sourit et en plus de ne pas être mort sur le coup, il guérit de sa blessure et est considéré comme un miracle aux yeux des médecins de son époque. Il parle, ses capacités physiques et intellectuelles semblent intactes, il répond à toutes les questions et démontre des réflexions profondes. Or sa famille et ses employés considèrent qu'il n'est plus le même; sa personnalité et son *humeur* ont changé. Mais plus important encore, il semble qu'il n'arrive plus à prendre de décisions. Pourtant, tous les tests démontrent que ses facultés cognitives n'ont pas été touchées. Le cas reste irrésolu et tombe presque dans l'oubli. Près d'un siècle plus tard, Antonio R. Damasio observe et effectue plusieurs tests sur un homme adulte qui présente les mêmes symptômes que Phinéas Gage; ses capacités physiques et intellectuelles sont intactes, mais sa personnalité a changé et il n'est plus en mesure de prendre des décisions. Dans *Descartes's Error*, Damasio le renommera Elliot. Elliot fut opéré au cerveau pour une tumeur bénigne mais d'une grosseur inquiétante. À la suite de l'observation des comportements d'Elliot et de plusieurs tests sur ses capacités intellectuelles et physiques, Damasio en vient à se questionner sur les émotions de son patient. Ce dernier lui confie explicitement qu'il ne ressent plus rien et Damasio en vient à la conclusion que son patient est en mesure de « connaître » mais non de « ressentir »[47].

[44] Voir Descartes, R. *Les Passions de l'âme [1649]*, Vrin, Paris, 1994. ; Descartes, R. *Méditations métaphysiques [1641],* Flammarion, Paris, 2011.
[45] Damasio, A. *L'erreur de Descartes*, Odile Jacob, Paris, 2010, p. 83.
[46] Ibid.
[47] Ibid., p. 74.

Cette révélation du patient est suivie par plusieurs autres cas similaires qui mèneront Damasio à requestionner le rôle de l'émotion dans les processus de prises de décision. Damasio en vient à poser l'hypothèse que les émotions se trouveraient en fait à la base du rouage de la faculté de raisonner.

Ce détachant de la tradition qui persistait à isoler la raison de tout contact avec les états émotionnels, Damasio est l'un des premiers à tenter d'expliquer la relation entre les émotions et la raison comme un processus unifié plutôt qu'en parties séparées et indépendantes l'une de l'autre. Endossant la démarche et le point de vue d'un neuropsychologue, il avance l'hypothèse que la faculté de raisonnement dépendrait de plusieurs systèmes de neurones œuvrant simultanément à de nombreux niveaux de l'organisation cérébrale, et non pas d'un seul centre cérébral. Du cortex préfrontal à l'hypothalamus et au tronc cérébral, de nombreux centres cérébraux, de « haut niveau » aussi bien que de « bas niveau », contribueraient au fonctionnement de la faculté de raisonnement[48]. Les recherches menées par Damasio vinrent confirmer sa première hypothèse selon laquelle les émotions et la raison étaient en fait étroitement liées. En fait, plus précisément, les recherches de Damasio venaient suggérer « l'existence d'interactions entre le système neuronal responsable de la capacité de ressentir des émotions et celui sous-tendant la faculté de raisonnement et de prise de décision[49] ». Par ailleurs, ces recherches le menèrent à observer qu'une seule région spécifique du cerveau, le cortex préfrontal ventro-médian, si lésé, entraînait des répercussions néfastes à la fois sur les processus de raisonnement et de prise de décision ainsi que sur l'expression et la perception des émotions[50]. Cette découverte empirique venait suggérer que les émotions et la « raison » ne viendraient pas même de régions distinctes du cerveau.

Cette découverte empirique vint remettre en question la pureté de la faculté de raisonnement puisque la capacité à exprimer et ressentir des émotions est finalement inséparable des comportements rationnels[51] : « la perception des émotions est à la base de ce que les êtres humains appellent, depuis des millénaires l'âme ou l'esprit[52] ». Ainsi, en plus de ne pas être, au quotidien, des ennemis de la raison – comme des théories tel que celles de Descartes[53] et Kant[54] le laissaient croire – les émotions

[48] Ibid., p. 10.
[49] Ibid., p. 85.
[50] Ibid., p. 106.
[51] Ibid., p. 9.
[52] Ibid., p. 13-14.
[53] Descartes, R. *Méditations métaphysiques, op. cit*.
[54] Kant, E. *Fondements de la métaphysique des mœurs [1785]* (trad. V. Delbos), Éditions Les Échos du Marquis, France, 2013

se trouvent être des constituants nécessaires de celle-ci. Si cette thèse a pris du temps avant d'être partagée dans le monde scientifique, c'est peut-être dû au fait qu'elle est contre-intuitive et ce, à deux niveaux[55]. Premièrement, Damasio n'oppose plus la raison aux émotions comme deux entités distinctes (ainsi séparée en premier lieu par la philosophie) mais plutôt comme des capacités interdépendantes d'un même système. Deuxièmement, il avance que le processus décisionnel chez l'humain est dépendant de la capacité à exprimer et ressentir des émotions au sens où sans la capacité à exprimer et ressentir des émotions, la raison est incapable de prendre des décisions « en accords avec nos projets personnels, les conventions sociales et les principes moraux[56] ». Cela signifie donc que, si une émotion vécue trop fortement peut causer un comportement irrationnel, un *affaiblissement* de la capacité de réagir émotionnellement peut également être la cause de comportements irrationnels. Il est donc nécessaire de déconstruire l'idée amenée par l'idéal rationaliste classique, selon laquelle nous arrivons à des raisonnements rationnels par une raison qui est épurée de toutes émotions.

Les émotions sont, au contraire, un aspect à la fois naturel et essentiel de notre être. Les émotions permettent de peser les choix qui s'offrent à nous ainsi que de limiter ces choix à des nombres restreint selon nos intérêts[57]. Par ailleurs, elles nous permettent de communiquer à nous-même et aux autres ce que nous ressentons. C'est en ce sens que, découvertes après découvertes, les scientifiques ont été amené à reconnaître qu'au quotidien, les émotions ne sont généralement pas un obstacle à la prise de décision rationnelle mais sont à l'inverse l'une des conditions nécessaires[58]. Ainsi, de nos jours, il y a consensus pour dire que les émotions font parties de la cognition.

[55] Selon Pichon et Vuilleumier, « bien qu'il soit encore courant de distinguer des régions cérébrales « émotionnelles » et d'autres plutôt « cognitives », cette vision dichotomique est relativement schématique. En effet, l'un des apports de la NiF (Neuro-imagerie fonctionnelle) a été de montrer l'importance des interactions entre régions néocorticales traditionnellement associées aux fonctions cognitives et régions dites « limbiques » associées aux émotions, dont certaines (comme l'amygdale) entretiennent des relations directes avec l'ensemble du cerveau, y compris les cortex sensoriels et moteurs ». Dans Pichon, S. et Vuilleumier, P. « Imagerie et cognition, Neuro-imagerie et neuroscience des émotions », *Médecine/Sciences*, 27(8-9), 763-769, 2011, https://medweb4.unige.ch/labnic/papers/SP_PV_MedSci2011.pdf. p. 764.
[56] Damasio, A. *op. cit*., p. 9.
[57] Schenk, F. « Les émotions de la raison », *Revue européenne des sciences sociales*, XLVII-144, 2009, 151-162, https://journals.openedition.org/ress/75#citedby, p. 161.
[58] Ibid., p. 160-161.

1.2.1. Conclusion partielle

Dans cette section, nous avons exposé la déconstruction de l'idéal rationaliste. La « rationalité » devient dorénavant un concept beaucoup plus complexe qui demande une redéfinition. Les recherches de Damasio nous ont permis de comprendre le rôle plus précis des émotions dans le processus de prises de décision. Ainsi, pour nous, non seulement la raison est désormais délogée de sa tour d'ivoire mais, elle est aussi en partie dépendante des émotions pour assurer son bon fonctionnement. En somme, affirmer que la raison possède comme conditions *sine qua none* de ressentir et vivre des émotions ne signifie pas que nous renonçons à la rationalité mais plutôt que nous déconstruisons l'idéal rationnel pour mieux le reconstruire.

L'idéal rationnel semble en ce sens être plus près de la théorie d'Aristote, qui soutenait qu'un être rationnel est un être qui ressent les bonnes émotions au bon moment et au degré approprié à la situation[59], que de celle de Kant par exemple, qui soutiendrait qu'un être rationnel devait être en mesure d'éloigner sa raison de ses émotions (ou passions) pour ne suivre que la loi morale[60].

1.3. L'affect comme principale cause dans le changement de paradigme des schémas conceptuels

« *Comprendre les émotions, les sentiments du consommateur, est aussi fondamental que comprendre ses pensées* »
(Edell et Burke, 1987)

La psychologie n'est pas le seul domaine de recherche à avoir effectué un changement de paradigme drastique (la philosophie, étant, dès Aristote, en débat sur la question) quant au rôle de l'émotion dans les processus de prise de décision. La finance, le marketing, la politique, la sécurité, le *care*, l'éducation, etc., sont tous des domaines qui ont récemment eu un changement de paradigme au sein de leur recherche quant à la place de l'émotion dans la prise de décision.

---

[59] Aristote. *Éthique à Nicomaque*, *op. cit*., livre 2, 1104b à 1106b

[60] « Il ne faut pas du tout se mettre en tête de vouloir dériver la réalité du principe [moral] à partir de la constitution particulière de la nature humaine. Ce qui est dérivé de la disposition naturelle propre de l'humanité, ce qui est dérivé de certains sentiments et de certains penchants, [...] tout cela peut bien nous fournir une maxime à notre usage mais non une loi, un principe subjectif selon lequel nous pouvons agir par penchant et inclination, non un principe objectif d'après lequel nous aurions l'ordre d'agir, alors même que tous nos penchants, nos inclinations et les dispositions de notre nature y seraient contraires ». Dans Kant, E. *op. cit.,* p. 100-101.

Par exemple, en finance, ce sont les travaux de Daniel Kahneman et Amos Tversky, qui initièrent les premiers changements quant à la perception des comportements et des décisions « rationnels »[61]. Étudiant les marchés financiers des années 1960 jusqu'aux débuts des années 1970, ils se penchèrent sur la finance comportementale[62]. En étudiant le comportement des investisseurs lors de leurs prises de décision, ils en viennent à relever ce que l'on appelle aujourd'hui des biais cognitifs[63], et prouvent que ces investisseurs, qui croyaient fermement agir de manière purement rationnelle, agissaient fréquemment en fonction d'heuristiques[64] dans les jugements de probabilités[65].

Dans le but de réfuter le modèle de l'*homo œconomicus* rationnel, qui présuppose que le marché financier aboutirait naturellement « aux équilibres les plus efficaces économiquement comme s'ils obéissaient à des règles purement rationnelles[66] », Kahneman et Tversky prouvèrent que les investisseurs possédaient tous différents travers comportementaux cognitifs qui les poussaient, à un moment ou à un autre, à prendre une décision qui ne pouvait s'expliquer comme étant une décision *purement rationnelle,* mais plutôt une décision émotive. En effet, l'analyse des deux chercheurs sur les cycles financiers ont démontré, entre autres choses, que les grandes fluctuations du moral des investisseurs suivaient souvent le même cycle que celui des marchés boursiers[67]. Selon leur recherche, l'investisseur a tendance à croire qu'il existe une raison qui justifie qu'il priorise une décision x qui n'est pas en cohérence avec la décision qu'il aurait dû prendre s'il s'était basé uniquement sur son jugement rationnel (par exemple, le fait qu'il fasse deux bonnes transactions de suite pourrait l'influencer et lui faire croire qu'on n'a « jamais deux sans trois » et ce, inconsciemment). Un

---

[61] Lorsque nous mettons le terme « rationnel » entre guillemets, nous faisons référence à la rationalité tel que défendue par la conception idéale rationaliste classique, qui tente de maintenir une séparation drastique entre la raison et les émotions.

[62] La finance comportementale est une branche de l'économie comportementale qui consiste à appliquer la psychologie cognitive à la finance.

[63] Jusqu'à présent, 250 biais cognitifs sont recensés, généralement classés dans les catégories suivantes : Biais sensori-moteurs (illusions liées aux sens et à la motricité), Biais attentionnels ou biais d'attention (problèmes d'attention), Biais mnésique (en rapport avec la mémoire), Biais de jugement (déformation de la capacité de juger), Biais de raisonnement (paradoxes dans le raisonnement) et Biais liés à la personnalité (en rapport avec la culture, la langue, l'influence sociale…). Voir Usabilis. « Définition biais cognitifs », *Usabilis*, 2018, https://www.usabilis.com/definition-biais-cognitifs/.

[64] Les heuristiques sont des raccourcis cognitifs que le cerveau humain prend dans l'objectif d'économiser de l'« énergie mentale ». Toutefois, les raccourcis peuvent mener à des erreurs de jugements qui causent la prise de mauvaises décisions (ou simplement de ne pas prendre la « meilleure » décision).

[65] Martinez, F. « L'individu face au risque : l'apport de Kahneman et Tversky », *Idées économiques et sociales*, 3(3), 15-23, 2010, https://doi.org/10.3917/idee.161.0015

[66] Finance comportementale. « La finance comportementale, les apports de la psychologie », *Finance comportementale*, Abc bourse, 2020, https://www.abcbourse.com/apprendre/19_finance_comportementale.html

[67] Kahneman, D., & Tversky, A. « Prospect Theory: An Analysis of Decision under Risk », *Econometrica, 47*(2), 263-291, 1979, www.jstor.org/stable/1914185

investisseur qui vient d'enchaîner une série de plusieurs échecs aura tendance à craindre de plus en plus les valeurs risquées et, inversement pour un excès de confiance, après une série de bonnes transactions, l'investisseur aura tendance à sous-estimer les risques[68]. Sans s'en apercevoir, l'investisseur ne prendra donc plus de décisions rationnelles puisqu'il accordera un poids de plus en plus important à la suite que forment ses échecs ou ses succès, alors qu'un algorithme, par exemple, ne changerait jamais la valeur du poids accordé à chaque situation en fonction de la suite que forment ses précédentes décisions (échecs ou succès).

C'est sur la base de cette hypothèse qu'ils créèrent la théorie des perspectives[69], qui est une théorie du choix face au risque[70]. Ainsi, alors que la théorie économique classique postule que les individus évaluent les différents états du monde de manière absolue et objective, Kahneman et Tversky proposent que les individus évaluent les situations de manière *relative*, par rapport à un point de référence qui peut être *subjectif*[71]. Cela signifie que selon eux, la décision prise par l'individu ne porte pas sur les états finaux (comme le propose la théorie classique) mais sur les changements en termes de richesse ou de bien-être par rapport à une position initiale conventionnellement définie. Ainsi, devant un choix risqué conduisant à des gains, l'individu a tendance à ressentir une forte aversion au risque et favorisera les options conduisant à une utilité espérée qui sera plus sûre, mais inférieure. À l'opposé, devant un choix risqué conduisant à des pertes, l'individu aura tendance à tendre vers une forte envie de recherche de risque (l'envie étant considérée comme une émotion), préférant les solutions conduisant à une utilité espérée supérieure mais moins sûre, lorsqu'il y a une chance de diminuer les pertes. Il y a donc une double tendance chez l'individu qui se divise en (1) l'aversion pour le risque, qui l'entraîne à se soustraire aux situations qu'il juge déjà favorable et (2) la recherche de réalisation du potentiel, qui entraîne l'individu à adopter des comportements qui accentuent la prise de risque lorsque la situation de base est jugée non-satisfaisante[72]. Cette variation dans la façon d'évaluer les situations selon notre point de référence personnel a de grandes répercussions. En effet, en acceptant cette idée, il est alors impossible de croire que les investisseurs prendraient toujours les décisions qui les

[68] Kahneman, D., & Tversky, A. *op. cit*., p. 269.
[69] Ibid.
[70] Il est important de noter que Kahneman publia ses recherches dans les années 1970, près de vingt ans avant l'ère du « Algorithmic Trading », qui s'installa au début des années 1990. À cette époque, l'humain avait encore un rôle important dans le processus de décision au sein des marchés financiers, ce qui est moins le cas maintenant.
[71] Gollier, C., Hilton, D. & Raufaste, É. « Daniel Kahneman et l'analyse de la décision face au risque », *Revue d'économie politique*, 113(3), 295-307, 2003, https://www.cairn.info/revue-d-economie-politique-2003-3-page-295.htm
[72] Ibid., p. 296.

mèneraient vers le choix le plus efficace économiquement. Ainsi, l'aversion et l'envie du risque sont deux émotions à la fois complexes et précises qui influencent les choix quotidiens.

De manière similaire, le monde du marketing fut longtemps mené par l'idée que la raison était (et devait) être isolée pour prendre les décisions les plus rationnelles possibles. Dans cet élan, les théories de l'approche cognitive, en marketing, affirmaient entre autres choses que la préférence du consommateur était toujours le résultat d'une pensée consciente et « rationnelle ». Le consommateur était illustré comme une personne objective qui formait le plus souvent sa décision sur le résultat d'un calcul utilitariste de base pour décider quel produit, parmi ceux mis à sa disposition, serait le « meilleur » selon ses propres critères arbitraires[73]. Ainsi, les théories de l'approche cognitive dérivaient de cette démarche des classements préférentiels et des critères de choix, uniquement basé sur l'idée que le consommateur arriverait avec une liste préconstruite de critères « rationnels » qui lui importait personnellement[74]. Par exemple, si un consommateur arrivait dans un magasin de meubles pour acheter un divan, l'approche cognitive estimait que les critères de base seraient probablement du type : (1) Est-il abordable ? (2) Est-il confortable? (3) Quelle est sa durée de vie ? (4) Est-il aisément nettoyable ? Le divan se faisant attribuer le meilleur résultat serait le divan choisi par le consommateur.

Toutefois, les théories de l'approche cognitive se trouvaient incapables d'expliquer le choix d'une personne qui dirait par exemple « Je prendrai ce divan car je l'aime bien ». Comment rattacher cette information à un choix rationnel ? Ce type de commentaires, généralement assez globaux et peu précis, est en effet difficile à rattacher à un antécédent cognitif et conséquemment, difficile à expliquer par les théories de l'approche cognitive. De même, les « achats impulsifs[75] »[76], c'est-à-dire des achats non-calculés qui sont fait par le consommateur de manière spontanée, échappaient aux théories de l'approche cognitive.

---

[73] Martinez, F. « L'individu face au risque : l'apport de Kahneman et Tversky », *Idées économiques et sociales*, 3(3), 15-23, 2010, https://doi.org/10.3917/idee.161.0015 ; Derbaix, C. et Pham, M. T. « Pour un développement des mesures de l'affectif en marketing : synthèse des prérequis », *Recherche et Applications en Marketing*, SAGE, Association française du marketing, 4(4), 71-87, 1989, https://www-jstor-org.acces.bibl.ulaval.ca/stable/pdf/40588767.pdf?ab_segments=0%252Fbasic_search%252Fcontrol&refreqid=excelsior%3A4659bdda8ad2106fcdc2b8a4b08c2b63, p. 72.

[74] Les théories de l'approche cognitive en marketing semblent parfois utiliser le mot « rationnel » comme synonyme de « pragmatique »

[75] Derbaix, C. et Pham, M. T. *op. cit.*, p. 75.

[76] Et avec eux, l'apparition des « nudges », notamment développés en économie comportementale par les chercheurs Thaler et Sunstein

Quoique l'approche cognitive tenta plusieurs alternatives afin de parvenir à justifier ces types de comportements, comme l'incorporation d'un *rattachement affectif* au modèle cognitiviste[77], une autre approche fut développée : l'approche affective. L'approche affective se base sur l'intuition que les processus de prise de décision ne doivent pas toujours être issus d'un choix froid et « rationnel ». Le processus de prise de décision n'est plus regardé comme un calcul éclairé des coûts-bénéfices pour le client sur un produit quelconque et il est désormais vu comme reposant en grande partie sur des composantes émotionnelles et inconscientes.

L'approche affective permet donc de donner à la réponse « j'achète ce divan parce que je l'aime bien », une explication qui met l'emphase sur le côté émotionnel du choix. La motivation de l'achat peut aussi bien-être un coup de cœur que liée à des désirs plus profonds. Par exemple, un client pourrait vouloir un divan spécifique dans l'objectif de suivre les tendances et ainsi de faire partie d'un groupe précis. Conséquemment, le désir de faire partie d'un groupe (et d'être accepté par les membres de ce groupe) est ce qui motive le client. De leur côté, les achats impulsifs sont intéressants au sens où ils mobilisent les émotions d'une autre manière. Acheté un divan spécifique pour suivre les tendances et acheter subitement un divan « coup de cœur » sont deux achats qui sont motivés par des types d'émotions différents. Selon l'approche affective de Derbaix et Pham[78] par exemple, l'affect est présent dans la majorité des comportements de consommation selon une intensité, une durabilité et un degré de conscience distincts[79]. Ainsi, l'achat du divan peut mobiliser des émotions qui ont une durabilité variable dépendamment de la cause qui motive l'achat et ce, même si ces deux achats mobilisent des émotions.

1.3.1. Utilisation malveillante des SRÉ, un risque à la liberté de choix

Par ailleurs, avec la commercialisation de l'Internet et de l'IA, un engouement considérable vis-à-vis de la recherche en modélisation informatique des émotions[80,81] est apparu ces dernières années et a

[77] Derbaix, C. et Pham, M. T. *op. cit*.
[78] Derbaix, C. et Pham, M. T. *op. cit*.
[79] Voir Annexe 1 pour la typologie de l'affectif qu'ils proposent.
[80] Delbrouck, P. « Les émotions humaines peuvent-elles être discrètes ? » *La Lettre du Psychiatre*, 7(1), Nouvelles technologies, 16-19, 2016, https://www.edimark.fr/Front/frontpost/getfiles/23843.pdf, p. 18.
[81] La modélisation informatique des émotions entre dans le domaine de l'informatique affective (affective computing), domaine développé par Rosalind Picard.

révolutionné plusieurs domaines, dont le marketing. La reconnaissance des émotions des individus est devenue l'un des principaux objectifs des collectes massives de données personnelles puisqu'elle permettrait, entre autres choses, de guider le consommateur à acheter un produit en particulier aussi bien que de guider les vendeurs afin que ceux-ci puissent plus facilement identifier les clients potentiels. Les vendeurs sur Facebook peuvent par exemple utiliser la « DataVisualisation[82] », qui a principalement pour fonction d'expliquer, de clarifier et d'éclairer les relations entre les données. Facebook rend cette fonction accessible, entre autres, aux marques et aux détenteurs de pages publiques. Lorsque ces derniers désirent accroître la portée de leur contenu, ils ont la possibilité d'acheter des espaces publicitaires qui assurent un « ciblage précis d'usagers[83] »[84]. Cet outil de publicité « permet à Facebook de proposer des critères précis et invite [les marques et les propriétaires de pages publiques] à spéculer sur leur public potentiel[85] ». Ce faisant, les propriétaires de pages publiques et les marques contribuent à perfectionner les corrélations significatives de l'algorithme en ayant l'opportunité de sélectionner les principaux traits des acheteurs potentiels.

Les données récoltées sur un individu peuvent en dire beaucoup sur sa personne. À titre d'exemple, les recherches de Kosinski, Stillwell et Graepel ont démontré que seulement à partir des « j'aime » sur le réseau social Facebook, leur algorithme était en mesure de créer un profil « psycho-démographique » de l'utilisateur qui leur permettait de savoir, entre autres choses, le genre, l'orientation sexuelle, l'ethnie, l'appartenance religieuse, les tendances politiques et la consommation de drogues ou de cigarette de l'utilisateur[86,87]. Ces « j'aime » sont donc vus et utilisés par les compagnies comme une forme de quantification des états émotionnels de l'utilisateur (qui deviendra rapidement un client).

Par ailleurs, les systèmes de reconnaissance des émotions peuvent prendre des formes variées et ainsi reposer sur plusieurs autres systèmes que celui de la récolte de données sur Internet. Certains systèmes (comme la reconnaissance faciale, vocale et des signaux physiologiques[88] de l'individu),

---

[82] Monino J.-L. et Sedkaoui, S. *Big Data, Open Data et valorisation des données*, 4, *ISTE Édition*, Londres, 2016, p. 129.
[83] Alcantara, C., Charest, F. et Agnostinelli, S. (Dir.). *Big Data et visibilité en ligne : un enjeu pluridisciplinaire de l'économie numérique*. Paris, Presse des Mines, 2018, p. 222.
[84] Kumar, P. « Corporate Privacy Policy Changes during PRISM and the Rise of Surveillance Capitalism », *Media and Communication*, 5(1), 2017, p. 74.
[85] Alcantara, C., Charest, F. et Agnostinelli, S. (Dir.), *op. cit*., p. 225.
[86] Kosinski, M., Stillwell, D. et Graepel, T. « Private traits and attributes are predictable from digital records of human behavior », *National Academy of Sciences*, 5802-5805, 2013, https://www.pnas.org/content/pnas/110/15/5802.full.pdf?3=
[87] Voir Annexe B pour le tableau détaillé.
[88] Delbrouck, P. *op. cit.,* p. 18.

utilisés simultanément, permettent de « reconnaître l'état émotionnel à partir des signaux issus des différents capteurs, en mettant en œuvre des mécanismes permettant d'exploiter conjointement les informations recueillies[89] ». Une fois ces informations traitées, il est possible d'en tirer des savoirs sur les individus et leurs particularités respectives qui pourront guider les branches marketing des compagnies à viser des clientèles précises via une gestion des états affectifs du consommateur physiquement présent à l'intérieur du magasin. En général, le vendeur sait qu'il devra tout d'abord réussir à contrôler l'anxiété de départ du client à la vue de la possibilité qu'il dépense une somme monétaire[90]. Le vendeur peut conséquemment inciter le consommateur à oublier son anxiété en lui précisant qu'un certain modèle est en grande demande mais que les stocks seront bientôt épuisés ou encore que le rabais sur le modèle sélectionné n'est effectif que pour les vingt-quatre prochaines heures, etc.

Évidemment, la modélisation informatique des émotions dépasse le champ d'expertise du marketing[91]. Elle peut en effet participer au processus d'embauche des nouveaux employés, détecter les conducteurs distraits et dangereux, mesurer la réponse émotionnelle d'un client face à une publicité ou encore analyser le niveau d'attention des étudiants en classe. Toutefois, la modélisation des émotions, combinée au ciblage émotionnel, peut aussi avoir pour objectif d'influencer les décisions d'une personne[92]. Dans ce cas précis, l'usage de la technologie serait un usage malveillant qui viendrait brimer la liberté de choix des individus.

Nous avons vu avec les recherches de Damasio, Kahneman et Tversky que nos émotions font partie et contribuent à notre processus de décision, ainsi, reconnaître les émotions d'un individu pourrait être la première étape qui permettrait d'influencer sa décision. L'utilisation de SRÉ pourrait donc brimer la liberté de choix des individus lorsqu'elle fait l'objet d'une utilisation malveillante. La liberté de choix renvoie à la capacité d'un individu de choisir lui-même ce qui est bon (ou mauvais) pour lui[93]. La liberté de choix est ici vue comme une partie de l'autonomie de l'agent rationnel, qui est autonome lorsqu'il

---

[89] Hamdi, H. « Plate-forme multimodale pour la reconnaissance d'émotions via l'analyse de signaux physiologiques: application à la simulation d'entretiens d'embauche ». *Thèse de doctorat*. Université d'Angers, France, 2012.
[90] Darpy, D. et Guillard, V. *Comportements du consommateur; concepts et outils,* Dunod (4e éd.), 2016, https://books.google.ca/books?hl=fr&lr=&id=LbMcDQAAQBAJ&oi=fnd&pg=PP3&dq=GAFA+et+%C3%A9motions&ots=nh-Kh2iHw4&sig=7zcjNYpe5ksa4mCtbqNVsjtbY9Y&redir_esc=y#v=onepage&q=%C3%A9motion&f=false, p. 243.
[91] Le ciblage émotionnel peut être utilisé dans des domaines aussi variés que la sécurité, la politique, les processus d'embauche, l'éducation, le *care*, la recherche, etc. Voir Crawford, K. *The Atlas of AI : Power, Politics and the planetary cost of Artificial Intelligence*, Yale University Press, New Haven and London, 2021, p. 152.
[92] Crawford, K. *The Atlas of AI : Power, Politics and the planetary cost of Artificial Intelligence*, *op. cit*., p. 153.
[93] Arpaly, N. *Unprincipled Virtue: An Inquiry Into Moral Agency*, Oxford University Press, New York, 2003, p. 118.

lui est possible de faire le choix qui lui convient le mieux, parmi ceux à sa disposition[94]. Ainsi, le risque que les SRÉ soient en mesure de brimer la liberté de choix des individus n'est pas seulement un risque envisagé de manière prospective. Ce risque, de manipulation des décisions d'autrui à travers une analyse de ses états émotifs peut, par exemple, menacer la démocratie.

Lors des élections américaines de 2010, Facebook avait lancé une étude expérimentale sur sa plateforme pour estimer son pouvoir d'influence sur le comportement de ses utilisateurs[95]. En créant le bouton « I voted »[96], Facebook était non seulement en mesure de recenser le taux de participation aux élections américaines mais aussi l'influence qu'un tel ajout pouvait avoir sur les individus. Les résultats obtenus démontrèrent que les utilisateurs étaient plus susceptibles d'aller voter s'ils savaient que leurs amis et les membres de leur famille y étaient déjà allés[97], ce qu'ils voyaient via les publications de ceux-ci. Le simple fait de voir que l'acte d'aller voter semblait populaire était suffisant pour inciter les individus à le faire aussi. Suite à cette étude, certains chercheurs, dont le professeur de droit de l'Internet et de droit international de Harvard, Jonathan Zittrain, ont ouvertement manifesté leurs inquiétudes quant à la possibilité que des compagnies comme Facebook « aient le pouvoir de décider du résultat d'une élection sans que personne ne s'en aperçoive[98] ». Autrement dit, la modélisation informatique des émotions pourrait être utilisée à des fins malveillantes et non démocratiques, comme celle de biaiser les élections.

En juin 2016, l'équipe de campagne de Donald Trump engagea la firme Cambridge Analytica dans l'objectif d'utiliser le ciblage émotionnel sur 220 millions d'électeurs américains[99]. À la fois utilisateurs de Facebook et perçus comme « indécis » quant à leur choix politique, la firme leur adressait des « messages politiques personnalisés[100] » via Facebook, dans le but d'influencer les élections américaines en la faveur du parti républicain. Similairement aux objectifs marketing, la firme Cambridge Analytica aurait « saturé les électeurs [via des publicités ciblées] avec des messages soigneusement

---

[94] Ibid.
[95] O'Neil, C. *Weapons of math destruction*, Broadway Books, New York, 2017, p. 180.
[96] Bond, R., Fariss, C., Jones, J. *et al.* « A 61-million-person experiment in social influence and political mobilization », *Nature* 489**,** 295–298, 2012, https://doi.org/10.1038/nature11421
[97] O'Neil, C. *op. cit.,* p. 181.
[98] Zittrain, J. « Facebook Could Decide an Election Without Anyone Ever Finding Out ». *The New Republic*, 2014. In : Chander, A. « The Racist Algorithm? », *Michigan Law Review*, 115(6), 1023-1045, 2017, http://michiganlawreview.org/wp-content/uploads/2017/04/115MichL Rev 1023_Chander.pdf, p. 1024.
[99] Gonzalez, R. J. « Hacking the citizenry? Personality profiling, big data and the election of Donald Trump », *Anthropology Today*, 33(3), 2017, p. 10.
[100] Manokha, I. « Le scandale Cambridge Analytica contextualisé: le capital de plateforme, la surveillance et les données comme nouvelle « marchandise fictive » », *Cultures & Conflits*, 109, 2018, p. 40.

conçus[101] » en fonction de leur profil psycho-démographique[102]. Ce scandale met en avant, d'une part, le succès des collaborations interdisciplinaires entre par exemple des « data scientists » et des psychologues et, d'autre part, le potentiel malveillant de certaines de ces collaborations[103]. En effet, l'utilisation de la modélisation informatique des émotions et du ciblage émotionnel peut venir brimer la possibilité de faire librement des choix individuels. En démocratie plus particulièrement, l'adhésion libre et éclairée d'un électeur à un parti de son choix est un droit qui est menacé par le ciblage émotionnel en contexte d'élection. Ainsi, il est possible d'envisager des utilisations malveillantes des systèmes de reconnaissance des émotions qui briment la liberté de choix des individus. Cette utilisation des SRÉ peut causer des torts importants aux individus ainsi qu'à la société en général lorsqu'elle concerne des aspects fondamentaux de celle-ci comme sa démocratie. Une société démocratique ne peut rester démocratique si les citoyens n'ont pas la possibilité d'avoir une libre adhésion au parti de leur choix. Créer des fausses nouvelles, dans l'objectif de décevoir, de frustrer ou encore d'apeurer certains citoyens qui n'ont pas encore pris de décision concernant le parti auquel ils voudraient voter vient brimer leur liberté de choix en influençant leur décision de manière malveillante.

1.3.2. Bilan pour les SRÉ

L'utilisation malveillante des SRÉ n'est pas un risque prospectif mais un risque actuel, qu'il est nécessaire d'évaluer. Ainsi, de prime abord, il est évident qu'une utilisation malveillante des SRÉ ne réussirait pas à satisfaire les exigences de la justice sociale, soit (1) celle que chaque individu possède un certain nombre de droits et libertés comprenant notamment le droit à la vie, la sécurité et l'égalité pour tous et (2) que chaque individu doit recevoir un traitement juste, c'est-à-dire un traitement qui inclut l'impartialité et la constance dans son application, peu importe le processus d'application. Une utilisation malveillante des SRÉ tel que nous l'avons soulevé dans la section précédente peut porter atteintes aux droits et libertés fondamentaux des individus. En ce cas, le droit à l'autonomie et le droit de voter librement sont les droits bafoués par une utilisation malveillante des SRÉ. Par ailleurs, la malveillance est en contradiction avec la justice sociale. La malveillance étant une volonté d'action dirigée sur un individu ou un groupe d'individu, elle ne respecte pas le critère de traitement juste des

[101] Gonzalez, R. J. *op. cit.*, p. 9.
[102] Ibid., p. 11.
[103] O'Neil, C. *op. cit.*, p. 181.

individus en plus d'être intrinsèquement en incohérence avec la notion de justice sociale, qui est intimement liée au principe de bienfaisance.

1.4. Conclusion

Dans ce chapitre, nous avons pu attester à la fois de la place importante des émotions dans notre rationalité mais aussi des pouvoirs peut-être sous-estimés des SRÉ sur nos processus de prise de décision dans des champs très variés. Les philosophes réfléchissent depuis plusieurs millénaires à la manière dont nous devrions percevoir nos émotions. Certains croient qu'il est nécessaire de les ignorer si l'on désire acquérir une faculté de juger objective et rationnelle[104], d'autres croient au contraire que nous devrions apprendre le plus tôt possible à les connaître pour mieux les contrôler[105]. Ainsi, si les émotions sont nécessaires au raisonnement et qu'elles peuvent aujourd'hui, à l'ère de la révolution numérique, être stratégiquement manipulées par des SRÉ, il faut alors admettre que ces systèmes de reconnaissance des émotions représentent un risque non-négligeable quant à notre liberté de choix. Par ailleurs, l'utilisation des SRÉ dans l'objectif d'influencer nos choix peut facilement ouvrir la porte à une utilisation malveillante des SRÉ. Cette utilisation comporte évidemment plusieurs enjeux éthiques – que nous aborderons au chapitre 3 – et vient se positionner à l'encontre des exigences de la justice sociale.

[104] Voir notamment Kant, E. *Anthropology from a pragmatic point of view [1796]* (trad. M. J. Gregor), The Hague, Pays-Bas, 1974, book III « On The Appetitive Power », § 73, p. 119.
[105] Voir notamment Aristote, *Éthique à Nicomaque*, *op. cit*., livre 2, 1378a

Chapitre 2 Sur les théories de l'universalité des émotions

Dans ce chapitre, nous étudierons l'une des grandes prétentions des SRÉ, qui est celle de reconnaître « universellement » les états émotionnels intérieurs en analysant le visage des individus. Nous concentrerons en ce sens notre attention sur la question de l'universalité des émotions ainsi que sur la relation des émotions avec les expressions faciales. Existe-t-il quelque chose de l'ordre des émotions « fondamentales » qui serait présent chez tous les êtres humains? Si oui, s'affichent-elles fidèlement sur nos visages en expressions prototypiques ? Dans un premier temps, nous observerons les recherches de Darwin et sa théorie des émotions ainsi que les deux fonctionnalités qu'il observe aux expressions faciales d'émotions. Selon Darwin, les expressions faciales d'émotions auraient à la fois une cause biologique et une cause sociale. Dans un second temps, nous nous attarderons sur la recherche menée par Ekman et Friesen qui servit de preuve de l'universalité des émotions dans l'expression et la reconnaissance ainsi que les raisons pragmatiques pour lesquelles plusieurs SRÉ se fondent sur la théorie de base des émotions endossée par Ekman. Dans un troisième temps, nous expliquerons les différences entre deux théories de l'universalité de l'émotion que nous séparerons en première et deuxième « couche », soit la couche représentant l'intériorité et la couche représentant l'extériorité.

2.1. La théorie de l'universalité des émotions

Depuis plusieurs années – et notamment depuis le succès de la reconnaissance faciale – les systèmes de reconnaissance des émotions ont augmenté en popularité. Mais pour reconnaître des émotions, il faut dans un premier temps savoir ce qu'*est* une émotion. Il semble que nous sachions tous individuellement et *intuitivement* ce qu'est une émotion, sans pour autant être en mesure de l'expliquer de manière objective. En effet, il semble y avoir, dans le cas de l'émotion, une scission forte entre les connaissances scientifiques et phénoménologiques; des connaissances qui peuvent être calculées, mesurées et partagées et des connaissances qui sont vécues. Les manquements quant aux connaissances scientifiques à son égard semblent avoir influencé négativement les réflexions à son

sujet en perpétuant des fausses croyances, voire des mythes[106]. Sur des millénaires de réflexions humaines, nous n'avons pas été en mesure de circonscrire une définition des émotions communément acceptée[107] :

> « There is **no consensus** […] Indeed, there is every appearance of disagreement: Some researchers use categories, some dimensions; some use bipolar concepts, some unipolar ones; and some presuppose simple structure, some a circumplex, and some a hierarchy. The key culprit in this mess is the concept of emotion, or affect, as it is now sometimes called. **Emotion is too broad a class of events to be a single scientific category**. As psychologists use the term, it includes the euphoria of winning an Olympic gold medal, a brief startle at an unexpected noise, unrelenting profound grief […][108] » (La mise en gras des termes est de nous)[109]

Cette affirmation de notre incapacité à circonscrire, définir et enfin comprendre ce qu'est une émotion est déstabilisante puisque tous les êtres humains vivent, expérimentent et ressentent des émotions ; il semblerait donc plus que normal d'être, à tout le moins, en mesure de définir le phénomène. Mais comme Russelll et Barrett le soulignent, l'émotion est un concept large qui inclut des phénomènes parfois très différents les uns des autres ce qui rend leur étude difficile. Ce flou présent dans la définition du phénomène est peut-être la cause de plusieurs mésententes quant à ce qu'est une émotion et le rôle que nous devons lui accorder dans notre vie.

Dans son livre *The Expression of the Emotion in Man and Animals* (1872), Darwin défendait l'idée selon laquelle les expressions faciales seraient des manifestations de nos émotions[110]. Selon lui, les expressions faciales « émotionnelles » seraient issues de l'évolution et auraient, par conséquent, des origines biologiques qui expliqueraient et prouveraient leur caractère inné et conséquemment, leur

---

[106] À titre d'exemple, les émotions ont très longtemps été associé aux femmes et utilisées comme un argument appuyant l'idée que les femmes seraient, de nature, des êtres irrationnels puisque muent par leurs émotions plutôt que leur raison. Voir Clément, C. et Cixous, H. *La jeune née*, Union générale d'Éditions, Paris, 10/18, série Féminin Futur, 1975, p. 125.

[107] Fehr et Russel. « Concept of Emotion Viewed from a Prototype Perspective », *Journal of experimental psychology*, 113(3), 464-486. 1984, https://psycnet.apa.org/doi/10.1037/0096-3445.113.3.464. "Many have sought but no one has found a commonly acceptable definition for the concept of emotion." p. 464.

[108] Russel, J. A., et Barrett, L. F. « Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant », *Journal of Personality and Social Psychology*, 76(5), 805-819, 1999, https://psycnet.apa.org/doi/10.1037/0022-3514.76.5.805. p. 805.

[109] Malheureusement, aucun progrès significatif quant à une définition consensuelle de l'émotion n'a vu le jour depuis la publication de l'article de Russel et Barrett comme le témoigne les propos de Kate Crawford dans son livre *The Atlas of AI : Power, Politics and the planetary cost of Artificial Intelligence*, *op. cit.*, p. 17 : « But there is considerable scientific controversy around emotion detection, which is at best incomplete and at worst misleading »

[110] Ekman, P. « The argument and evidence about universals in facial expressions of emotion », Chap. 6 in H. Wagner and A. Manstead (Eds.), *Wiley Handbook of social psychophysiology*, *Handbook of social psychophysiology*, University of California, San Francisco, 1989, https://psycnet.apa.org/record/1989-97735-006, p. 144.

universalité[111] chez les humains. Ainsi, nous pourrions étudier et comprendre les émotions en tournant notre attention sur les expressions faciales.

Pour Darwin, il ne fait aucun doute que les expressions émotionnelles jouent une part importante dans la fonction communicative des espèces (humaine et autres). Darwin considèrera certaines expressions d'émotions comme universelles, du fait qu'elles sont transmises génétiquement à tous les membres d'une même espèce. Sa théorie, qui se classe dans l'approche des émotions de base (« basic emotions approach »), comprend notamment deux grands principes : le principe de « l'habitude serviable » (Principle of Serviceable Habits) et le principe de « l'antithèse ».

Le principe de « l'habitude serviable » de Darwin défend l'idée que tous les êtres humains partageraient des expressions faciales émotionnelles spécifiques entre eux – et même avec certains animaux qui auraient des expressions faciales émotionnelles similaires. L'un des rôles des expressions émotionnelles dans la fonction communicative serait la transmission des « états intérieurs » par une manifestation extérieure[112]. En ce sens, un membre d'une même espèce est en mesure de reconnaître certaines expressions émotionnelles (comme la joie, la colère, la tristesse, etc.) chez un autre membre simplement en observant ces expressions faciales. Toutefois, le principe de « l'habitude serviable », qui lui valide l'universalité des expressions émotionnelles, est suivi d'un autre principe qui lui, apporte des nuances importantes. En effet le principe de « l'antithèse » porte son attention sur des facteurs qui valident l'universalité des expressions mais qui ne valident pas nécessairement les expressions faciales émotionnelles comme étant une fonction de communication des états émotionnels intérieurs.

Le principe de l'antithèse démontre que les expressions faciales servent parfois à communiquer des traits plutôt que des états émotionnels. Darwin avançait l'idée par exemple que certaines expressions faciales avaient pour objectif de mettre de l'avant des traits souhaitables (desirable traits)[113]. Par exemple, une personne désirant avoir l'air menaçant devant un danger peut contracter ses muscles faciaux en une expression de colère, même si la personne ne ressent pas de la colère. Exprimer la colère plutôt que la peur par exemple peut être décisif lorsque l'objectif est d'intimider son rival pour

[111] Barrett, L. F. *How Emotions Are Made*. Houghton Mifflin Harcourt, New York, p. 4.
[112] Hess, U. et Thibault, P. « Darwin and Emotion Expression », *American Psychological Association*, 64(2), p. 120 –128, 2009, https://psycnet.apa.org/fulltext/2009-01602-003.pdf?auth_token=a75e8192b84781dc3dc4facff1a983dedf27d78e. p. 120.
[113] Darwin, C. « The expression of the emotions in man and animals », *University of Chicago Press*, Chicago, 1965, p. 103. In Hess, U. et Thibault, P. *op., cit*., p. 121.

éviter un affrontement ou pour se préparer à un affrontement inévitable. En effet, exprimer de la colère nous porte à amener notre corps vers l'avant tandis que la peur nous porte à se rétracter vers l'arrière.

En ce sens, Darwin soulève l'idée que les expressions faciales pourraient être à la fois (1) des expressions d'émotions (froncer les sourcils pour exprimer sa colère) et (2) des signaux de communications (froncer les sourcils pour démontrer que nous sommes prêts à l'affrontement même si nous éprouvons de la peur)[114]. Donc (c), pour Darwin, les expressions faciales ne communiquent pas toujours nos émotions intérieures. Les expressions faciales renverraient à la fois à des processus biologiques d'émotions et à des normes et règles sociales. Ces fonctions des expressions faciales permettent aussi de comprendre que si nous utilisons universellement des expressions faciales au sens où tout le monde utilise les muscles de son visage en des schèmes précis (froncement des sourcils, pincements des lèvres, écarquillements des yeux, etc.), il n'est toutefois pas toujours possible de savoir la différence entre une personne qui exprime une émotion intérieure et une personne qui désire communiquer un signal à autrui.

2.1.1. La preuve de l' « universalité » des émotions dans l'expression et la reconnaissance faciale

Dans plusieurs articles dont *The Argument and Evidence about Universals in Facial Expressions of Emotion*[115], le psychologue Paul Ekman défend la théorie de l'universalité des émotions en se basant sur les recherches de Darwin[116]. Déclarant suivre les traces des études darwiniennes sur l'expression et la reconnaissance d'expression faciales émotionnelles, Ekman défend l'idée selon laquelle les émotions seraient à la fois *exprimées* et *reconnues* de manière universelle par tous les êtres humains et ce, peu importe leur culture[117].

---

[114] Cette notion a été testée pour la première fois par Hess, U., Banse, R., & Kappas, A. « The intensity of facial expression is determined by underlying affective state and social situation », *Journal of Personality and Social Psychology, 69*(2), 280–288. https://doi.org/10.1037/0022-3514.69.2.280 qui ont montré que les sourires varient à la fois en fonction du contexte social (et donc des motifs sociaux) et du contenu émotionnel du stimulus. Voir Hess, U. et Thibault, P. *op., cit*., p. 122.

[115] Ekman, P. « The argument and evidence about universals in facial expressions of emotion », chap. 6 in H. Wagner and A. Manstead (Eds.), *Wiley Handbook of social psychophysiology*, *Handbook of social psychophysiology*, University of California, San Francisco, 1989, https://psycnet.apa.org/record/1989-97735-006

[116] Crawford, K. *The Atlas of AI : Power, Politics and the planetary cost of Artificial Intelligence*, Yale University Press, New Haven and London, 2021, p. 158.

[117] Ibid., p. 151.

Dès 1972, il postule un modèle de six émotions de base où chacune d'entre elles est associée à un prototype d'expression faciale spécifique dans l'objectif d'établir l'existence de liens entre les émotions dites de base (joie, tristesse, peur, dégoût, colère, surprise) et des expressions faciales spécifiques (ex. froncement des sourcils = colère, sourire = joie, etc.). Cette théorie fut empiriquement testée par Ekman et son collègue, Friesen, qui réalisèrent une étude expérimentale auprès des Dani de Nouvelle-Guinée. Les Dani avaient été favorisés par Ekman et Friesen sur les autres cultures en raison de l'absence d'échange (ou presque) que ce groupe d'individus avait eu avec toutes autres cultures[118] et plus spécifiquement avec les Américains.

Dans cette étude, Ekman et Friesen[119] utilisèrent la méthode appelée « susciter des postures[120] ». Cette méthode consistait à demander à des individus d'une même culture – les Dani (aussi parfois nommés Ndani) de Nouvelle-Guinée – d'exprimer sur leur visage l'émotion appropriée à la phrase qu'ils recevaient. Par exemple, si le chercheur disait « tu es triste parce que ton enfant vient de mourir », l'individu devait exprimer une expression faciale représentant sa tristesse fictive. Au moment de l'expression, les chercheurs prenaient des photographies du visage de l'individu. Par la suite, ils retournèrent aux États-Unis où ils présentèrent les photographies à des Américains, leur demandant d'identifier l'émotion qu'ils voyaient sur le visage du Dani. L'idée était que si deux cultures qui n'avaient jamais eu de contact antérieur exprimaient et reconnaissaient des émotions spécifiques, la double universalité des émotions serait prouvée. De fait, l'étude fut considérée comme une réussite prouvant l'universalité des expressions faciales émotionnelles compte tenu que les individus de la deuxième culture – les Américains – avaient réussi à identifier les émotions adéquates à un pourcentage plus élevé que celui de la simple chance[121]. En effet, puisque pour Ekman, les émotions de base se limitent au nombre de six (joie, colère, tristesse, peur, dégoût et surprise), les individus qui tentaient de deviner les émotions sur les photographies avaient une liste comprenant six choix devant eux. Cette étude fut un véritable succès aux États-Unis ; l'absence de contact prouvait que l'expression des émotions n'avait pas pu être transmise d'une culture à l'autre et donc que les expressions émotionnelles étaient universelles et innées.

[118] Ekman, P. *op. cit*., p. 150.
[119] Ibid., p. 150.
[120] Traduction libre de : « *eliciting poses* »
[121] Ekman, P. *op. cit*., p. 150.

Illustration 2.1. Les expressions faciales émotionnelles universellement reconnues

Source image : https://www.paulekman.com/blog/darwins-claim-universals-facial-expression-challenged/.

Toutefois, il est important de noter que trois problèmes ont été identifiés par les auteurs comme limitant le succès de cette étude. Dans un premier temps, ce type d'étude n'avait jamais été fait auparavant en ce sens qu'aucune étude n'avait jamais été effectuée auprès d'une culture aussi isolée[122]. Ensuite, des six émotions de bases exprimées par les Dani, seules quatre d'entre elles étaient réellement reconnues par les Américains. En effet, les Américains ne semblaient pas en mesure de distinguer les expressions faciales de peur et de surprise sur les photographies. Il est d'ailleurs possible de remarquer que sur l'illustration 2.1. – qui sont les photographies originelles d'Ekman et Friesen – n'apparaissent que quatre photographies et que la peur et la surprise ont été retirées. Enfin, les expressions faciales n'étaient pas des expressions spontanées réelles mais plutôt le résultat d'un « acte de théâtre » qui n'impliquait aucun engagement émotionnel réel[123]. Conséquemment, les Dani ne réagissaient que de la manière dont ils *croyaient* qu'ils *devaient* réagir. Ces trois observations permettent de réaliser que l'étude d'Ekman et Friesen comporte des limites non-négligeables. L'incapacité des Américains à différencier l'expression prototypique[124] de la peur et de la surprise, par exemple, vient remettre en question les prétentions à l'universalité, qui auraient exigées une identification complète des six émotions de base. Ainsi, si les preuves quant à une reconnaissance des six émotions identifiées par Ekman sont généralement acceptées par la communauté scientifique au sein des membres d'une même culture, les preuves quant à une reconnaissance de ces six émotions de base à travers les cultures ne fait pas l'objet d'une telle acceptation[125].

[122] Ekman, P. *op. cit.*, p. 150.
[123] Barrett, L. F. *op cit.*, p. 4.
[124] Le terme prototypique signifie ici que l'expression émotionnelle est le stéréotype ou le symbole parfait de l'émotion en question selon les normes sociales d'une culture donnée.
[125] Schenk, F. « Les émotions de la raison », *Revue européenne des sciences sociales*, XLVII-144, 2009, 151-162, https://journals.openedition.org/ress/75#citedby. p. 153.

2.1.2. SRÉ basés sur la théorie de l'universalité

L'idée d'une double universalité des émotions, au sens où elles seraient vécues d'un côté et exprimées et reconnaissables universellement de l'autre, est une idée qui peut servir d'assise théorique aux SRÉ[126]. En effet, la théorie des émotions de base donne les fondements nécessaires (c'est-à-dire les six émotions ainsi que les expressions faciales y correspondant) aux systèmes de reconnaissance des émotions[127] pour que ces derniers puissent s'entraîner avec des jeux de données précis. Avec ces six émotions de base, il apparaît possible d'entraîner des logiciels à reconnaître les expressions faciales prototypiques. Étant donné l' « adéquation[128] » soutenue par Ekman entre les expressions faciales prototypiques et les émotions vécues par les personnes, il est possible de conclure qu'à partir d'une analyse attentive des expressions du visage, il est possible d'avoir accès aux émotions intérieures des individus.

Il est d'ailleurs possible de constater que les modèles de SRÉ se basant sur la théorie de l'universalité (en ignorant les autres fonctions des expressions faciales émotionnelles) constituent la grande majorité des modèles existants à ce jour[129]. Si plusieurs techniques existent afin d'établir une corrélation entre une expression faciale et une émotion, certains modèles sont plus utilisés que d'autres comme le système FACS d'Ekman et Friezen (*Facial Action Coding System*)[130] qui permet d'identifier les points déterminants à surveiller lors d'expressions faciales d'émotions. L'idée est qu'en identifiant des points d'intérêts du visage (yeux, bouche, sourcils, nez), puis en codant ces points d'intérêts à surveiller, il est par la suite possible de comparer ces points d'intérêts entre le visage neutre et l'une des six expressions d'émotions prototypiques. En calculant la variation entre les distances de points spécifiques (lieu des points d'intérêts sur visage neutre versus sur visage expressif), il est possible de

---

[126] McStay, A. « Empathic Media: The Rise of Emotion AI », *Arts & Humanities Research Concil*, 2016, p. 4
[127] Ibid., p. 4
[128] Nous mettons le terme entre guillemets puisque comme nous l'avons expliqué dans la conclusion de la section précédente, l'adéquation est questionnable.
[129] Khadija Lekdioui. *Reconnaissance d'états émotionnels par analyse visuelle du visage et apprentissage machine. Synthèse d'image et réalité virtuelle*, Université Bourgogne Franche-Comté, Université Ibn Tofail. Faculté des sciences de Kénitra, 2018. ; Faiza Khalfi. *Reconnaissance automatique des émotions par données multimodales : expressions faciales et des signaux physiologiques*, Université Paul Verlaine - Metz, 2010.
[130] S.A. « Facial Action Coding System », *Paul Ekman Group*, https://www.paulekman.com/facial-action-coding-system/. (Page consultée le 4 avril 2021)

classer par la suite les expressions faciales dans l'une des six catégories[131] en moins de 0.03132 secondes[132]. Cette technique permettrait donc d'identifier l'émotion vécue par la personne.

Illustration 2.2. Les points spécifiques pour la détection des expressions faciales émotionnelles

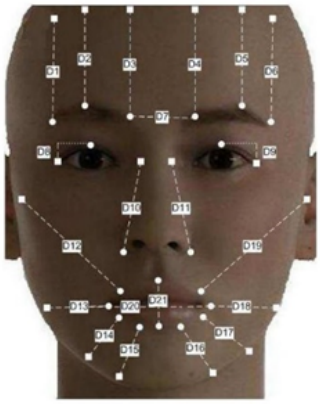


FIGURE 2.14 – Distances utilisées pour le codage des expressions faciales

Source image : Khalfi, F. *op. cit.* p. 66.

Conséquemment, plusieurs technologies de reconnaissance des émotions ont été créés à diverses fins (sécurité, marketing, politique, processus d'embauche, le *care*, l'éducation, etc.[133]) dans l'objectif d'avoir accès aux états émotionnels intérieurs des individus ciblés.

Les aéroports chinois, par exemple, sont depuis quelques années équipés de systèmes de reconnaissance des émotions qui viennent s'ajouter à ceux déjà présents de la reconnaissance faciale[134]. Dans une volonté d'augmenter la sécurité, le gouvernement chinois utilise des systèmes de reconnaissance des émotions dans l'objectif d'identifier des individus recherchés mais aussi de détecter à l'avance des individus démontrant les signes précurseurs d'intentions malveillantes ou agressives[135]. Dans plusieurs pays dont la Chine, les États-Unis et la Grande-Bretagne, il en va de même avec les stations de trains et d'autobus, les banques, les hôpitaux[136], les parcs, les écoles[137],

[131] Faiza Khalfi. *op. cit.*, p. 9.
[132] Ibid., p. 75.
[133] Crawford, K. *The Atlas of AI*, p. 152.
[134] McStay, A. « Emotional AI, Soft Biometric and the Surveillance of Emotional Life: an Unusual consensus on privacy », Big Data & Society, *Journal Sage Pub*, 2020, p. 8.
[135] Wong et Liu 2019 dans McStay, A. Emotional AI, Soft Biometric and the Surveillance of Emotional Life, p. 8.
[136] Gillum, J. et Kao, J. « Aggression Detectors : The Unproven, Invasive Surveillance Technology Scholls Are Using to Monitor Students », *ProPublica*, 2019, https://features.propublica.org/aggression-detector/the-unproven-invasive-surveillance-technology-schools-are-using-to-monitor-students/.
[137] Ibid.

les centres d'achat[138], les compagnies privées[139], les événements, etc., où les technologies de reconnaissance des émotions deviennent une tendance de plus en plus populaire[140] pour assurer la sécurité. Une compagnie japonaise, la compagnie Vaak, prétend que son système de reconnaissance des émotions est en mesure d'identifier un vol à l'étalage avant même que l'individu ne commette son crime. Cette reconnaissance des signes précurseurs serait possible grâce à l'algorithme, qui aurait analysé, à partir de vidéos enregistrés de vol à l'étalage, les expressions faciales, les comportements et l'habillement « habituel » des voleurs[141].

Cette possibilité, de reconnaître une action chez un individu avant que celle-ci soit commise, peut sembler comporter des avantages non-négligeables. Devant le nombre élevé de fusillades chaque année dans les écoles aux États-Unis par exemple, certains établissements scolaires américains seraient susceptibles de considérer les technologies de reconnaissance des émotions comme une option raisonnable, qui pourrait sauver la vie de plusieurs personnes. Toutefois, d'autres pays, notamment plusieurs pays sous le RGPD[142] (comme la Suisse ou encore la France), s'opposent à l'utilisation des technologies de reconnaissance des émotions dans plusieurs situations comme à l'intérieur des écoles. Une tension est présente entre la volonté d'accroître la sécurité et l'augmentation des risques d'atteintes à certains droits et libertés fondamentaux des individus. Le point de tension entre ces décisions opposées réside dans le fait que ces technologies de reconnaissance des émotions détiennent des données sensibles tel que des données biométriques faciales, vocales, physiologiques, etc. La collecte de telles données est, d'un côté, nécessaire et inévitable pour l'utilisation de SRÉ et, de l'autre, elle a, par définition, des répercussions importantes sur la vie des individus[143], qui se verraient dans certains cas obligés de s'y soumettre. Il est évident que certaines compagnies et

[138] Voir les compagnies tel que EyeSee, « Making behavioral insights accessible », https://eyesee-research.com/ (Page consultée le 7 avril 2021) Et Nippon Electric Company, « NeoFace Watch », https://www.nec.com/en/global/solutions/biometrics/face/neofacewatch.html (Page consultée le 3 mars 2021)
[139] Calistra, C. « Emotion Analysis in the real world », Kairos, 2015, https://www.kairos.com/blog/emotion-analysis-in-the-real-world (Page consultée le 25 avril 2021)
[140] Article 19. « Emotional Entanglement: China's emotion recognition market and its implications for human rights », Article 19, Free World Center, Londres, 2021, https://www.article19.org/emotion-recognition-technology-report/. p. 13.
[141] Vaak. « Automatiser les opérations avec l'IA d'analyse », VAAK, https://vaak.co/ (Page consultée le 4 avril 2021)
[142] Journal officiel de l'Union européenne. « Règlement du parlement européen et du conseil », Journal officiel de l'Union européenne, 2016, https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32016R0679 (Page consultée le 3 février 2020)
[143] Commissariat du Canada à la protection de la vie privée. « Des données au bout des doigts : La biométrie et les défis qu'elle pose à la protection de la vie privée » Commissariat du Canada à la protection de la vie privée, 2011, https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privee/renseignements-sur-la-sante-renseignements-genetiques-et-autres-renseignements-sur-le-corps/gd_bio_201102/ (Page consultée le 14 avril 2021)

certains pays sont prêts à privilégier la sécurité sur d'autres valeurs comme l'autonomie ou la vie privée, mais pour que le gain en sécurité soit acceptable, la perte au chapitre de la vie privée ou de l'autonomie doit être à tout le moins proportionnelle à ce gain[144]. Conséquemment, certaines écoles américaines pourraient juger que le gain est proportionnel à la perte si les SRÉ démontrent qu'ils peuvent prévenir les attaques à mains armées à l'intérieur des établissements scolaires en détectant les signes précurseurs d'une agression par l'analyse des expressions faciales, du ton de voix, de la gestuelle, etc[145].

2.1.3. Les différentes « couches » d'universalité

Les théories de l'universalité des émotions ne portent pas toutes sur le même aspect. En effet, dans un premier temps, il y a l'idée selon laquelle nous ressentons tous, de manière universelle, certaines émotions précises de base comme la peur, la tristesse, la joie, la colère, le dégoût et la surprise. Dans un second temps, il y a l'idée que nous exprimons *et* reconnaissons ces émotions « universelles » sur le visage des autres et ce, de manière universelle. Les théories qui nous intéressent dans ce projet de recherche – que nous nommons ici les théories de la double-universalité des émotions – appartiennent à cette deuxième idée et en ce sens, s'intéressent à la « couche supérieure » qui est elle-même bâtie sur une première catégorie de théories de l'universalité des émotions.

Dans l'idée que tous les êtres humains ressentent certaines émotions dites de « base », il y a aussi l'idée que ces émotions de base seraient des émotions innées. Toutefois, s'il y avait réellement quelque chose de l'ordre de l'universalité des émotions de base, cette universalité devrait en principe engendrer une sélection unanime et consensuelle d'un nombre précis d'émotions, ce qui n'est pas le cas. Les émotions de base, prétendument innées, sont chez les uns au nombre de six[146], chez les autres au nombre de quatre, cinq, huit, onze, dix-sept, etc.[147]. Le défi de prouver l'existence d'une universalité

---

[144] Commissariat du Canada à la protection de la vie privée « Des données au bout des doigts : La biométrie et les défis qu'elle pose à la protection de la vie privée ».
[145] Nous reviendrons sur la légitimité de l'utilisation des SRÉ dans le chapitre suivant.
[146] Ekman, P. « An argument for basic emotions », *Cognition and Emotions*, 6(3/4), University of California, San Francisco, 1992, https://www.tandfonline.com/doi/abs/10.1080/02699939208411068
[147] Plutchik (Plutchik, 1980) en compte huit (acceptation, colère, anticipation, dégoût, joie, peur, tristesse, surprise). Schaver et Schwarz (Schaver, Schwartz et al., 1987) en compte cinq (peur, surprise, joie, colère, tristesse), Kemper (Kemper, 1987) en compte 4 (peur, colère, dépression, satisfaction, Izard (Izard, 1977) en compte onze (joie, surprise, colère, peur, tristesse, mépris, détresse, intérêt, culpabilité, honte, amour), Frijda (Frijda, 1986) en compte dix-sept (parmi

dans les émotions de base réside donc dans la preuve. Pour prouver leur existence, les chercheurs n'ont d'autres choix que d'observer des personnes « cobayes ». Cette technique consiste à tenter d'apercevoir la première couche, la couche intérieure (les émotions innées, produits du biologique) à travers la deuxième couche, le couche extérieure (l'expressions des émotions influencées par les normes sociales et la volonté de communiquer un signal quelconque[148]). Cette difficulté importante que représente l'obligation de passer par la couche extérieure peut être en partie la cause de l'irréductible diversité dans les résultats des chercheurs sur le nombre d'émotions dites de base[149].

Toutefois, avec des psychologues comme Lisa Feldman Barrett, une autre hypothèse est possible. Selon la théorie des émotions construites (*Theory of constructed emotion*), nous ne « ressentons » pas – d'une manière innée et biologique – des émotions précises comme la joie ou la tristesse, mais nous ressentons plutôt ce qui se rapproche davantage à des excitations (ou des absences d'excitations) accompagnés de plaisir ou de déplaisir[150], que notre corps identifiera par la suite à une émotion précise – comme la joie – selon notre classification personnelle des émotions (classification qui est influencée par notre culture, notre époque, les normes sociales, etc.). Ainsi, dans une situation donnée, nous pouvons lier une grande excitation accompagnée d'un plaisir à une émotion telle que la joie alors que dans une autre situation elle peut être liée à la satisfaction, la surprise (positive), l'amour, la confiance, l'intérêt, etc. Au contraire, une excitation accompagnée d'un déplaisir peut être liée dans une situation donnée à l'émotion du dégoût alors que dans un autre contexte elle aurait pu être associée par la même personne à la colère, à la peur, à la honte, à la détresse, etc. Suivant cette théorie, il est possible de dire que tous les êtres humains, de manière universelle, ressentent des sensations d'excitations accompagnée de plaisir ou de déplaisir, et qu'ils les associent dans un second temps à une émotion particulière[151].

---

lesquelles l'arrogance, la confiance, la peine, l'effort, etc.). Voir Le Breton, D. « Sociologie des émotions : Critique de la raison darwinienne », *Recherches sociologique*, 1, 1998, https://sharepoint.uclouvain.be/sites/rsa/Articles/1998-XXIX-1_05.pdf, p. 46.

[148] Voir la section 2.1. sur Darwin

[149] Crawford, K. Atlas of AI, *op.cit*., p. 153.

[150] Barrett, L. F. *op. cit*., p. 32-33.

[151] La question quant à savoir si ce que nous ressentons est universel n'est toutefois pas l'objet de ce mémoire. Nous nous concentrerons plutôt sur les théories de la deuxième couche ; nous tenterons de démontrer que l'endossement d'une version « simplifiée » de la théorie de la double-universalité des émotions par les SRÉ peut avoir des impacts négatifs importants sur la vie des individus qui y sont soumis puisque sa simplification nie l'existence des limites contextuelles et culturelles au sein de nos sociétés. Ce refus de prendre en considération les limites contextuelles et culturelles engendre des risques éthiques et sociaux considérables qui s'appliquent à des groupes souvent déjà identifiés comme vulnérables dans la société (tel les femmes et les minorités visibles).

L'universalisme permet de dire qu'une chose est vraie, peu importe le point de vue, peu importe le cadre de référence. Cela signifie que la théorie de l'universalité des émotions demanderait ici que tous les êtres humains expriment les mêmes émotions de la même manière et reconnaissent les émotions des autres. Simplifier la théorie de la double universalité des émotions vient affirmer un universalisme dans la première et dans la deuxième couche soit dans la possession innée d'un nombre précis d'émotions et d'une relation unique entre une expression faciale et une émotion vécue précise. Autrement dit, la simplification de la théorie de la double universalité des émotions vient nier l'idée que nous ressentons, exprimons et reconnaissons les émotions de manières différentes[152].

2.2. Conclusion

Dans cette section, nous avons regardé avec Darwin les fondements de la théorie de l'universalité des émotions et avons mis l'accent sur les deux fonctionnalités de l'expression de l'émotion, ce qui nous a permis de comprendre la première grande difficulté dans la quête de l'observation et l'analyse des états émotionnels intérieurs. En effet, les états émotionnels intérieurs ne sont pas la seule cause de l'expression faciale d'émotion puisqu'ils se situent aussi aux côtés des signaux de communication. Ensuite, nous avons expliqué que la théorie de l'universalité des émotions n'a pas encore été prouvé en dépit de certaines recherches. Par ailleurs, la simplification d'une théorie peut mener à de fausses croyances comme celle de l'universalité de l'émotion dans l'expression faciale et la reconnaissance des expressions faciales. Sur ces fausses croyances se basent des SRÉ qui, en plus de menacer la vie privée par la récolte de renseignements sensibles (telles les données biométriques faciales), serviront d'aide à la décision dans des aspects importants de la vie des individus et peuvent conséquemment nuire à l'égalité des chances et l'égalité dans les opportunités de vie.

[152] Notre but ici n'est pas de réfuter la théorie d'Ekman mais plutôt l'idée simplifiée qui en découle et qui est à la source des développements des systèmes de reconnaissance des émotions.

Chapitre 3 Les risques d'une théorie de l'universalité des émotions

Dans ce chapitre, nous partirons de la prémisse qu'il est vrai que tous les êtres humains ressentent universellement des émotions – sans pour autant nous positionner quant à ce que ce soient des émotions précises (joie, tristesse) où des excitations accompagnées de plaisir ou de déplaisir que nous lierons par la suite à une émotion spécifique comme la joie ou la tristesse. Dans un premier temps, pour réfuter la prétention à l'universalité, non dans le vécu ressenti mais dans l'expression et la reconnaissance des émotions, nous identifions la complexité du contexte ainsi que la diversité culturelle comme principaux facteurs déterminants venant influencer drastiquement l'expression et la reconnaissance des émotions des individus particuliers. Dans un deuxième temps, nous démontrerons que l'influence du contexte sur nos expressions faciales émotionnelles apporte des difficultés non négligeables aux SRÉ qui ont pour objectif d'identifier les états émotionnels intérieurs en situations concrètes. Dans un troisième temps, nous démontrerons que, de même, l'universalité des SRÉ est, tout au plus, une universalité néo-occidentale qui expose l'incapacité à la fois de reconnaître et de prendre en compte les différences culturelles entre les sociétés et à l'intérieur de nos sociétés.

3.1 Le contexte

Les SRÉ ne sont pas, pour le moment, en mesure de tenir compte du contexte entourant une situation donnée. Or si l'on se fie au processus de reconnaissance des émotions de l'humain, il est possible de constater que le contexte est une partie incontournable de notre reconnaissance des émotions[153,154,155]. L'être humain fait sens d'une situation en se référant à plusieurs données variées. Dans une situation donnée impliquant des individus, l'humain prendra en compte les expressions faciales des individus, leur ton de voix, leur gestuel, leur propos, mais aussi les liens qui lient chaque individu à un autre si de telles données lui sont accessibles. Un visage expressif dans une situation

[153] Barrett, L. F. *op cit.*, p. 12-13.
[154] Chen, Z. et Whitney, D. « Tracking the Affective State of Unseen Persons », *Proceedings of the National Academy of Sciences*, 2019, https://www.pnas.org/content/pnas/early/2019/02/26/1812250116.full.pdf, p. 5.
[155] Ven, R. « Choose How You Feel; You Have Seven Options », *Institute of Network Cultures*, Amsterdam, 2017, https://networkcultures.org/longform/2017/01/25/choose-how-you-feel-you-have-seven-options/

particulière et dans un rapport social donné influence notre reconnaissance de l'émotion d'autrui. Le contexte à lui seul peut donc influencer notre manière individuelle d'exprimer des émotions mais aussi d'en reconnaître.

Cela signifie que, par exemple, je peux sourire dans plusieurs contextes sans nécessairement exprimer une émotion de joie à chaque fois. En effet, plusieurs études ont démontré que les expressions faciales n'étaient pas toujours liées à une émotion précise[156]. Je peux, dans certains contextes et pour des raisons diverses (personnelle, culturelle, etc.) exprimer de la gêne, de la honte, de la colère, du mépris, de la tristesse etc. et ce, par un sourire. Certaines personnes de nature plus « gênée » auront parfois tendance à sourire davantage alors que d'autres auront tendance à garder un air neutre, voire sérieux, en raison de la gêne qui les fige. En tant qu'être humain qui ressent, exprime et reconnaît des émotions dans une société donnée, il nous est possible de moduler notre expression faciale émotionnelle en (1) l'intensifiant, (2) l'atténuant ou (3) la neutralisant[157]. Ce libre-arbitre dans l'expression faciale émotionnelle peut être utilisé pour des raisons diverses mais est en général utilisé pour se plier aux règles d'expressivité qui sont issues des normes sociales. Certains contextes – enterrement, entrevue, réunion d'équipe, etc. – sont davantage régis par les normes sociales et sont plus susceptibles d'influencer les expressions faciales émotionnelles des individus[158].

Les règles sociales dictent certains comportements à adopter selon notre position sociale dans la société. Par exemple en Occident, il est généralement attendu des employés qu'ils masquent leurs émotions négatives[159] alors que les employeurs se donnent la « permission » de ressentir ouvertement ces mêmes émotions négatives[160]. Le sourire est l'expression faciale la plus utilisée pour masquer ses émotions négatives. Il est aussi un signe de déférence, qui affirme un respect envers la personne devant nous[161]. Ainsi, un employé sait qu'il est dans une position sociale hiérarchique inférieure à son employeur et cet écart dans les positions hiérarchiques peut moduler son expression faciale. En général, l'individu dans la position inférieure aura tendance à moduler son expression faciale vers une atténuation ou une neutralisation alors que l'individu dans la position supérieure aura tendance à

[156] Barrett, L. F. *op. cit*., p. 16.
[157] Tcherkassof, A. « Le sens dessus dessous des expressions faciales d'émotions : vers un nouveau tournant paradigmatique », *HAL*, Université Grenoble Alpes, France, 2018, https://hal.archives-ouvertes.fr/tel-01868279/document. p. 15.
[158] ibid.
[159] Ibid., p. 33.
[160] Ibid.
[161] Ibid.

accentuer son expression faciale[162]. Par exemple, un employé qui reçoit un refus de promotion de la part de son patron pourrait garder un sourire sur son visage alors qu'il est en colère tandis qu'inversement, un patron qui n'est pas satisfait du travail de son employé peut ouvertement lui manifester son insatisfaction. Dans les contextes où l'individu est dans une position d'infériorité sociale, les expressions faciales émotionnelles sont fréquemment utilisées pour communiquer avec les autres plutôt que de simplement afficher des émotions « intérieures ».

Ainsi, les règles sociales ne demandent pas seulement de dissimuler certaines émotions dans certains contextes (ex. la colère envers son patron) mais elles demandent aussi d'afficher des émotions spécifiques dans certains contextes (ex. afficher la majorité du temps un sourire lorsque l'on est un employé). L'expression faciale émotionnelle a vraisemblablement une fonction sociale accolée à sa fonction biologique de base[163] qui démontre que l'expression faciale émotionnelle permet à l'individu d'affirmer (ou réaffirmer), d'établir (ou rétablir) « à ses partenaires sociaux une certaine relation interpersonnelle dans un contexte interactionnel donné[164] ».

À l'intérieur des règles sociales qui permettent d'affirmer la relation interpersonnelle dans un contexte donné existent des structures de pouvoir entre individus à des niveaux hiérarchiques sociaux divergents, comme nous venons de le voir avec un employé et son patron, mais aussi entre les genres (homme et femme). En effet, les règles sociales étant apprises dès un jeune âge, il est possible d'observer l'apparition d'« émotions sexuées ». Ces stéréotypes émotionnels viendront classer certaines émotions comme étant « typiquement » masculines ou féminines et seront davantage présentes chez les individus de ce genre. Par exemple, les femmes exprimeraient – selon ces stéréotypes – plus fréquemment des émotions de bonheur, de tristesse ou de peur alors que les hommes exprimeraient davantage de la colère, du mépris ou du dégoût. Ces émotions « sexuées » ainsi que les expressivités émotionnelles leur étant associées sont d'ailleurs fréquemment utilisées pour justifier un classement arbitraire entre les genres[165] (les femmes étant identifiées comme le genre inférieur et les hommes comme le genre supérieur). Les femmes expriment plus fréquemment des émotions qui impliquent la vulnérabilité comme la tristesse, la peur ou la honte mais aussi des émotions qui sont dites pro-sociales comme l'empathie, l'enthousiasme ou le bonheur[166]. En étant identifiée

[162] Ibid., p. 32.
[163] Ibid., p. 33.
[164] Ibid.
[165] Ibid.
[166] Ibid., p. 34.

comme le genre « inférieur », il est davantage toléré que les femmes expriment des émotions de faiblesse comme la tristesse ou la peur dans des contextes sociaux particuliers alors que de telles expressions émotionnelles sont en général moins tolérées chez les hommes. Pleurer à la suite d'un refus pour une promotion serait donc davantage acceptée d'une femme que d'un homme pour cette raison. De même, la colère est une expression émotionnelle moins socialement acceptée chez les femmes que chez les hommes[167]. Toutefois, les règles sociales sont en adéquation avec les stéréotypes sexués de chaque culture et peuvent, en ce sens, diverger grandement d'une culture à l'autre. Par exemple, une femme qui exprime une grande timidité en rougissant devant un homme est une expressions faciale émotionnelle acceptée dans les pays occidentaux. Selon les règles sociales, la timidité d'une femme devant un homme « montre qu'elle reconnaît la supériorité de son interlocuteur[168] ». Cette expression faciale émotionnelle est cependant considérée comme défavorable dans les sociétés qui ne valorisent pas la différence de statut[169]. Dans la vie quotidienne, les expressions faciales émotionnelles sont fréquemment effacées, perturbées, limitées, etc.[170], par l'adhérence collective et individuelle aux normes sociales. Les normes sociales peuvent nous inciter à exprimer des émotions qui divergent de notre expérience vécue intérieurement et qui transformeront donc nos expressions faciales émotionnelles en une fonction de communication volontaire d'un signal à autrui (communiquer l'expression qui est en cohérence avec la norme), plutôt que celle d'expression des états émotionnels intérieurs (qui n'aurait pas été en cohérence avec la norme).

Par ailleurs, plusieurs recherches indiquent qu'un même individu exprimera une même émotion de différentes manières selon les situations[171]. En effet, les recherches de la psychologue Barrett et son équipe viennent confirmer que, s'il est vrai qu'il arrive que nous souriions lorsque nous sommes heureux, nous fronçons les sourcils lorsque nous sommes en colère, nous relevons les sourcils lorsque nous sommes surpris, etc., notre façon de communiquer notre joie, notre colère ou notre surprise varie selon plusieurs facteurs tels que la situation, notre genre, notre âge, notre culture d'appartenance, etc. De plus, une même personne peut exprimer une même émotion de différentes manières selon le contexte. Par exemple, il est vraisemblable de penser qu'à certains moments, nous avons froncé les

[167] Ibid.
[168] Ibid.
[169] Ibid., p. 34-35. Shweder (1994),
[170] Ekman, P. « An argument for basic emotions », *op. cit.*, p. 175-176.
[171] Voir Barrett, L. F. et al. « Emotional Expressions Reconsidered : Challenges to Inferring Emotion From Human Facial Movements », *Psychological Science in the Public Interest*, 20(1), 1-68, 2019, https://journals.sagepub.com/doi/10.1177/1529100619832930

sourcils et pincé les lèvres en une expression prototypique de colère alors qu'à d'autres moments nous souriions en répliquant sèchement[172]. À ces deux moments, nous étions en colère mais selon le contexte dans lequel nous étions placés, nous avons choisi d'exprimer (ouvertement) notre colère de différentes façons. Les recherches[173] en viennent donc à démontrer qu'un nombre diversifié d'expressions émotionnelles peuvent être exprimées dans un contexte émotionnel donné, ce qui signifie que chaque émotion (joie, colère, tristesse, etc.) peut être exprimée selon une variété d'expressions faciales. Une personne n'exprime donc pas nécessairement ces états émotionnels intérieurs selon les modèles prototypiques (sourire = joie, froncement de sourcils = colère, etc.). Au contraire, les expressions prototypiques des émotions ne correspondraient pas nécessairement aux expressions faciales émotionnelles que les personnes expriment lorsqu'elles manifestent des expressions faciales dans l'objectif d'exprimer leurs états émotionnels intérieures. Conséquemment, il est possible de dire qu'aucune émotion est directement liée à un état corporel spécifique ce qui signifie qu'aucune expression faciale précise ne peut être associée à une seule émotion.

3.1.1. Bilan pour les SRÉ

L'influence des structures de pouvoir sur les expressions faciales émotionnelles démontre la complexité de nos expressions faciales ainsi que leur double fonction qui est à la fois liée à notre ressenti émotionnel et notre volonté de communiquer avec autrui. Les normes sociales d'expressivité étant apprises dès un jeune âge, leur intériorisation se fait rapidement et devient dans la majorité des cas des actes expressifs inconscients. Les hommes n'ont pas nécessairement conscience de suivre des normes sociales lorsqu'ils expriment de la colère dans les moments où ils ressentent de la tristesse. De même, les femmes n'ont pas toujours conscience qu'elles ressentent une pression sociale à sourire davantage[174].

Cette influence des structures de pouvoir vient complexifier la reconnaissance des émotions réellement vécues par un individu et, en ce sens, rend la tâche des SRÉ plus difficile. Le phénomène d'intériorisation des règles sociales engendre une soumission inconsciente de la part des individus à

[172] Barrett, L. F. *op. cit*., p. 16.
[173] Hess, U. et Thibault, P. « Darwin and Emotion Expression », *American Psychological Association*, 64(2), p. 120 –128, 2009. p. 124.
[174] Tcherkassof, A. « Le sens dessus-dessous de l'expression faciale des émotions », *op. cit*., p. 34.

manifester des expressions faciales émotionnelles conformes au contexte dans lequel ils sont placés (ex. atténuer ou masquer ces émotions négatives lorsque l'on est un employé). Il semble qu'un SRÉ qui n'est pas en mesure de tenir compte du contexte ne pourrait espérer identifier correctement les émotions vécues par une personne. Comme nous l'avons vu dans cette section, plusieurs facteurs mènent un individu à masquer ses émotions intérieures sous des expressions faciales émotionnelles qui ne concordent pas avec celles-ci et ce, pour différentes raisons. Par ailleurs, il est aussi clair qu'un SRÉ ne peut prétendre être « universel » au sens où il serait en mesure de reconnaître les émotions de tous les individus, peu importe par exemple leur classement social vis-à-vis de leur interlocuteur et peu importe leur genre. Ces deux facteurs sont des éléments contextuels importants qui motivent les individus à manifester des expressions émotionnelles qui se conforment à la situation dans laquelle ils se trouvent plutôt qu'à exprimer leurs émotions intérieures réelles. Ainsi, la tentative d'accéder à la réalité émotionnelle intérieure à travers des mouvements mécaniques faciaux extérieurs est à questionner sérieusement. La signification de l'expression n'est donc finalement que dans de rares cas une extériorisation fidèle des émotions vécues intérieurement. Si l'on prend en compte les recherches sur les expressions faciales émotionnelles sexuées par exemple – et donc stéréotypées – un SRÉ qui aurait pour objectif de détecter les signes avant-coureurs d'une dépression au travail aurait vraisemblablement des défis supérieurs à relever avec les femmes puisque ces dernières ressentent une pression sociale à sourire davantage dans des contextes sociaux comme le lieu de travail, ce qui viendrait fausser les résultats[175]. Les tentatives d'identification des émotions intérieures à travers une analyse minutieuse des mouvements des muscles faciaux semblent donc faire face à des défis importants du côté du contexte.

3.2. Les groupes sociaux culturels

Si le débat concernant la cause de l'émotion, à savoir si elle est biologique ou sociale, n'est toujours pas résolu, un nombre grandissant d'études permet d'avancer que nos émotions sont influencées par notre environnement[176]. Celui-ci modifie la manière dont nous ressentons, exprimons et reconnaissons

[175] Tcherkassof, A. « Le sens dessus-dessous de l'expression faciale des émotions », *op. cit.*, p. 34.
[176] Des chercheurs tel que Dumas par exemple constatèrent que les aveugles de naissance ne sont pas en mesure de sourire. Il en déduit donc que les expressions faciales d'émotions sont socialement apprises et qu'elles sont donc susceptibles de différer d'une culture à l'autre dans des contextes où les normes sociales diffèrent. Voir Le Breton, D. «

des émotions[177]. Les expressions d'émotion sur nos visages sont affectées par notre culture[178], et celle-ci peut conditionner les membres d'une même société à exprimer des émotions particulières (ex. joie, tristesse, peur, etc.) dans un contexte donné. À travers une culture donnée, des règles sociales forgent et influencent nos comportements sociaux et mènent les individus d'une même société à exprimer une émotion précise (ex. peur) d'une manière précise (ex. écarquillement des yeux = peur). Ainsi, il est possible de constater que nous n'exprimons pas tous de la même façon une même émotion et parfois, la raison est le fait que nous provenons de différentes cultures qui nous ont enseigné des règles sociales différentes[179].

Ces différences culturelles ont des impacts importants sur les processus d'expression et de reconnaissance des émotions puisqu'elles créent un cadre de référence donné à un groupe de personnes spécifiques. Le cadre de référence d'une société donnée implique par exemple la langue parlée, les comportements acceptés et prohibés et les idées véhiculées[180]. Ainsi, dépendamment de la culture à laquelle il appartient, un individu particulier ne concevra pas, n'appréhendera pas et ne vivra pas nécessairement ses émotions de la même façon qu'un individu d'une autre culture pour une même situation donnée. Le vocabulaire introspectif (le nombre d'émotions qu'il connaît) de l'individu, qui est issu d'un cadre de référence donné, le mènera à classer les émotions d'une manière singulière[181,182]. Cette classification peut varier dans le nombre d'émotions reconnues mais aussi quant au champ d'action de l'émotion (c'est-à-dire, à ce qu'elle réfère, les moments où sa présence fait du sens et où elle n'en fait pas, etc.), selon la culture à laquelle l'individu appartient[183]. Ainsi, le nombre d'émotions ainsi que la signification à laquelle chaque émotion se rattache[184] varient selon les cultures. Le terme « colère » par exemple, est pensé et compris diffèremment en anglais (anger) et en polonais

Sociologie des émotions : Critique de la raison darwinienne », *Recherches sociologique*, 1, 1998, https://sharepoint.uclouvain.be/sites/rsa/Articles/1998-XXIX-1_05.pdf, p. 41.

[177] Barrett L. F. et al. *op cit*.

[178] Le Breton, D. *op. cit*., p. 41.

[179] Barrett, L. F. *op. cit.,* p. 13.

[180] Nous ne voulons pas déterminer d'ordre particulier entre la culture, la langue, le vocabulaire et les idées, nous tenons simplement à spécifier qu'ils sont liés les uns aux autres et conséquemment qu'ils impactent les uns et les autres.

[181] Wierzbicka, A. « Emotions across Languages and Cultures: Diversity and Universals », Studies in Emotion and Social Interaction, *Cambridge University Press*, Cambridge, 1999, https://www.cambridge.org/core/books/emotions-across-languages-and-cultures/7C03D03C6DF34ACBD7155B6555381715, p. 31.

[182] Ibid.

[183] Russell, J. A. « Culture and the categorization of Emotions », *Psychological Bulletin*, 110(3), 426-450, 1991, https://pubmed.ncbi.nlm.nih.gov/1758918/, p. 435.

[184] Schenk, F. « Les émotions de la raison », *Revue européenne des sciences sociales*, XLVII-144, 2009, 151-162, https://journals.openedition.org/ress/75#citedby. p. 152.

(złość ou gniew)[185]. Si à travers la langue anglaise, il est possible de ressentir de la colère face à une situation injuste (ex. : être en colère contre un médecin qui ne trouve pas de « solution miracle » pour la maladie incurable d'un proche) à travers le polonais, il ne fait aucun sens d'être en colère par rapport à ce type de situation. La colère n'est donc pas parmi les émotions applicables à cet exemple de situations en polonais[186] mais, pourtant, elle l'est en anglais. Un Polonais qui voit un Américain se fâcher dans ce type de situation (éprouver de la colère), pourrait rester surpris par ce comportement qui serait à ses yeux une expression d'émotion non attendue dans une telle situation alors qu'un autre Américain ne serait probablement pas surpris devant la manifestation de cette colère de la part de l'individu. Ainsi, si les Polonais ressentent de la colère et ont un terme clair pour définir la colère de manière similaire aux Américains (« złość » et/ou « gniew »), leurs concepts de la colère n'incluent pas une réaction de la sorte dans ce type de situation. Cela est dû à la différence culturelle qu'il y a entre la classification polonaise et américaine de cette émotion. Les différences existantes quant aux classifications d'un terme d'une culture à l'autre ne viennent pas remettre en doute l'affirmation que tous les individus ressentent des « émotions » au sens où nous ressentons peut-être tous quelque chose comme de la colère ou de l'excitation accompagnée de plaisir ou de déplaisir devant une situation. Seulement, nous ne donnons pas à cette sensation la même signification. Devant la situation où un médecin ne trouve pas de solution miracle pour la maladie incurable d'un proche, les Polonais interprètent les signaux de leur corps d'une autre façon que les Américains puisque leur classification est différente. Par ailleurs, il existe plusieurs mots pour décrire des émotions contextualisées qui n'ont pas leur référent en anglais[187]. Puisque chaque langue permet de déployer un certain cadre de référence qui n'est pas nécessairement réciproque ni identique d'une langue à l'autre et d'une culture à l'autre.

Si, à travers les cultures, certaines divergences existent quant aux émotions dites de base comme la colère ou la joie, il existe des cultures chez qui certaines de ces émotions n'existent pas du tout. C'est le cas des Utkuhikhalingmiut au Nord du Canada[188], qui ne sont jamais en colère[189]. La colère est pour eux un concept vide de sens. Les situations où, par exemple, un Américain et un Polonais

[185] Wierzbicka, A. *op. cit.,* p. 31.
[186] Ibid.
[187] Voir entre autres « Litost » en Tchèque, « Schadenfreud », « Angst » en Allemand dans Russell, J. A. « Culture and the categorization of Emotions », *op. cit*., p. 426.
[188] Briggs, J. L. *Never in Anger: Portrait of an Eskimo Family*, Harvard University Press, Harvard, 1971. p. 1.
[189] Ibid.

ressentiraient tous les deux de la colère seraient des situations où un Utkuhikhalingmiut ressentirait une émotion autre que celle de la colère.

Ces exemples démontrent donc que la colère, tout comme la joie, la tristesse, la peur, etc., sont des concepts culturels que nous appliquons à des changements d'états corporels (augmentation du rythme cardiaque, sudation, etc.)[190]. En tant que concept culturel, les émotions sont identifiées différemment et abordent des situations qui varient d'une culture à l'autre.

Ainsi, catégoriser et diviser les « émotions de base » en six catégories distinctes comprenant précisément la joie, la tristesse, la peur, le dégoût, la colère et la surprise ne permet pas d'englober les « émotions de base » de toutes les cultures. Pour certaines cultures, les émotions de base[191] incluront par exemple la honte, la fierté et la culpabilité[192]. Conséquemment, la théorie de base des émotions qui prend les six émotions citées plus haut comme émotions de base de *toutes* cultures n'est pas en mesure de représenter un cadre conceptuel (conceptual space) adéquat des émotions de base de plusieurs cultures, puisque les émotions de base données par la théorie ne sont pas celles reconnues pas leur culture.

L'illustration 3.1. permet de voir que la théorie des émotions de base « classique » est adaptée au cadre conceptuel des cultures caucasiennes occidentales mais l'est beaucoup moins en ce qui concerne par exemple le cadre conceptuel des cultures de l'Asie de l'Est.

---

[190] Barrett, L. F. *op. cit*. p. 28.
[191] Rachael E. J. et al. « Facial expressions of emotion are not culturally universal », *Proceedings of the National Academy of Sciences of the United States of America*, 2012, https://doi.org/10.1073/pnas.1200155109.
[192] Li, J., Wang, L. et Fischer, K. « The organisation of Chinese shame concepts? », *Cognition and Emotion,* 18(6), 767–797, 2004, https://www.tandfonline.com/doi/abs/10.1080/02699930341000202. ; Tracy, J.L. et Robins, R.W. « Show your pride: Evidence for a discrete emotion expression », *Psychological Science*, 15(3), 194–197, 2004 ; Bedford O. et Hwang K.-K. « Guilt and shame in Chinese culture: A cross-cultural framework from the perspective of morality and identity », *Journal for the Theory of Social Behaviour*, 33(2), 127–144, 2003.

Illustration 3.1. Analyse des différences dans la reconnaissance des émotions

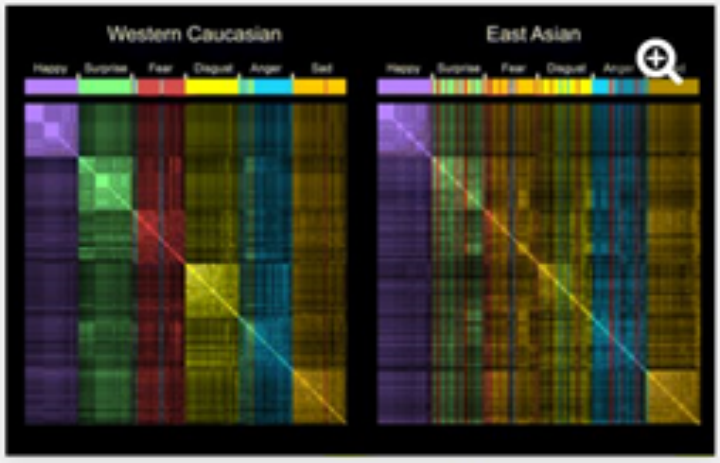


Fig. 2.

Download figure | Open in new tab | Download powerpoint

Cluster analysis and dissimilarity matrices of the Western Caucasian and East Asian models of facial expressions. In each panel, vertical color-coded bars show the *k* means (*k* = 6) cluster membership of each model. Each 41-dimensional model (*n* = 180 per culture) corresponds to the emotion category labeled *Above* (30 models per emotion). The underlying gray-scale dissimilarity matrices represent the Euclidean distances between each pair of models, used as inputs to *k*-means clustering. Note that, in the Western Caucasian group, the lighter squares along the diagonal indicate higher model similarity within each of the six emotion categories compared with the East Asian models. Correspondingly, *k*-means cluster analysis shows that the Western Caucasian models form six emotionally homogenous clusters (e.g., all 30 "happy" models belong to the same cluster, color-coded in purple). In contrast, the East Asian models show considerable model dissimilarity within each emotion category and overlap between categories, particularly for "surprise," "fear," "disgust," "anger," and "sad" (note the heterogeneous color coding of these models).

Source image : Rachael E. J. et al. *op. cit.*, p. 7243.

Il est en effet possible de voir que dans cette recherche, les individus des cultures occidentales caucasiennes ont facilement réussi à classer les images d'expressions faciales émotionnelles prototypiques (et plus précisément « prototypique » selon les normes sociales occidentales caucasiennes) c'est pourquoi les couleurs de chaque classe sont homogènes (ex. : mauve pour la joie, vert pour la surprise, rouge pour la peur, etc.). Dans le tableau des cultures de l'Asie de l'Est, la joie est la seule émotion facilement reconnue alors que les cinq autres émotions sont considérablement mélangées entre elles comme le démontrent les couleurs hétérogènes de chaque classe. Ainsi, il est possible de constater que plusieurs émotions sont exprimées différemment selon la culture à laquelle nous appartenons et ne peuvent pas, en ce sens, être reliées à une expression faciale précise et

universelle[193,194,195,196]. Le classement émotionnel varie considérablement entre cultures et ne permet pas de dire que l'expression faciale ou la reconnaissance des émotions sont universelles et ce, même en ce qui concerne les émotions dites « fondamentales ». En ce qui concerne plus particulièrement les cultures de l'Asie de l'Est, il est possible de noter que dans les émotions de base de cette culture, les individus incluent instinctivement et cherchent à retrouver des émotions comme la fierté, la honte et la culpabilité, alors que ces émotions ne sont pas reconnues en tant qu'émotions de base par les chercheurs occidentaux-caucasiens[197]. C'est donc pour cette raison – parce que le cadre de référence ainsi que la classification des termes et concepts changent d'une culture à l'autre – que les individus issus des cultures de l'Asie de l'Est ont eu tant de difficulté à associer les expressions faciales prototypiques à l'une des six émotions suggérées par le modèle occidental-caucasien.

Une technologie de reconnaissance des émotions qui se fonde sur la théorie de base des émotions occidentales-caucasiennes par exemple ne sera pas en mesure de tenir compte de la diversité culturelle existante et ne pourra reconnaître efficacement que les expressions d'émotions prototypiques caucasiennes et occidentales[198]. Autrement dit, un modèle unique qui se fonde sur une théorie des émotions située est un modèle non-universel. Toutefois, en prétendant être un modèle universel, et donc en prétendant être en mesure de reconnaître les expressions d'émotions de toutes les cultures sur la base d'un modèle situé culturellement, certains systèmes de reconnaissance des émotions peuvent causer des torts importants à l'égard des individus appartenant à des cultures qui se distinguent de celle sur lequel le modèle est basé.

Tenter d'interpréter avec une « grille de lecture » issue de sa propre culture les actes d'autrui, des actes qui ont été réfléchis à partir d'une autre grille, engendre le risque de prêter des significations bien

---

[193] Sarwari, K. « You Think You Can Read Facial Expressions? You're Wrong », *Northeasteran University Media*, 2019, https://news.northeastern.edu/2019/07/19/northeastern-university-professor-says-we-cant-gauge-emotions-from-facial-expressions-alone/ (Page consultée le 4 avril 2021)
[194] Barrett, L. F. *op. cit.,* p. 11.
[195] Russell, J.A. « Is There Universal Recognition of Emotion from Facial Expression? A Review of the Cross-Cultural Studies », *Psychological Bulletin*, 115(1), 102-141, 1994, https://doi.org/10.1037/0033-2909.115.1.102
[196] Chen, C. et al. « Distinct Facial Expressions Represent Pain and Pleasure Across Cultures », *Proceedings of the National Academy of Sciences of the United States of America*, 115(43), 2018, E10013–E10021, https://www.pnas.org/content/115/43/E10013
[197] Rachael E. J. et al. *op. cit*.
[198] Barrett, L. F. et al., *op. cit*., p. 51.

différentes des intentions initiales de leurs auteurs[199]. En effet, les règles d'expressivité des émotions (display rules) sont différentes d'une culture à l'autre[200]. Selon la psychologue Anna Tcherkassof :

> « ces règles spécifiques d'expressivité sont des normes sociales caractéristiques des différentes cultures imposées en matière d'expressivité émotionnelle. Ce sont des prescriptions culturelles dictant comment une personne d'une culture donnée doit exprimer ses émotions, ces normes sociales étant acquises très tôt[201] »

Les émotions vécues ne sont pas les seules à diverger d'une culture à l'autre, les règles sociales aussi. Conséquemment, la décision d'atténuer, de neutraliser ou d'accentuer une expression faciale d'émotion sera modulé selon les règles sociales de notre culture d'appartenance. Par exemple, certains chercheurs ont constaté que certaines émotions – souvent négatives – étaient plus fréquemment « masquées » par un sourire dans plusieurs cultures asiatiques comme le Japon que dans les cultures occidentales caucasiennes[202]. Cela signifie qu'il est possible que deux individus issus de deux cultures très différentes, par exemple un Américain et un Japonais, pleurent un proche aimé seul dans leur salon en esquissant une expression faciale émotionnelle similaire de tristesse. Toutefois, à l'extérieur de leur maison, où lorsqu'ils ne sont pas seuls, les deux individus n'exprimeront possiblement pas leur tristesse de la même manière. Il est possible que l'Américain pleure devant des personnes qu'il ne connaît pas, parce qu'il est triste et qu'il ressent un « besoin » de pleurer alors que de son côté le Japonais affiche un grand sourire parce que la personne devant lui est un inconnu[203]. Le sourire du Japonais est construit par une convention sociale : « un Japonais annonçant la mort d'un proche à un tiers garde le visage souriant ; il marque ainsi le respect de l'intimité de l'autre, le refus ritualisé de l'impliquer dans le partage d'une douleur qui ne le concerne pas[204] ». Dans un contexte où l'individu est non-soumis aux règles d'expressivité sociale de sa culture – lorsqu'il est seul chez lui – il est possible que l'individu exprime sur son visage une expression faciale émotionnelle non altérée par les normes sociales. Cependant, dans certains contextes, les expressions faciales émotionnelles sont provoquées et forgées par les normes sociales de notre culture, ce qui fait en sorte que les individus

---

[199] Chevrier, S. « Le management des équipes interculturelles », International Management, 8(3), 2004, http://www.managementinternational.ca/catalog/revue/achat-articles-et-de-numeros/themes/gestion-interculturelle/le-management-des-equipes-interculturelles-texte-en-francais.html.
[200] Tcherkassof, A. *op. cit.*, p. 14.
[201] Ibid.
[202] Le Breton, D. *op. cit.*, p. 41.
[203] Ibid.
[204] Ibid.

de certaines cultures réagiront d'une certaine manière dans une situation donnée. Lorsque cela arrive, les expressions faciales ne sont pas toujours liées à une émotion précise[205] ni à l'émotion attendue (ex. sourire = joie).

3.2.1. Bilan pour les SRÉ

Un SRÉ qui prétend être culturellement universel alors qu'il ne l'est pas est un SRÉ qui risque de causer des torts disproportionnés vis-à-vis des individus qui n'appartiennent pas à la culture à partir de laquelle le module a été entrainé. Le risque d'arriver à des résultats erronés ou désavantageux pour des groupes précis de la population est un risque qui avait déjà été soulevé dans les débuts des systèmes de reconnaissance faciale, lorsque plusieurs modèles n'avaient pas été en mesure de démontrer des performances égales entre tous les groupes sociaux. Plusieurs systèmes éprouvaient, par exemple, des difficultés à performer avec les visages d'individus non-caucasiens. Ce même risque est aussi présent dans les systèmes de reconnaissance des émotions. En ce qui concerne les SRÉ, les lacunes importantes engendrées par le peu de prise en considération des différences culturelles représentent un facteur qui peut être la cause d'un traitement erroné. Utilisé dans des contextes d'embauche par exemple, il pourrait mener à la disqualification de candidats présentant les compétences requises[206].

Les contextes sociaux peuvent influencer les expressions faciales émotionnelles selon notre position sociale vis-à-vis de notre interlocuteur, selon notre genre mais aussi dépendamment de notre culture d'appartenance. Nous ne réagirons pas tous de la même façon à un contexte social donné et notre culture peut être la source d'une réaction précise. Ces différences ont une influence importante sur la manière dont nous exprimons nos émotions et sur la manière dont les autres reconnaîtront (ou non) nos émotions. La reconnaissance des émotions par l'IA peut être biaisée par ces différences si l'IA est basée sur un modèle qui ne tient pas compte de celles-ci. Conséquemment, les SRÉ qui sont basés sur un système situé culturellement ne sont pas adéquats pour reconnaître les expressions d'émotions des autres cultures.

[205] Barrett, L. F. *op. cit*., p. 163.
[206] Ibid., p. 192.

La Chine est l'un des premiers pays à avoir grandement investit dans des recherches sur les systèmes de reconnaissance des émotions[207]. Les systèmes présentement utilisés par la Chine se sont révélés être « volontairement » culturellement limités puisqu'ils sont basés sur les expressions prototypiques chinoises[208], dans l'objectif d'arriver à des résultats plus précis. Le choix de la Chine de se limiter aux expressions faciales chinoises est une autre preuve de l'existence de différences culturelles et de leurs impacts sur les résultats des SRÉ.

Des torts importants peuvent découler d'une reconnaissance des émotions erronées[209], tel que des arrestations, des expulsions, des rejets à l'embauche et même des refus de libération conditionnelle. Aux États-Unis, le programme SPOT (Screening of Passengers by Observation Techniques) fut utilisé un certain temps pour surveiller et analyser les expressions faciales des voyageurs dans les années suivant les attaques du 11 septembre 2001 dans l'objectif de prévenir des attaques terroristes[210]. Le programme SPOT était composé de 94 critères qui permettaient de détecter les signes de stress, d'anxiété et de peur[211]. Toutes les personnes identifiées comme stressées, inconfortables ou anxieuses étaient mises sous interrogatoires. Ce programme de 900 millions de dollars fut finalement réduit à un programme de profilage racial puisque les groupes sociaux (exemple communautés afro-américaines et arabes) vivant d'avantage de stress et d'anxiété lors d'interactions avec la police ou d'autres formes d'autorités compétentes (ex. sécurité, douaniers, etc.) étaient détectés comme menaces potentielles[212]. Conséquemment le programme participait à renforcer un cercle vicieux entre les groupes historiquement opprimés et discriminés (communauté afro-américaine, communautés arabes) et les stéréotypes des autorités compétentes. D'ailleurs, le programme a fait l'objet de critiques sévères de la part du *U.S. Government Accountability Office* et des Associations juridiques pour la défense des droits et libertés aux États-Unis, qui reposèrent la majorité de leurs critiques sur l'absence de preuves et de méthodes scientifiques pour appuyer le programme[213]. De tels programmes causent en effet des torts importants à ces groupes ciblés par les effets discriminatoires.

---

[207] Article 19. « Emotional Entanglement: China's emotion recognition market and its implications for human rights », *op. cit*., p. 13-14.
[208] Ibid., p. 41.
[209] Crawford, K. et al. « AI Now 2019 Report », *op. cit*., p. 6.
[210] Heaven, D. « Why Faces Don't Always Tell the Truth about Feelings », *Nature*, 2020, https://www.nature.com/articles/d41586-020-00507-5.
[211] Crawford, K. *The Atlas of AI : Power, Politics and the planetary cost of Artificial Intelligence*, p. 170-171.
[212] Ibid.
[213] Ibid.

En 2018, la chercheuse Lauren Rhue avait démontré que deux programmes de reconnaissance des émotions, soit le programme Face++ et le programme Microsoft AI présentaient des biais raciaux importants[214]. Sur un ensemble de données composé de 400 joueurs de la NBA, les deux programmes attribuaient systématiquement aux joueurs à la peau noire des scores émotionnels plus négatifs en moyenne que les joueurs à la peau blanche. Le programme Face++ identifiait les joueurs à la peau noire comme étant en moyenne davantage « en colère » (angrier) que les joueurs à la peau blanche et le programme Microsoft AI identifiait les joueurs à la peau noire comme étant en moyenne plus « méprisants » (contemptuous) que les joueurs à la peau blanche. Cette étude nous permet d'illustrer la gravité des erreurs présentes dans l'utilisation de plusieurs systèmes de reconnaissance des émotions. Le rapport de 2019 du AI Now Institutes, affilié à l'Université de New York, confirmait qu'aux États-Unis, les systèmes de reconnaissance des émotions étaient déjà utilisés par les services de police, dans les aéroports, les banques, les restaurants, les magasins, les casinos, les lieux publics, etc[215]. Il est indéniable que si ces systèmes s'avéraient biaisés de manière similaire à la manière dont les programmes Face++ et AI Microsoft l'étaient, les impacts sur la vie des personnes désavantagées seraient importants. L'utilisation des SRÉ dans les aéroports, les hôpitaux, les banques et par la police a en général pour principal objectif d'identifier des « personnes agressives » avant que ces dernières ne commettent un acte agressif[216,217]. Si ces systèmes attribuent par exemple systématiquement des scores émotionnels plus négatifs aux individus à la peau noire, il est évident que les risques qu'ils soient identifiés comme « agressifs » soient plus élevés que les risques qu'une personne à la peau blanche soit identifiée comme « agressive ». Une identification de la sorte peut vraisemblablement mener à des arrestations biaisées, à l'utilisation d'une force excessive basée sur un résultat erroné, à une expulsion d'un lieu sous prétexte que l'individu émet les signes précurseurs d'une agression, etc.[218,219] Dans d'autres cas, l'identification erronée de scores émotionnels plus négatifs peut biaiser négativement les entrevues d'embauche, mener à un renvoi[220], se faire refuser une mise en libération

[214] Rhue, L. « Racial Influence on Automated Perceptions of Emotions », *SSRN*, 2018, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3281765, p. 3.
[215] Crawford, K. et al. « AI Now 2019 Report », *op. cit.,* p. 6.
[216] Ibid.
[217] Gillum, J. et Kao, J. « Aggression Detectors : The Unproven, Invasive Surveillance Technology Scholls Are Using to Monitor Students », *ProPublica*, 2019, https://features.propublica.org/aggression-detector/the-unproven-invasive-surveillance-technology-schools-are-using-to-monitor-students/
[218] Ibid.
[219] Thomas, D. « The Cameras that Know if You're Happy – or a Threat », *BBC News*, 2018, https://www.bbc.com/news/business-44799239
[220] Crawford, K. et al. « AI Now 2019 Report », *op. cit.*, p. 6.

conditionnelle, etc[221]. Aux États-Unis, par exemple, les émotions des accusés sont des informations précieuses et une attention particulière est portée par les juges et le jury à « lire les émotions d'un accusé[222] ». Les acteurs juridiques tels que les jurys et les juges s'appuient régulièrement sur les mouvements du visage pour déterminer la culpabilité et les remords d'un accusé[223]. Les accusés perçus comme indignes de confiance sont condamnés à des peines plus sévères qu'ils ne l'auraient été autrement[224]. L'utilisation d'un système de reconnaissance des émotions qui attribue des émotions en moyenne plus négatives aux personnes à la peau noire peut mener à des inférences erronées sur l'état émotionnel des accusés et cette erreur peut leur coûter la garde de leurs enfants, leur liberté, voire leur vie[225].

Ces exemples appuient l'idée qu'un SRÉ basé sur une théorie occidentale-caucasienne de l'expression de l'émotion n'est pas un SRÉ adéquat à la reconnaissance d'émotion d'autres cultures. Évidemment, les systèmes caucasiens-occidentaux ne sont pas les seuls à ne pas être « culturellement » universels, comme nous l'avons vu avec les SRÉ utilisés en Chine.

Ces lacunes des systèmes culturellement limités viennent remettre en cause l'efficacité ainsi que la fiabilité de ces systèmes qui se prétendaient « universels »[226]. Enfin, volontairement choisir un système basé sur un modèle culturellement limité des émotions pour, par exemple, aider les juges et les jurys d'un pays à fonder leurs jugements sur les accusés n'est pas acceptable dans nos sociétés contemporaines où la diversité culturelle est la norme. Plusieurs chercheurs et groupes de chercheurs affirment déjà leurs désaccords quant à l'utilisation des systèmes de reconnaissance des émotions, notamment dans le cadre de situations « importantes dans la vie d'un individu », ce qui inclut notamment l'éducation, l'accès aux soins de santé, les entrevues d'embauche et les décisions juridiques[227,228].

---

[221] Ibid.
[222] Barrett, L. F. et al. *op. cit.*, p. 3.
[223] Ibid.
[224] Ibid.
[225] Ibid.
[226] Korte, A. « Facial Recognition Technology Cannot Read Emotions, Scientists Say », *American Association for the Advancement of Science*, 2020, https://www.aaas.org/news/facial-recognition-technology-cannot-read-emotions-scientists-say (Page consultée le 4 mai 2021)
[227] Crawford, K. et al. « AI Now 2019 Report », *op. cit.*, p. 3.
[228] Article 19. « Emotional Entanglement: China's emotion recognition market and its implications for human rights », *Article 19*, 2021, https://www.article19.org/wp-content/uploads/2021/01/ER-Tech-China-Report.pdf, p. 3.

3.3. Conclusion

Dans ce chapitre, nous avons réitéré les fonctions biologiques et sociales des expressions faciales ainsi que notre incapacité collective à différencier ces trois fonctions par la simple observation des expressions faciales des individus. Nous avons identifié deux principaux facteurs importants venant influencer la manifestation des expressions faciales ainsi que notre compréhension de celle-ci : le contexte dans lequel l'individu manifeste une expression faciale émotionnelle et la culture à laquelle il appartient. Le contexte peut influencer de diverses manière nos expressions faciales et de nombreuses études ont permis d'établir des preuves convaincantes quant à la modulation de nos expressions en fonction de facteurs précis comme notre position sociale vis-à-vis de notre interlocuteur ou notre genre. Le contexte est si important qu'il incite un même individu à exprimer une même émotion différemment selon la situation dans laquelle il se trouve. De son côté, la culture a une influence non négligeable sur le nombre d'émotions que nous avons à notre disposition ainsi que sur le champ d'action de nos émotions. Ce nombre varie largement selon les cultures et vient remettre en question l'universalité de la première « couche », à savoir, l'idée selon laquelle nous aurions un nombre prédéfini d'émotions innées. Le champ d'action de nos émotions varie lui aussi selon les cultures et peut nous inciter à ressentir certaines émotions dans certains contextes. Il peut aussi être influencé par les normes sociales de notre culture et en ce cas nous inciter à atténuer, accentuer ou neutraliser l'expression de certaines émotions dans certains contextes. Cette variance vient remettre en question la prétention à l'universalité de la deuxième « couche », soit l'idée que les êtres humains expriment et reconnaissent certaines émotions de base chez tous les êtres humains. Tel que nous l'avons vu plus haut, l'existence de limites contextuelles et culturelles vient miner les prétentions à l'universalité des SRÉ et discrédite leur efficacité ainsi que leur fiabilité. Conséquemment, la remise en question de ces deux « couches » d'universalité est au cœur de notre projet qui vise à questionner la participation des SRÉ à la justice sociale.

Chapitre 4 Sur la participation des SRÉ à la justice sociale

Dans ce quatrième chapitre, nous procèderons à l'analyse de l'utilisation des SRÉ dans l'objectif de déterminer si cette utilisation est en cohérence avec la justice sociale. Pour ce faire, dans un premier temps, nous utiliserons trois critères spécifiquement identifiés pour être en mesure d'analyser les SRÉ soit la non-discrimination, l'efficacité et la fiabilité. Ces trois critères nous permettront de déterminer si les SRÉ répondent aux exigences de la justice sociale. À la suite de cette analyse, nous tenterons de faire ressortir, dans un deuxième temps, plus explicitement les risques et les bénéfices de l'utilisation des SRÉ afin de déterminer si ceux-ci sont en cohérence avec la justice sociale. Enfin, dans un troisième temps, nous expliquerons à l'aide de trois types de SRÉ que certains d'entre eux ne sont peut-être pas incompatibles avec la notion de justice sociale, même si les enjeux éthiques que leur utilisation soulève sont considérables.

4.1. L'utilisation des SRÉ répond-t-elle aux exigences de la justice sociale ?

La société actuelle – et les autorités compétentes qui la dirigent – déclare avoir la volonté de protéger et respecter les droits et libertés fondamentaux des individus et ce, à travers les changements sociaux et collectifs que représente la révolution numérique. Toutefois, l'augmentation de l'utilisation des SRÉ dans de nombreux pays[229] (Chine[230], États-Unis[231], les États membres de l'Union Européenne[232]) et par de nombreuses compagnies privées[233] à des fins de sécurité, de marketing, de politique,

[229] Article 19. « Emotional Entanglement: China's emotion recognition market and its implications for human rights », *Article 19*, 2021, https://www.article19.org/wp-content/uploads/2021/01/ER-Tech-China-Report.pdf.
[230] Mou, X. « Artificial Intelligence: Investment Trends and Selected Industry Uses », *International Finance Corporation*, 2019 https://www.ifc.org/wps/wcm/connect/7898d957-69b5-4727-9226-277e8ae28711/EM Compass-Note-71-AI-Investment Trends.pdf?MOD=AJPERES&CVID=mR5Jvd6.
[231] NITRD. « Supplement to the President's FY2020 Budget », *The Networking and Information Technology Research and Development Program*, 2019, https://www.nitrd.gov/pubs/FY2020-NITRD-Supplement.pdf#page=17. Et Cornillie, C. « Finding Artificial Intelligence Money in the Fiscal 2020 Budget », *Bloomberg government*, 2019, https://about.bgov.com/news/finding-artificial-intelligence-money-fiscal-2020-budget/ (Page consultée le 19 mars 2021)
[232] Commission européenne. « Façonner l'avenir numérique de l'Europe: la Commission présente des stratégies en matière de données et d'intelligence artificielle », *Commission Européenne*, 2020, https://ec.europa.eu/commission/presscorner/detail/fr/ip_20_273 (Page consultée le 15 mars 2021)
[233] Lewis, T. « AI can read your emotions. Should it? », *The Gardian*, 2019, https://www.theguardian.com/technology/2019/aug/17/emotion-ai-artificial-intelligence-mood-realeyes-amazon-facebook-emotient (Page consultée le 17 mars 2021)

d'embauche, d'éducation, de *care*, etc.[234], semble créer une tension avec cette volonté de protéger et respecter les droits et libertés fondamentaux, droits et libertés qui, comme nous l'avons vu, constituent les piliers de la justice sociale. Par ailleurs, les lacunes importantes quant à l'iniquité des performances des SRÉ vis-à-vis de certains groupes de la société vient aussi remettre en question leur capacité à octroyer un traitement juste, c'est-à-dire un traitement constant et impartial.

4.1.1. Les enjeux éthiques des SRÉ

Dans ce contexte qui révèle des tensions internes, nous croyons qu'il est nécessaire d'analyser l'utilisation des SRÉ afin de déterminer si cette utilisation est en cohérence avec les principales exigences de la justice sociale – exigences que nous avons illustrées à l'aide de nos deux constances ; (1) celle que chaque individu possède un certain nombre de droits et libertés et (2) qu'un traitement juste est un traitement impartial et constant.

L'objectif commun qui sous-tend les diverses utilisations des SRÉ est celui d'accéder à des données sur les états intérieurs qui n'auraient pas été autrement accessibles. Cet objectif peut par la suite être utilisé dans divers contextes tel que le marketing, la sécurité, la politique, l'éducation, etc. Notre analyse éthique se divisera en deux grandes étapes, dans un premier temps, nous évaluerons si l'utilisation des SRÉ est en mesure de se faire dans le respect de trois principes – qui endossent ici le rôle de critères de satisfaction à la justice sociale – et dans un deuxième temps, nous soupèserons les risques et les bénéfices qu'apportent l'utilisation des SRÉ. Notre analyse se basera sur trois critères dont le premier sera un principe éthique (la non-discrimination) et les deux suivants (l'efficacité et la fiabilité) des principes dont la valeur éthique est dérivée de leur pertinence quant à une analyse de systèmes de reconnaissance des émotions ainsi que de leurs liens étroits avec le principe éthique précédent. Ces trois critères représenteront les deux exigences de la justice sociale en étant spécifiquement liés à l'utilisation des SRÉ.

[234] Crawford, K. *The Atlas of AI : Power, Politics and the planetary cost of Artificial Intelligence*, Yale University Press, New Haven and London, 2021, p. 152.

4.1.1.1. La non-discrimination

Prétendre octroyer le même traitement à tous et chacun alors que ce n'est pas le cas est susceptible de causer des torts importants aux groupes sociaux désavantagés par cette utilisation. Ces torts incluent entre autres choses des arrestations, des expulsions de lieux, des rejets à l'embauche, des renvois et des refus de libération conditionnelle, sans compter l'augmentation du stress et de l'anxiété que des expériences négatives répétées d'un traitement partial peut engendrer chez l'individu[235]. Un traitement injuste peut aussi aller jusqu'à brimer un individu de ses droits et libertés fondamentaux. À travers la valeur de non-discrimination, nous soulèverons les risques et les menaces que représente l'utilisation des SRÉ pour divers groupes sociaux.

4.1.1.1.1. La non-discrimination et le statut social

Le statut social est régi par des structures de pouvoir qui influencent les comportements et les expressions faciales des individus dans un contexte donné. Les interactions entre les individus s'effectuent toujours sous l'ombre de cette structure sociale de pouvoir qui vient classer les individus selon une hiérarchie sociale déterminée. Ces structures de pouvoir forment ainsi des normes sociales qui influencent le comportement des individus. L'utilisation des SRÉ vient poser un risque accru de traitement inéquitable vis-à-vis des individus ayant différents statuts sociaux.

Un SRÉ qui prétend s'élever à un universalisme de la reconnaissance des émotions ne tient pas compte de l'influence des structures de pouvoir sur l'expression faciale des individus ce qui peut engendrer, selon le contexte, une discrimination à l'encontre des individus de certaines classes. Les structures de pouvoir modulent les comportements que les individus auront vis-à-vis des autres individus. Par exemple, un employé n'adoptera pas le même comportement avec un autre employé versus avec son employeur. Avec un autre employé, l'individu risque de s'exprimer plus librement, de laisser libre court à ses expressions faciales émotionnelles, qu'elles soient positives ou négatives, tandis qu'avec son employeur, l'individu aura tendance à atténuer ses expressions faciales émotionnelles puisqu'il est devant une autorité compétente et que les normes sociales dictent à

[235] Holmes, S. C., Facemire, V. C. et DaFonseca, A. M. « Expanding criterion a for posttraumatic stress disorder: Considering the deleterious impact of oppression ». *Traumatology, 22*(4), 314–321, https://doi.org/10.1037/trm0000104

l'employé d'atténuer ses expressions faciales émotionnelles en plus de les rendre le plus positif possible[236]. Cette modulation des comportements incite les individus à modifier leurs expressions faciales, ce qui peut engendrer des biais dans l'analyse des SRÉ et causer de la discrimination. Un employé qui est en fréquent contact avec son employeur pourrait obtenir des scores émotionnels plus positifs en moyenne qu'un employé qui est très peu en contact avec son employeur. Toutefois, atténuer les expressions faciales émotionnelles négatives devant une autorité compétente n'implique pas que l'individu ressent des émotions positives. Ainsi, un SRÉ qui aurait pour objectif de détecter les signes précurseurs d'une dépression pourrait voir ses résultats biaisés par l'implication des normes sociales issues des relations de pouvoir entre un employé et son employeur et ne pas détecter que l'employé vit une dépression grave puisqu'il est incité à afficher des émotions positives et à atténuer ses émotions négatives devant son employeur. Conséquemment, le statut social peut être la cause de biais discriminatoires qui viennent fausser les résultats des SRÉ et désavantager certains individus en raison de leur position sociale.

4.1.1.1.2. La non-discrimination et les genres

La Loi sur le ministère des Femmes et de l'Égalité tente de renforcer l'égalité des genres en stipulant que « tous ont droit à la même protection et au même bénéfice de la loi, indépendamment de toute discrimination fondée sur le genre[237] ». Toutefois, la discrimination des genres semble être un risque potentiel dû à l'utilisation des systèmes de reconnaissance des émotions. Dû à l'endossement de certaines normes sociales dès un jeune âge, les individus d'une société donnée adoptent des comportements dans des contextes précis qui sont susceptibles d'influencer leurs expressions faciales[238], de flouer les SRÉ et conséquemment, d'être sujets de discrimination vis-à-vis de l'analyse des SRÉ. En Occident par exemple, une pression sociale est mise sur les femmes pour les inciter à sourire davantage[239] alors que les hommes sont incités à exprimer plus librement leur colère ou leur mécontentement[240]. Or, comme nous l'avons vu, dans le cas où un SRÉ qui serait utilisé dans l'objectif

[236] Tcherkassof, A. « Le sens dessus dessous des expressions faciales d'émotions : vers un nouveau tournant paradigmatique », *op cit.,* p. 34.
[237] Gouvernement du Canada. « Loi sur le ministère des Femmes et de l'Égalité des genres », Gouvernement du Canada, https://laws-lois.justice.gc.ca/fra/lois/W-11.3/page-1.html (Page consultée le 4 juin 2021)
[238] Tcherkassof, A. *op. cit*., p. 34.
[239] Ibid., p. 33.
[240] Ibid., p. 34.

de détecter les signes précurseurs d'une dépression au travail, il semble que le système aurait des défis supérieurs à relever avec les femmes puisque ces dernières ressentent une pression sociale à sourire davantage dans des contextes sociaux comme le lieu de travail, ce qui viendrait fausser les résultats[241].

Ainsi, il est possible de constater que les genres, combinés aux normes sociales, sont un schéma complexe – basé généralement sur la binarité sexuelle – de règles, de normes, de comportements et de systèmes d'influences qui sont plus ou moins suivis par les individus eux-mêmes et qui peuvent aussi parfois être supplantés par d'autres normes sociales qui peuvent provenir de leur noyau familial, de leur personnalité ou encore de leurs cercles d'amis. Les SRÉ qui ne tiennent pas compte de cette différence contextuelle de genres ne seront pas en mesure d'identifier les signes précurseurs de la dépression chez les hommes et les femmes avec un pourcentage de réussite équivalent.

De plus, l'enjeu autour des « émotions sexuées » pose de sérieux problèmes à la reconnaissance émotionnelle « universelle » puisque le phénomène d'émotions sexuées exige de tenir compte du particulier et du contextuel. Si les femmes et les hommes ont « tendance » à exprimer certaines émotions et à en refouler d'autres, cette influence des normes sociales sur les expressions faciales diffère grandement selon le contexte, le milieu, les individus présents, le vécu de chaque individu présent, l'état émotionnel dans lequel tous et chacun sont lors de l'événement X, etc.

Pour le moment, l'utilisation des SRÉ ne semble pas en mesure d'octroyer un traitement égal des genres parce qu'il n'est pas en mesure de tenir compte des normes sociales qui s'appliquent différemment aux individus d'une même situation selon leur genre. Par ailleurs, quoique les normes sociales puissent, à prime abord, sembler être des guides pour la programmation des SRÉ, c'est-à-dire qu'ils pourraient fournir les codes sociaux à travers lesquels la programmation des SRÉ pourrait être bonifiée, il est important de souligner que ces normes sont constamment déclassées par le particulier et le contingent. En effet, même si les individus ont « tendance » à suivre ces normes sociales selon leur genre, ils n'ont pas conscience qu'ils suivent ces normes et, conséquemment, n'ont pas toujours conscience qu'ils dérogent de la norme aussi. Chaque individu suit certaines normes plus

[241] Ibid.

que d'autres et, conséquemment, baser la programmation sur ces normes sociales causerait certainement de la discrimination en plus de ne pas atteindre la fin visée.

Ainsi, le diagnostic d'un SRÉ peut désavantager certains individus dans certains contextes selon leur genre. Le genre est donc pour le moment un facteur de discrimination lors de l'utilisation des SRÉ. Il peut entrainer une discrimination négative aussi bien chez les hommes que chez les femmes selon le contexte dans lequel il opère.

4.1.1.1.3. La non-discrimination et la diversité culturelle

Dans la section sur le principe d'inclusion de la diversité de la Déclaration de Montréal, il est possible d'y lire : « le développement et le déploiement des SIA doivent prendre en considération les multiples expressions des diversités sociales et culturelles, et cela dès la conception des algorithmes[242] ». Le respect de la valeur de non-discrimination, au cœur de la Déclaration de Montréal pour un développement responsable, est une valeur qui est aussi partagée par la société québécoise qui interdit la discrimination par la Charte des droits et libertés du Québec[243].

Or l'utilisation des SRÉ ne se fait généralement pas dans le respect du principe de non-discrimination. En effet, les SRÉ qui prétendent à une universalité dans la reconnaissance des émotions nient l'existence des différences culturelles ainsi que les impacts des différences culturelles sur le pourcentage d'efficacité de la reconnaissance des émotions d'individus issus de cultures se situant à l'extérieur des limites culturelles du modèle. Un modèle basé sur les émotions et les expressions émotionnelles occidentales-caucasiennes n'octroiera pas un traitement égal entre les individus faisant partie de la culture sur laquelle le modèle a été basé et les individus se situant à l'extérieur de cette culture. Ainsi, les personnes issues de cultures se situant à l'extérieur des limites du SRÉ recevront un traitement différent, qui viendra biaiser l'analyse ainsi que les diagnostiques possibles. Dans ce projet de recherche, notamment au chapitre 2, nous avons vu que les émotions ainsi que les expressions

[242] Déclaration de Montréal. « La Déclaration de Montréal pour un développement responsable de l'intelligence artificielle ». Principe d'inclusion de la diversité, 2018, https://www.declarationmontreal-iaresponsable.com/la-declaration.
[243] Commission des droits de la personne et des droits de la jeunesse. « La discrimination », Commission des droits de la personne et des droits de la jeunesse, Québec, https://www.cdpdj.qc.ca/fr/vos-obligations/ce-qui-est-interdit/la-discrimination (Page consultée le 3 juin 2021)

émotionnelles sont influencées par notre culture d'appartenance et varient donc selon elle. Conséquemment, les SRÉ basés sur un modèle occidental-caucasien, par exemple, ne donneront pas à un homme blanc les mêmes chances qu'à un homme noir, et il en sera de même avec tous les individus se situant à l'extérieur des limites culturelles du modèle. De même, en Chine, un homme occidental-caucasien n'aura pas les mêmes chances qu'un homme chinois Han. Les compagnies de SRÉ qui prétendent être en mesure d'analyser les émotions de tous les groupes culturels nient la discrimination systématique de leur système de reconnaissance des émotions, ce qui représente une menace pour la justice sociale. Nier que l'analyse sera discriminatoire envers certains groupes culturels est une menace pour la justice sociale puisqu'elle vient s'opposer à l'une des deux exigences fondamentales : celle qui exige un traitement juste, c'est-à-dire un traitement qui est impartial et constant. Notre analyse du traitement des SRÉ démontre que le traitement n'est pas, dans bien des cas, impartial. Ainsi, l'utilisation des SRÉ ne satisfait pas les exigences du critère de non-discrimination sur plusieurs niveaux. Au niveau technique, les SRÉ ont des limites qui les empêchent pour le moment d'avoir des modèles qui soient en mesure de considérer tous les groupes culturels sans discrimination. Au niveau moral, nier que l'utilisation des SRÉ demeure pour le moment une utilisation limitée à certains groupes culturels est un manquement à la responsabilité morale de la compagnie qui engendre des risques éthiques importants pour les individus qui subissent la discrimination négative.

Par ailleurs, la tension présente entre l'utilisation des SRÉ et la valeur de non-discrimination nécessite un questionnement plus en profondeur concernant plus particulièrement la légitimité d'un modèle culturellement situé dans nos sociétés pluriculturelles : comment accepter les limitations culturelles d'un système conçu pour analyser les émotions des individus d'une société pluriculturelle ? Comme nous l'avons vu, le critère de non-discrimination ne peut vraisemblablement pas être respecté lorsqu'un SRÉ est utilisé sur un groupe culturel qui se situe à l'extérieur des limites de ce SRÉ. Or la limite culturelle du SRÉ semble nécessairement impliquer une discrimination lorsque son utilisation s'effectue dans une société pluriculturelle, puisque cette dernière n'est pas constituée d'un seul groupe culturel. Au deuxième chapitre, nous avons souligné que la Chine utilise volontairement des SRÉ « culturellement limités » qui sont basés sur les expressions prototypiques « chinoises »[244], dans l'objectif d'arriver à des résultats plus précis. Le choix de la République populaire de Chine (RPC) de se limiter aux expressions faciales « chinoises » signifie que la RPC utilise consciemment des SRÉ

[244] Article 19. *op. cit*., p. 41.

limités qui se spécialisent dans la reconnaissance des émotions du groupe culturel Han, qui est le groupe dominant en RPC. Toutefois, la RPC admet aussi être une société pluriculturelle, dans laquelle nous retrouvons d'autres groupes culturels dont notamment les Hui, les Mongols, les Tibétains et les Ouïghours[245]. Les SRÉ conçus pour reconnaître les émotions des Han causeront probablement de la discrimination lors de la reconnaissance des émotions d'individus issus d'autres groupes culturels. Tout comme en RPC, l'utilisation de SRÉ au Canada engendrerait systématiquement de la discrimination puisque le Canada est une société pluriculturelle. Conséquemment, il n'est pas possible de soutenir que les limitations culturelles des SRÉ ne sont pas problématiques si l'utilisation de SRÉ se fait à l'intérieur des frontières d'une société pluriculturelle ou qu'un SRÉ conçu pour reconnaître les émotions des individus issus d'une certaine culture est utilisée sur les individus d'une autre culture.

Ainsi, à l'heure actuelle, les SRÉ ne semblent pas en mesure de satisfaire à l'exigence de la valeur de non-discrimination qui nécessite que tous et chacun reçoivent le même traitement, indifféremment des caractéristiques personnelles, culturelles, de genres, etc. D'une part, des systèmes situés culturellement ne peuvent prétendre à un universalisme dans leur utilisation et doivent reconnaître que leur développement est basé sur des choix discriminatoires, qui privilégient un groupe social en particulier. D'autre part, des systèmes situés culturellement ne peuvent être utilisés dans une société pluriculturelle puisque l'utilisation de ces systèmes discriminent certains groupes culturels ce qui avantage certains groupes et en désavantage d'autres, créant de la discrimination et, par le fait même, des injustices.

Sous l'analyse de la valeur de non-discrimination, nous pouvons dire que l'utilisation des SRÉ n'est pas universelle parce que les SRÉ discrimineront systématiquement les individus selon leurs classes sociales, leurs genres ou leur culture. Il semble que les SRÉ se trouvent pris entre l'évidence technique – que pour le moment ils ne peuvent qu'être situés culturellement et contextuellement – et l'exigence de non-discrimination – qui demande à ce qu'ils ne soient pas discriminatoires envers les différents groupes sociaux.

[245] Rekacewicz, P. « Chine, une mosaïque d'ethnies », *Le monde diplomatique*, 85, 2006, https://www.monde-diplomatique.fr/cartes/chineethnies ; Allès, E. et Robin, F. « L'impasse au Tibet », Journal OpenEdition, Perspectives Chinoises, 3, 2009, https://journals.openedition.org/perspectiveschinoises/5299#authors.

4.1.1.2. L'efficacité

L'efficacité est le rapport entre les résultats obtenus et les objectifs fixés[246]. Ainsi, pour être considérée efficace, une technologie doit prouver qu'elle est en mesure de réaliser les objectifs annoncés. Les SRÉ doivent, conséquemment, prouver qu'ils sont en mesure de reconnaître les émotions d'une personne. Or cet objectif en est un difficile à réaliser puisqu'il exige une inférence au sujet des émotions vécues à partir des expressions faciales ; une capacité qui n'est pas pour l'instant supportée par des preuves scientifiques reconnues[247].

Il est difficile d'être en mesure de prouver qu'il existe bel et bien un lien de causalité entre les émotions vécues et les expressions faciales. Affirmer que les émotions vécues provoqueraient des expressions faciales implique que certaines émotions vécues seraient en corrélation directe et significative avec certaines expressions faciales précises ; c'est-à-dire que, par exemple, lorsque l'humain est en colère, il fronce les sourcils.

Tel que nous l'avons vu au chapitre 1 et 2, les émotions sont des phénomènes complexes qui ne sont pas encore bien identifiés par les scientifiques. Plusieurs méthodes existent pour les observer et selon la méthode, les résultats diffèrent grandement[248]. En ce qui a trait plus spécifiquement au lien de causalité entre les émotions et les expressions faciales, nous avions soulevé l'argument darwinien selon lequel les expressions faciales peuvent être à la fois le symptôme des émotions vécues et le symptôme d'une volonté de communiquer[249], et ce, sans qu'une personne extérieure soit en mesure de pouvoir faire la différence entre les deux. Cette thèse pose un problème de taille en matière d'efficacité des SRÉ puisqu'il n'est pas possible de prouver que l'émotion identifiée par le système est une émotion vécue plutôt qu'un signal de communication et, conséquemment, il est impossible de prouver l'existence de l'inférence causale entre les émotions et les expressions faciales.

---

[246] Charbonneau, S., Cliche, D. et Rocheleau-Houle, D. « Les enjeux éthiques soulevés par la reconnaissance faciale », *Commission de l'éthique en science et en technologie*, 2020, https://www.ethique.gouv.qc.ca/media/2wqngchp/cest-j_2020_reconnaissance_faciale_acc_web.pdf. p. 17.
[247] Crawford, K. et al. « AI Now 2019 Report », *op. cit.* ; Article 19. « Emotional Entanglement: China's emotion recognition market and its implications for human rights », *op. cit.*
[248] Russell, J.A. et Barrett, L.F. « Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant », *op. cit.*, p. 805.
[249] Hess, U. et Thibault, P. op. cit., p. 122.

Par ailleurs, comme nous l'avons souligné plus haut, les expressions faciales sont aussi influencées par les normes sociales, le contexte et la culture. En ce sens, une personne aura tendance à modifier ses expressions faciales en fonction du contexte social dans lequel elle se trouve. Il faut donc prendre en compte qu'une même personne peut vivre une même émotion (ex. colère) deux fois dans la même journée et l'exprimer de deux façons très différentes, chaque fois selon le contexte dans lequel elle se trouve au moment où elle ressent l'émotion. Par ailleurs, la culture pose elle aussi un sérieux problème à l'efficacité des SRÉ. Un SRÉ ne peut espérer identifier les émotions vécues à partir d'une analyse du visage des individus en utilisant les mêmes règles de base dans toutes les sociétés, puisque selon les cultures, (1) les émotions vécues divergent et (2) la manière d'exprimer une émotion diverge. Cela signifie que les « émotions de base » sélectionnées au début de la phase de développement des SRÉ ainsi que les expressions faciales, associées aux émotions de base, varieront selon les cultures, créant un défi supplémentaire aux SRÉ afin de respecter le critère d'efficacité.

Les normes sociales, le contexte et la culture viennent poser des défis importants en ce qui concerne l'efficacité des SRÉ puisqu'ils viennent ajouter des variables difficiles à mesurer et augmentent les possibilités que les SRÉ soient moins efficaces sur certaines personnes et dans certains contextes. Or l'efficacité est pourtant l'un des critères fondamentaux que tout outil technologique doit être en mesure d'atteindre afin d'être considéré comme un outil fiable sur lequel nous pouvons reposer certains de nos jugements.

Conséquemment, devant les défis importants (et non relevés) que présentent le phénomène de l'émotion ainsi que les variables (contexte, normes sociales, culture) qui ont trait à ses (possibles) symptômes extérieurs que sont les expressions faciales, il semble que les SRÉ ne soient pas en mesure de répondre au critère d'efficacité. Tous les SRÉ qui sont utilisés en sécurité, marketing, éducation, politique, etc., ne sont pas des outils technologiques qui ont fait leur preuve au niveau de l'efficacité en plus d'avoir démontré une incapacité à incorporer trois aspects essentiels aux interactions sociales (contexte, normes sociales, cultures), ce qui vient ajouter une inquiétude sérieuse quant à leur utilisation. Cette incapacité de prendre en compte ces variables, qui sont pourtant des éléments centraux de nos interactions sociales quotidiennes, s'ajoute au premier problème, qui est celui de démontrer la véracité de l'inférence causale entre l'émotion et l'expression faciale. Tous ces

enjeux discréditent l'efficacité des SRÉ à apporter des résultats véridiques et fondés sur des preuves scientifiques fiables.

4.1.1.3. La fiabilité

Une technologie est fiable si les résultats qu'elle génère comportent un très faible pourcentage d'erreurs. Ici, les SRÉ se positionnent dans une situation exceptionnelle où l'évaluation même des résultats est problématique. En effet, puisqu'il n'est pas encore possible de prouver l'existence de l'inférence causale entre les émotions et les expressions faciales par des preuves convaincantes et satisfaisantes, il devient par la suite problématique de donner une quelconque crédibilité aux résultats qui succèdent l'analyse. Si le SRÉ n'a pour objectif que de reconnaître les expressions faciales prototypiques et non les expressions faciales d'émotions (réellement vécues par la personne), alors l'analyse des résultats peut être plus crédible. En effet, la question devient simplement du type : « est-ce que le SRÉ est en mesure de reconnaître un sourire et d'attribuer ce schéma de contractions musculaires donné à l'émotion de la joie ? ». Ce type de question est plus simple mais surtout moins pertinent. Étant donné que la fiabilité du système ne repose pas sur la concordance d'une émotion réellement vécue (joie) avec une expression faciale (sourire) mais simplement d'une association entre une expression faciale prototypique et une émotion exprimée (sourire égale toujours joie).

Cependant, même dans ces cas d'association simple, certains biais discriminatoires viennent remettre en cause la fiabilité des SRÉ à simplement reconnaître des expressions faciales prototypiques. En effet, pour mesure la fiabilité, il est possible de noter le nombre de « faux positif » et de « faux négatif ». Lorsque le SRÉ effectue une correspondance erronée entre, par exemple, une expression faciale et une émotion, le SRÉ effectue un « faux positif », c'est-à-dire qu'il identifie et reconnaît de manière fautive un lien qui n'est pas présent. Au contraire, en échouant à identifier une corrélation existante entre une expression faciale et une émotion, le SRÉ effectue un « faux négatif », c'est-à-dire qu'il ne parvient pas à identifier un lien qui est bel et bien présent. Conséquemment, plus une technologie génère de « faux positifs » et de « faux négatifs », moins elle est fiable[250]. En ce sens, lorsqu'un système comme Face++ attribue systématiquement des émotions négatives aux joueurs de la NBA à

[250] Charbonneau, S., Cliche, D et Rocheleau-Houle, D. *op. cit*. p. 17.

la peau noire, le système effectue des « faux positifs » puisqu'il identifie une corrélation qui n'est pas correcte. Or ces « faux positifs » ne sont pas anodins puisqu'ils sont reliés entre eux par un biais discriminatoire. Ces « faux positifs » révèlent donc deux lacunes importantes : (1) le SRÉ comporte un biais discriminatoire envers les personnes à la peau noire et (2) le SRÉ ne répond pas au critère de fiabilité puisque les résultats qu'il génère ne comporte pas qu'un très faible pourcentage d'erreurs.

Les impacts que les normes sociales et le contexte ont sur la variabilité des expressions faciales (en fonction du genre, du statut social, etc.), cumulé à l'impact des différences culturelles posent un doute sérieux quant à la fiabilité de cette technologie. Certaines recherches portent même à croire que les SRÉ génèrent des « faux positifs » et des « faux négatifs » de manière systématique et discriminatoire, ce qui est un risque accru pour les groupes sociaux visés par cette discrimination.

L'idée de l'existence de modèles contextuellement limités et culturellement situés pose un défi important pour un grand nombre de SRÉ. Les SRÉ qui auront des impacts importants sur la vie des individus comme ceux qui participent à l'évaluation des scores émotionnels et qui peuvent biaiser négativement les entrevues d'embauche, mener à un renvoi[251], se faire refuser une mise en libération conditionnelle, etc[252], comportent des risques accrus de causer des torts aux individus faisant l'objet de leur analyse. Nous avions dit qu'une technologie est fiable si les résultats qu'elle génère comportent un très faible pourcentage d'erreurs. Or les lacunes présentes dans les SRÉ sur le plan de la non-discrimination et de l'efficacité semblent suffisantes pour démontrer que les résultats générés par ces types de systèmes comportent en général un nombre d'erreurs significatif qui discrédite la fiabilité du système.

4.2. Les risques et les bénéfices de l'utilisation des SRÉ

Après l'analyse des SRÉ selon les trois critères de la non-discrimination, de l'efficacité et de la fiabilité, il semble que les SRÉ éprouvent de sérieuses difficultés à satisfaire aux exigences particulièrement en raison (1) du manque de preuves scientifiques quant à l'existence de corrélations significatives

[251] Crawford, K. et al. « AI Now 2019 Report », *op. cit.*, p. 6
[252] Ibid.

entre les expressions faciales et les émotions (2) de l'incapacité à performer de façon équitable à travers les classes sociales, les genres, les cultures ainsi que (3) des erreurs discriminatoires qui causent des torts importants aux personnes se situant à l'extérieur des limites du cadre de performance des SRÉ. Ces trois raisons sont des preuves suffisantes qui permettent d'affirmer que les SRÉ ne satisfont pas les critères de non-discrimination, d'efficacité et de fiabilité. Par ailleurs, comme nous l'avons vu au chapitre 1, les SRÉ constituent aussi un risque au libre-choix, en particulier ceux dont l'utilisation est malveillante.

Notre analyse, basée sur ces trois critères, devait démontrer que l'utilisation des SRÉ satisfaisait les critères pour répondre aux exigences de la justice sociale que nous avons relevées dans la section précédente, soit (1) celle d'avoir les mêmes droits et libertés comprenant notamment le droit à la vie, la sécurité et l'égalité pour tous et (2) celui d'avoir droit à un traitement juste qui inclut l'impartialité et la constance, peu importe la manière dont il est effectué.

Au début de notre analyse, nous avions dit que l'objectif sous-tendant les diverses utilisations des SRÉ était celui d'accéder à des données sur les états intérieurs qui n'auraient pas été autrement accessibles. Cet objectif se voit mis en péril par le manque de preuves scientifiques qui permettraient d'affirmer l'existence d'un lien de causalité entre les émotions et les expressions faciales. Ce manque de preuves vient discréditer la réussite des divers objectifs tel que la sécurité, le marketing, la politique, l'éducation, etc.

Puisque les possibilités d'utilisation des SRÉ sont très larges, nous prendrons l'exemple de la sécurité pour illustrer les risques de leur utilisation. Dans le contexte où l'utilisation des SRÉ avait pour objectif d'accroître la sécurité, les principaux risques seraient le risque de performer de façon inéquitable à travers les classes sociales, les genres et les cultures ainsi que le risque que les diagnostiques erronés engendrent des actes discriminatoires causant des torts importants aux personnes faisant l'objet du diagnostic.

Ces risques découlent d'une part d'une utilisation d'une technologie qui repose sur des méthodes infondées ou non-acceptées comme valables par la communauté scientifique. Cette utilisation de méthodes infondées ou non-acceptées par la communauté scientifique engendre le risque que l'utilisation des SRÉ ne permette pas d'atteindre l'objectif visé. Ici, nous visons une augmentation de la sécurité, par exemple, en utilisant les SRÉ dans l'objectif de détecter des signes précurseurs d'un comportement agressif afin d'agir de manière préventive plutôt que réactive. Or une méthode infondée

engendrera inévitablement des « faux-positifs » et des « faux négatifs », ce qui pourrait avoir pour effet de distraire continuellement les opérations de sécurité en détournant leur attention vers de fausses alertes et augmenter le risque qu'ils ne soient pas en mesure d'intercepter efficacement les véritables menaces lorsque celles-ci se présentent. De plus, comme nous l'avons expliqué, les « faux-positifs » risquent d'être corrélés à des groupes sociaux particuliers qui incluent les minorités visibles, les femmes et les personnes vulnérables. Un SRÉ qui a pour objectif de détecter les signes précurseurs d'un comportement agressif chez un individu devra corréler des expressions faciales à des intentions. En étant non fondé sur des preuves scientifiques fiables, les SRÉ auront pour objectif de détecter des expressions faciales de peur, de colère, d'anxiété ou de stress. Nous avons argumenté que les contextes sociaux influencent les expressions faciales émotionnelles des individus. Cette modulation de nos expressions faciales émotionnelles peut être causée par divers facteurs comme le genre et le statut social mais parfois les causes sont plus spécifiques. La tension entre les autorités et les groupes historiquement opprimés influence les expressions faciales émotionnelles de ces individus lorsque celles-ci se retrouvent en proximité[253]. Cette proximité peut engendrer du stress, de l'anxiété, de la peur ou de la colère sans pour autant que les individus l'exprimant aient l'intention d'adopter des comportements agressifs ou de causer du tort à autrui. L'utilisation de SRÉ à des fins de sécurité peut conséquemment être biaisé par des schémas historiques d'oppressions. Les SRÉ détectent à tort les individus appartenant à des groupes sociaux historiquement opprimés comme des menaces à la sécurité. Or, systématiquement identifier ces individus comme des menaces à la sécurité engendre le risque d'accentuer les tensions existantes entre les autorités compétentes et les groupes sociaux historiquement opprimés puisque l'identification répétée de ces individus vient confirmer les biais racistes et discriminatoires des membres de l'autorité et renforce conséquemment les stéréotypes négatifs à leur égard. Par ailleurs, être systématiquement identifié comme une menace à la sécurité renforce aussi les biais des groupes sociaux historiquement opprimés vis-à-vis des autorités compétentes. L'utilisation des SRÉ contribue donc aussi dans ce contexte à accroître les arrestations et les expulsions de lieux injustifiés du côté des autorités compétentes en plus d'augmenter le stress et l'anxiété que des expériences négatives répétées d'un traitement partial peut engendrer chez l'individu issu d'un groupe social historiquement opprimé. De manière générale, l'utilisation des SRÉ vient miner les efforts de confiance mutuelle et de résolution de tension entre ces groupes. L'utilisation des SRÉ à des fins de sécurité risque donc de renforcer un cercle vicieux et travaille dans le sens

[253] Holmes, S. C., Facemire, V. C. et DaFonseca, A. M. *op. cit.*

contraire de la justice sociale puisqu'elle vient brimer les individus des groupes sociaux historiquement opprimés de certains de leurs droits et libertés fondamentaux à travers des diagnostics basés sur des méthodes scientifiquement infondées. De plus, elle n'octroie pas un traitement juste puisqu'elle contribue à perpétuer des biais discriminatoires de manière systématique.

Ainsi, les bénéfices en sécurité sont ardus à reconnaître étant donné le manque de preuves scientifiques des méthodes utilisées. Toutefois, certains bénéfices peuvent être reconnus dans des contextes autres comme le marketing, la politique, etc. En marketing, l'analyse des expressions faciales émotionnelles des consommateurs peut par exemple donner des informations pertinentes sur l'expérience-client[254] et en politique, elle peut permettre aux campagnes politiques de confirmer ou d'infirmer la satisfaction des partisans à leur discours[255].

Conséquemment, devant les risques soulignés par notre analyse et devant le peu de bénéfices, l'utilisation des SRÉ ne semble pas, présentement, être en mesure de satisfaire les exigences de la justice sociale et il semble que leur utilisation ne soit, en général, pas en cohérence avec celle-ci.

4.3. Bienfaisance et utilisation des SRÉ

Compte tenu des enjeux éthiques importants que pose l'utilisation des SRÉ, il peut sembler que, pour le moment, les SRÉ ne sont simplement pas prêts, sur le plan technologique, à être utilisés, puisque leur utilisation pose des risques éthiques sérieux, engendre très peu de bénéfices et ne satisfait pas les exigences de la justice sociale. Malgré tout, nos recherches nous ont mené à découvrir certains types spécifiques de SRÉ qui semblent être en mesure d'apporter une contribution positive à la justice sociale en étant motivés par la bienfaisance. En effet, certaines compagnies tentent activement de déployer des SRÉ dont le développement est pensé et effectué sous le principe de bienfaisance. La bienfaisance est ici considérée une vertu publique, qui met en œuvre des actions dans l'objectif de contribuer au bien-être ainsi que d'apporter de l'aide et du support (physique, moral, psychologique, émotionnel, etc.) à autrui. La bienfaisance est un principe exigeant qui, à la différence de la bienveillance, qui se concentre sur la « volonté de » ou la « disposition à », porte son attention non sur

[254] Derbaix, C. et Pham, M. T. *op. cit.*
[255] Zittrain, J. *op. cit.*, In : Chander, A. *op. cit.* ; Gonzalez, R. J. *op. cit.*; Manokha, I. *op. cit.*

l'intention mais sur le résultat de l'action[256]. Conséquemment, un effort de plus est nécessaire pour respecter le principe de bienfaisance et cet effort doit être respecté pour être en mesure de juger si un type de SRÉ est en cohérence ou non à la justice sociale.

Ainsi, en servant d'outils technologiques à certains individus, ces types de SRÉ permettraient à des individus spécifiques de combler ou de réduire un écart existant entre eux et les autres membres de la société. De ce fait, ces SRÉ respecteraient le principe de bienfaisance et contribueraient à la justice sociale. Nous analyserons trois types de SRÉ dont le développement et le déploiement sont motivés par la bienfaisance dans l'objectif de déterminer si certains SRÉ pourraient activement participer à la justice sociale.

4.3.1. Robots sociaux pour personnes âgées vulnérables

Parmi les SRÉ conçus dans l'objectif d'être des outils technologiques d'aide à l'individu, nous croyons que certains robots sociaux pourraient participer à la justice sociale en apportant une aide et du support émotionnel. Cette aide engendrerait des bénéfices importants au niveau du bien-être général de l'individu ainsi qu'au niveau de sa santé psychologique et physique. Les robots sociaux sont des robots autonomes qui interagissent avec les êtres humains. Ces robots sont pour l'instant assez limités. Ils sont principalement guidés par des règles qui dictent leurs comportements sociaux ainsi que leurs mouvements.

Malgré leurs limitations, certains robots sociaux par exemple, peuvent, grâce à des interactions spécifiques, apporter une aide considérable à des personnes vulnérables ; une aide qui n'aurait pu être apportée par un être humain. Le robot PARO par exemple, a été créé dans l'objectif d'outiller les professionnels de la santé en leur procurant un robot qui, intégrant plusieurs technologies de pointes, permet de créer des bénéfices similaires à ceux de la thérapie animalière[257]. Sous l'apparence d'un phoque blanc et doux, ce robot a été pensé et conçu pour aider notamment les personnes atteintes de troubles du comportement et de la communication (symptômes de la maladie d'Alzheimer par exemple)

[256] Merlier, P. « Bienveillance, bienfaisance, bientraitance », *Philosophie et éthique en travail social. Manuel*, sous la direction de Merlier, P. « Politiques et interventions sociales », Presses de l'EHESP, France, 2013, 45-49.
[257] Inno3Med. « PARO », *Inno3Med*, 2018, https://www.phoque-paro.fr/.

« en procurant à ces personnes une amélioration de leur bien-être et de leur qualité de vie dans un cadre non médicamenteux[258]».

Illustration 4.1. Le robot PARO pour les personnes âgées vulnérables

Source image : https://www.phoque-paro.fr/.

PARO est en mesure de développer une interaction proactive puisqu'il réagit s'il est interpellé ou caressé. Ce robot social est en mesure de simuler certaines émotions spécifiques et reconnaissables par l'humain tel que la joie, le mécontentement et la surprise. Ces émotions sont discernables à travers la gestuelle et les sons émis par l'animal, ce qui est totalement différent d'une affirmation verbale tel que « Je suis heureux de jouer avec toi », qu'un robot humanoïde aurait pu utiliser. Plusieurs études ayant été effectuées sur le terrain avec des robots PARO démontrent que les émotions sont suffisamment bien simulées pour que la majorité des participants soient en mesure d'interagir avec le robot[259].

D'ailleurs, cette volonté du patient d'interagir par lui-même avec le robot a des effets thérapeutiques qui incluent, entre autres choses, une amélioration généralisée de l'humeur (qui contribue à la diminution des symptômes de dépression[260]) et un sentiment de détente (qui permet de contrer certaines douleurs physiques chroniques[261]). Ainsi, le robot PARO semble présenter certains avantages de la zoothérapie sans en subir les inconvénients pour l'humain. Avec un animal réel, on

[258] Ibid.
[259] Pu, L., Moyle, W., et Jones, C. « How people with dementia perceive a therapeutic robot called PARO in relation to their pain and mood: A qualitative study», *Journal of Clinical Nursing*, Volume 29, (3-4), p.437-446, 2020, https://doi.org/10.1111/jocn.15104.
[260] Pu, L., Moyle, W., et Jones, C. *op. cit.* ; Birks et al. « Robotic Seals as Therapeutic Tools in an Aged Care Facility: A Qualitative Study », *Journal of Aging Research*, vol. 2016, 2016, https://doi.org/10.1155/2016/8569602
[261] Voir Roger, K. et al. « Social commitment Robots and Dementia », *Canadian Journal on Aging*, vol. 31, n. 1, 2012, p. 87-94, https://muse.jhu.edu/article/468572

observe un accroissement de l'anxiété face au risque de griffure ou de morsure, des allergies, etc. D'ailleurs ces avantages ont aussi un impact sur les animaux, qui courraient des risques de maltraitance (involontaire ou volontaire) de la part du personnel comme de la part du patient.

L'idée innovante du robot PARO consiste, d'un côté, à contrer les limitations connues en ce qui concerne le langage et, de l'autre, à laisser le patient déduire lui-même des états du robot, de même que de sa réelle capacité de compréhension. Le robot devient dès lors une simulation complète d'un animal de compagnie qui serait d'une grande sympathie envers son maître. Parce qu'il expose une forme animale, le patient ne s'attend pas à ce que le robot soit en mesure de communiquer avec des mots et modifie ainsi ses méthodes de communication pour s'adapter à celles du robot-animal en optant plutôt pour le touché et le contact visuel.

Cette approche permet de contrer les principales limitations quant au contexte, puisqu'il n'est pas attendu du robot que celui-ci soit en mesure de le prendre en compte. En fait, les grandes limitations contextuelles du robot sont chez PARO tournées en forces. Tout comme il n'est pas attendu d'un animal qu'il fasse des distinctions entre les genres, les classes sociales ou les cultures, le patient ne s'attend pas à ce que le robot fasse de telles distinctions. Par ailleurs, les objectifs de ce robot le soustraient aussi à certaines limitations culturelles. La maladie d'Alzheimer est une maladie dont les patients présentent les mêmes symptômes sans qu'il n'y ait de différence observée à travers les cultures (malgré le fait que l'approche adoptée vis-à-vis de la maladie peut différer). Conséquemment, PARO est un outil d'aide thérapeutique qui transcende les différences culturelles, d'une part par son apparence de phoque et, d'autre part, de par sa communication par des sons et non par des mots. En ce sens, son langage tout comme ces gestes sont adéquats pour tous les êtres humains puisqu'il est conçu pour laisser place à l'interprétation de la part du patient. Cela dit, notons que PARO est fondé sur un système d'IA très rudimentaire.

4.3.1.1. Bilan pour les SRÉ

Par sa capacité démontrée à diminuer les symptômes de dépression ainsi qu'à contrer certaines douleurs physiques chroniques, ce SRÉ, sous la forme d'un robot social, contribue à améliorer le bien-

être des patients et, de ce fait, respecte le critère de bienfaisance. Par ailleurs, la simplicité de sa technologie permet d'éviter les principaux enjeux éthiques qui avaient été relevés avec les autres SRÉ soit l'incapacité à performer de façon équitable à travers les classes sociales, les genres et les cultures et les erreurs discriminatoires d'analyse qui causent des torts importants aux personnes ciblées. Conséquemment, les robots de ce type respectent à la fois les exigences de la justice sociale – soit celle d'avoir les mêmes droits et libertés pour tous et celle d'avoir droit à un traitement juste – et le principe de bienfaisance. Le respect de ces principes démontre que ces types de SRÉ sont non seulement en cohérence avec la justice sociale mais qu'ils participent activement à améliorer celle-ci.

4.3.2. Robots sociaux pour enfants sur le spectre de l'autisme

Du côté des robots sociaux plus complexes que PARO, certains robots sociaux sous des formes plus ou moins humanoïdes peuvent, grâce à certains jeux interactionnels et éducatifs, aider certains enfants présentant un trouble du spectre de l'autisme (TSA) à développer certaines capacités émotionnelles tel l'empathie et la compassion qui engendrent des impacts positifs sur leurs capacités à interagir socialement[262].

Le TSA étant un trouble très répandu et en constante augmentation (affectant près de 1% de la population mondiale et 1 enfant sur 40 aux États-Unis selon la Fédération Québécoise de l'Autisme[263]), un nombre considérable de recherches lui ont été consacré. Chez les individus se situant sur le spectre de l'autisme, la reconnaissance des émotions, qu'elle soit les siennes ou celles des autres, est généralement une tâche ardue. Alors que les êtres humains qui ont un développement typique[264] réussissent en général à interpréter les émotions d'autrui via un processus multisensoriel qu'effectue leur cerveau sans même qu'ils en aient conscience[265] (une interprétation comprend entre autres

[262] Voir notamment : APF France. « Leka, un robot pour les enfants en situation de handicap », APF France, 2019, https://www.apf-francehandicap.org/actualite/leka-un-robot-pour-les-enfants-en-situation-de-handicap-21349 (Page consultée le 2 mai 2021) ; S.A. « Buddy, le premier robot compagnon émotionnel », Blue Frog Robotics, 2018, https://buddytherobot.com/fr/buddy-le-robot-emotionnel-famille/ (Page consultée le 20 novembre 2019)
[263] Fédération Québécoise de l'Autisme. « L'autisme en chiffres », Fédération Québécoise de l'Autisme, https://www.autisme.qc.ca/tsa/lautisme-en-chiffres.html#:~:text=Gouvernement%20du%20Canada-,Dans%20le%20monde,environ%201%25%20de%20la%20population (Page consultée le 8 février 2021)
[264] Traduction libre de « typical development »
[265] Klucharev, V. et Sams, M. « Interaction of gaze direction and facial expressions processing: ERP study », *PubMed*, Volume 15, (4), p. 621–626, 2004, doi : doi:10.1097/00001756-200403220-00010.

choses une analyse de la prosodie verbale[266], des expressions faciales et des mouvements corporels[267]), les individus sur le spectre de l'autisme quant à eux évitent en général les contacts visuels et ont de la difficulté à reconnaître les expressions faciales[268], ce qui affaiblie l'efficacité de leurs interprétations[269]. De ce fait, ces individus éprouvent plus fréquemment que les autres de la difficulté à effectuer l'interprétation nécessaire pour être en mesure de comprendre les émotions des autres. C'est en partie pour cette raison que la communication et l'interaction avec d'autres êtres humains est souvent une tâche ardue, voire pénible. Ce désavantage au niveau des relations et des communications peut rendre par exemple difficile la création d'amitié et ce, dès l'enfance. De plus, les performances en relation sociale ainsi qu'en communication sont des performances recherchées dans plusieurs métiers, conséquemment, une faiblesse à ce niveau pourrait être la cause d'un refus à l'embauche. Ces difficultés peuvent engendrer des inégalités dans les chances ainsi que dans les possibilités de vies.

Les robots sociaux présentent plusieurs avantages que les humains ne peuvent offrir aux enfants se situant sur le spectre de l'autisme. En effet, les stimuli sur les plans de la prosodie verbale, des expressions faciales et des mouvements corporels sont drastiquement réduits et simplifiés chez le robot, ce qui apparaît comme un avantage chez l'enfant TSA. De plus, un robot peut répéter un nombre infini de fois la même phrase en ayant, à chaque répétition, exactement la même expression faciale, les mêmes mouvements corporels, le même ton de voix, la même intonation, etc., ce qui rend l'interprétation plus facile pour l'enfant sur le spectre de l'autisme. Leur constance est l'une des principales qualités que ces robots possèdent. Cette précision dans la répétition permet de gagner généralement la confiance des enfants, qui se sentent rassurés par la répétition qu'ils retrouvent dans les comportements du robot. Puisque cette répétition n'est pas quelque chose d'atteignable pour un être humain, elle devient un atout du robot. De plus, en complémentarité, le robot ne dérogera jamais des actions ni des phrases ni des émotions qu'il a apprises. Cette autre forme de précision est aussi

[266] La prosodie verbale fait référence à l'ensemble des traits oraux que nous utilisons dans l'expression verbale : inflexion, ton, intonation, accent, modulation, etc. Voir : CNRTL. « Prosodie », *Centre national de ressources textuelles et lexicales*, https://www.cnrtl.fr/definition/prosodie/substantif (Page consultée le 4 mai 2021)
[267] Kuusikko, S. et al. « Emotion Recognition in Children and Adolescents with Autism Spectrum Disorders » *Journal of Autism and Developmental Disorder,* vol. 39, p. 938–945, https://doi.org/10.1007/s10803-009-0700-0.
[268] Selon le Dr Nouchine Hadjikhani, cet évitement serait causé par une sur-activation du système subcortical du cerveau qui serait dû à un déséquilibre entre les neurotransmetteurs qui stimulent le cerveau et ceux qui l'apaisent. Voir SantéLog. « Autisme : Pourquoi le contact visuel est difficile », *Santé Log*, 2017, https://www.santelog.com/actualites/autisme-pourquoi-le-contact-visuel-est-difficile (Page consultée le 4 mai 2021)
[269] Kuusikko, S. *op. cit*.

hors d'atteinte pour une personne. En ce sens, alors que la répétition systématique d'une même phrase ou la démonstration d'une même émotion sont généralement perçues comme une lacune, elles sont, dans ce cas-ci, considérées comme une force par les enfants sous le spectre de l'autisme.

En thérapie, les enfants sur le spectre semblent démontrer en général des progrès remarquables lorsqu'ils travaillent avec un robot social, surtout en ce qui concerne leur langage spontané, leurs comportements stéréotypés et répétitifs et leur niveau d'attention[270]. L'étude menée par Pennisi et al. – qui est une recension de 29 études sur le développement des enfants sur le spectre de l'autisme avec un robot social – a démontré que les enfants qui avaient reçu des thérapies avec des robots sociaux avaient amélioré leur langage spontané, ce qui signifie que ces enfants avaient parlé avec plus d'aisance, de rapidité et de facilité avec le robot qu'ils ne l'avaient auparavant fait avec un humain[271]. De plus, les enfants qui avaient expérimenté la relation humain-machine avaient en général réduit le nombre ou la fréquence de comportements stéréotypés ou répétitifs[272]. Enfin, le niveau d'attention que portent les enfants envers le robot est en général plus élevé que celui qu'ils portent à un humain lorsqu'ils sont en thérapie[273].

4.3.2.1. Bilan pour les SRÉ

Ces robots sociaux ont démontré avoir des impacts positifs sur les individus avec lesquels ils interagissaient en améliorant les capacités interactionnelles et communicationnelles ainsi que la gestion émotionnelle des enfants sur le spectre. Toutefois, ces SRÉ n'ont pas démontré être en mesure de respecter les exigences fondamentales de la justice sociale. La plupart des recherches étant effectuées sur des petits groupes dont la culture ou les origines ethniques des participants ne sont pas révélées[274], ces SRÉ ne sont pas en mesure de prouver que leur système de reconnaissance des émotions reconnaît les émotions de tous les enfants, indépendamment de leur culture. Par exemple, une discrimination à l'égard d'enfants issus de certaines cultures pourrait mener les robots à noter à

[270] Pennisi, P. et al. « Autism and Social Robotics: A Systematic Review », *Autism Research*, Volume 9 (2), p. 165-183, 2015, https://doi.org/10.1002/aur.1527, p. 171.
[271] Ibid.
[272] Ibid., p. 171.
[273] Ibid., p. 169
[274] Voir la recension de Pennisi, P. et al., *op. cit*.

tort des émotions plus négatives en général sur certains d'entre eux. Ce biais discriminatoire pourrait biaiser les analyses des robots sociaux et conséquemment empêcher de créer une relation d'aide aussi performante. Ce risque éthique de discrimination met aussi en péril les bénéfices qui pourraient être acquis à travers la relation d'aide humain-machine. Si le robot identifie des émotions erronées chez l'enfant, dû à un biais discriminatoire, cette identification erronée serait, comme nous l'avons mentionné plus haut, répétée. La répétition d'une affirmation erronée sur les émotions de l'enfant (ex. « pourquoi es-tu fâché aujourd'hui? ») pourrait affecter négativement la gestion de ces émotions ainsi que ses performances communicationnelles.

En ce sens, si les résultats des études permettent de décréter que ces types de SRÉ rencontrent les exigences du principe de bienfaisance, il est important de préciser que les groupes à l'étude sont des petits groupes où la présence ou non de diversité culturelle n'est pas précisée, relativise le succès des résultats. Toutefois, puisqu'il n'y a pas non plus d'études qui avancent la possibilité de discrimination de la part du robot sur des groupes d'enfants, il nous est difficile de décréter si ces SRÉ satisfont aux exigences de la justice sociale. Toutefois, étant donné que notre recherche nous a mené à identifier la discrimination comme étant l'un des principaux enjeux de l'utilisation des SRÉ, nous croyons qu'il est nécessaire que les informations manquantes quant à la diversité culturelle soient dorénavant incluses dans les recherches sur les performances des robots sociaux auprès des enfants sur le spectre de l'autisme. Une analyse négative des émotions d'un enfant sur le spectre de l'autisme pourrait le mener à confondre davantage ses émotions et celles des autres et détériorer ses communications avec les autres. Ce biais culturel et discriminatoire aurait donc pour effet d'annuler les avantages de la bienfaisance en plus de ne pas respecter les exigences de justice sociale.

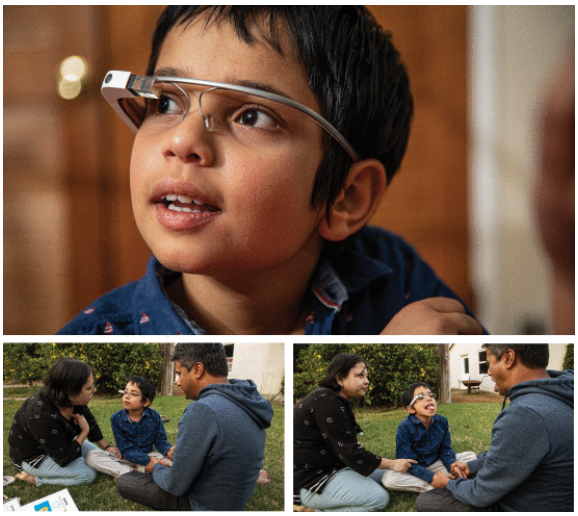
En conséquent, il nous est difficile de déterminer si nous considérons ce type de SRÉ comme satisfaisant les exigences de la justice sociale et de la bienfaisance. Nous considérons que l'idée derrière le développement de ces robots est une idée motivée par la bienfaisance qui, toutefois, ne satisfait que partiellement les exigences de la justice sociale. Alors que son objectif est de permettre aux enfants sur le spectre de l'autisme d'acquérir des compétences émotionnelles, sociales et communicationnelles qui lui permettrait, en théorie, d'augmenter par exemple ses possibilités d'accès à des emplois stables, les risques que ces SRÉ soient conçus pour performer sur certains groupes sociaux uniquement (par exemple les personnes caucasiennes) vient contrecarrer dans une certaine

mesure l'objectif de départ qui est un objectif de performance « universelle », c'est-à-dire, tous groupes sociaux confondus.

4.3.3. Objets technologiques d'assistance

Les robots sociaux ne sont toutefois pas les seuls outils technologiques qui doivent être considérés dans la catégorie des SRÉ qui servent d'outils d'aide aux individus. Les lunettes créées par la compagnie Google par exemple, sont spécifiquement conçues pour aider les individus sur le spectre de l'autisme à identifier les émotions de la personne avec laquelle ils interagissent[275,276].

Illustration 4.2. Les lunettes de reconnaissance des émotions de Google



Source image : https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9055973.

Le système à l'intérieur des lunettes fonctionne en deux grandes séquences. En premier lieu, il cherche des visages à l'intérieur du champ de vision de l'individu. Une fois qu'il identifie un visage, une lumière verte s'allume et vient encadrer le visage de l'individu pour signifier qu'il a été détecté[277]. En second

[275] Sahin, NT. et al. « Second Version of Google Glass as a Wearable Socio-Affective Aid: Positive School Desirability, High Usability, and Theoretical Framework in a Sample of Children with Autism », *JMIR Human Factors,* Volume 5 (1), 2018, 12 p., https://humanfactors.jmir.org/2018/1/e1/.
[276] Voir Haber, N., Voss, C. et Wall, D. « Making emotions transparent: Google Glass helps autistic kids understand facial expressions through augmented-reaiity therapy », *IEEE Spectrum*, vol. 57, (4), 2020, 46-52, 2020, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9055973., ainsi que Sahin, NT. et al. *op, cit.*
[277] Haber, N., Voss, C. et Wall, D. *op. cit.*

lieu, l'« affichage tête haute » (head-up display) fera apparaître un emoji, un émoticône ou un mot dans le champ de vision de l'individu afin qu'il puisse comprendre quelle émotion a été identifiée par le système[278].

Illustration 4.3. La détection des points faciaux spécifiques par les lunettes Google

Source image : https://wall-lab.stanford.edu/projects/autism-therapy-on-glass/

Cet outil peut se révéler très efficace afin d'aider les individus dans leur communication et leur compréhension de l'Autre. Posséder un outil technologique qui aurait pour fonction de reconnaître les émotions des autres pourrait être une opportunité non-négligeable pour ces personnes.

4.3.3.1. Bilan pour les SRÉ

Ce type de SRÉ présente des avantages considérables sur le plan des enjeux éthiques par rapport aux robots sociaux. En effet, dans un premier temps, ils ne sont pas conçus pour découvrir les émotions *vécues* par les personnes mais plutôt les émotions *exprimées*. Ce qui signifie que dans le cas de ces SRÉ, on ne cherche pas à atteindre les émotions vécues à travers les expressions faciales, mais on tente plutôt d'aider l'individu à analyser et reconnaître les émotions liées aux expressions faciales prototypiques, qu'elles soient des émotions ou des signaux de communication. En ce sens, étant donné l'incapacité généralisée des SRÉ à différencier les expressions faciales émotionnelles des signaux de

[278] Voir Annexe C.

communication, avoir pour objectif de reconnaître simplement les expressions faciales prototypiques est dans ce contexte un avantage qui permet d'augmenter l'efficacité et la fiabilité.

Dans un deuxième temps, le système est conçu pour être en mesure de se familiariser avec les expressions faciales des proches[279], ce qui signifie qu'il sera, dans une certaine mesure, capable d'adaptation. Cette adaptation restreinte permet de résoudre dans une certaine mesure la difficulté que représentait les différences culturelles en ce qui a trait aux individus qui évoluent dans la sphère sociale de l'enfant sous le spectre de l'autisme (parents, frères et sœurs, etc.). Ces deux premiers avantages viennent donc résoudre plusieurs aspects importants des enjeux éthiques que nous avons soulevés concernant la non-discrimination, l'efficacité et la fiabilité.

Toutefois, si en se limitant à la reconnaissance des expressions faciales prototypiques et en développant une capacité d'adaptation ce type de SRÉ réussit en théorie à augmenter son efficacité, sa fiabilité et à diminuer la discrimination, il est nécessaire de s'assurer que cette capacité d'adaptation obtient les résultats escomptés, et ce, pour tous les groupes sociaux. En dépit du fait que certaines compagnies laissent sous-entendre que les différences culturelles et ethniques ne posent pas de problème pour leur technologie[280] en raison de leur capacité d'adaptation, il reste que néanmoins l'enfant ira à l'école, se promènera dans les magasins, etc. Il est conséquemment indéniable que son champ d'interaction dépassera le cercle de ses proches et excèdera en ce sens les limites de la capacité d'adaptation. Or de ce fait, il est clair que l'adaptation de la technologie ne résout pas le problème des biais discriminatoires en ce qui concerne toutes les interactions qui excèdent les proches de l'enfant.

Cette limitation soulève ainsi des risques importants qu'il importe de considérer. Les lunettes pourraient par exemple ne pas être en mesure de reconnaître les expressions faciales des individus à la peau noire avec autant d'efficacité que ceux à la peau blanche. Cette analyse discriminatoire des expressions faciales d'individus de groupes sociaux spécifiques pourrait causer des torts importants à son développement et à sa propre compréhension de ses émotions et des émotions des autres.

[279] Haber, N., Voss, C., et Wall, D. *op. cit.*
[280] Haber, N., Voss, C., et Wall, D. *op. cit.* Affirme que leur lunette « s'adapte » aux expressions faciales des proches.

Enfin, l'objectif des lunettes respecte le principe de bienfaisance puisqu'il vise à aider les enfants se situant sous le spectre de l'autisme à améliorer leurs interactions sociales ainsi qu'à leur apporter un support technologique. Toutefois, nous ne pouvons qu'affirmer pour le moment que l'objectif respecte le principe, mais non la technologie. Il est clair que leur volonté est d'apporter de l'aide ainsi qu'un support aux enfants, mais cette aide et ce support, dans leur phase pratique, n'ont pas démontré pour le moment respecter le principe de bienfaisance. Par ailleurs, en raison des discriminations potentielles, ce type de SRÉ est sur la bonne voie pour respecter les exigences de la justice sociale mais n'a pas démontré qu'il y était arrivé.

4.4. Conclusion

En conclusion, après avoir analysé l'utilisation générale des SRÉ selon les trois critères de non-discrimination, d'efficacité et de fiabilité, il nous apparaît que l'utilisation générale des SRÉ ne semble pas satisfaire les exigences de la justice sociale, qui demandent que tous reçoivent un traitement juste, c'est-à-dire un traitement impartial et constant, et, que les droits et libertés fondamentaux de tous soient respectés. En effet, dans plusieurs cas, l'utilisation des SRÉ pose de graves enjeux de discrimination. De plus, à plusieurs reprises, nous avons pu constater que les cas de discrimination ciblaient des groupes sociaux spécifiques tel que des minorités visibles et des femmes. Lorsque nous avons sous-pesé les risques et les bénéfices, nous avons constaté que les risques dépassaient grandement les bénéfices, et ce, pour plusieurs raisons qui découlent en général de l'absence de preuves scientifiques qui appuieraient l'idée d'un lien entre les expressions faciales et les émotions ressenties. En l'absence de ce lien, il devient difficile de voir l'atteinte des bénéfices envisagés, tel qu'une sécurité accrue ou une meilleure approche-client. Ainsi, les risques sont bien présents et les bénéfices sont difficiles à évaluer. Toutefois, en allant dans le domaine du *care*, nous avons relevé trois types de SRÉ qui, selon nous, auraient le potentiel de respecter les exigences de la justice sociale, ainsi que de démontrer, avec le respect du critère de bienfaisance, qu'ils participent activement à améliorer celle-ci. Suite à notre analyse, nous pouvons seulement affirmer que le premier type, soit les SRÉ du type PARO, sont en mesure de satisfaire les exigences de la justice sociale ainsi que de respecter le principe de bienfaisance dans la volonté comme dans l'acte. En ce qui concerne les deux types de SRÉ, soit les

robots sociaux et les lunettes pour les enfants ayant un trouble du spectre de l'autisme, ces derniers démontrent clairement dans leurs objectifs qu'ils ont la volonté de répondre au critère de bienfaisance. Toutefois, selon nous, leurs résultats ne permettent pas de démontrer qu'ils satisfont les exigences de la justice sociale, et ce, pour plusieurs raisons qui incluent entre autres le nombre très restreint d'études à leur sujet, le nombre limité d'échantillons ainsi que les risques de discrimination qui sont toujours présents et qui ne semblent pas avoir été suffisamment considérés, dans ce cas-ci par le fait qu'il n'y avait pas d'identification des performances selon les différents groupes culturels.

Conclusion

Ce projet de recherche avait pour objectif de déterminer si l'utilisation des SRÉ était en cohérence avec la justice sociale. Pour ce faire, nous avons relevé deux exigences au fondement de la justice sociale sur lesquelles nous appuyer pour poser notre jugement. Dans le premier chapitre, nous avons démontré la pertinence de se pencher sur l'utilisation des SRÉ en soulevant le rôle des émotions dans le processus de prise de décision. En confirmant la place essentielle des émotions dans notre processus de prise de décision, nous avons pu observer et entrevoir les diverses utilisations malveillantes des SRÉ et les risques vis-à-vis de notre liberté de choix que ces types d'utilisation représentent. L'utilisation malveillante des SRÉ est une utilisation que nous pouvons aisément juger comme n'étant pas en cohérence avec la justice sociale, toutefois, elle ne représente pas l'utilisation générale des SRÉ. En ce sens, dans notre deuxième chapitre, nous avons relevé l'un des grands problèmes éthiques de l'utilisation des SRÉ, soit la prétention à l'universalité dans la reconnaissance des états émotionnels ressentis par l'individu. Les états émotionnels sont des états subjectifs très personnels et intimes qui se vivent différemment chez les individus et qui sont modulés par plusieurs facteurs.

Le chapitre 2 soulève l'hypothèse que la prétention à l'universalisme des SRÉ est une prétention fausse et infondée. En ce sens, nous remettons en question la possibilité même de prétendre à un universalisme dans la reconnaissance des émotions en avançant que les SRÉ ne sont pas en mesure de faire la distinction entre la reconnaissance des signaux de communication et la reconnaissance des émotions. Par ailleurs, nous remettons en question les bases scientifiques d'une théorie de l'universalité des émotions (UÉ) en soulignant que les scientifiques n'ont pas réussi à prouver une théorie de l'UÉ alors que des décennies de recherches scientifiques ont attesté des différences irréductibles entre les cultures en ce qui a trait aux émotions et à notre rapport avec elles. Par ailleurs, nous avons aussi tenté d'expliquer que la simplification d'une théorie – comme ce fut le cas avec la théorie d'Ekman – peut mener à de fausses croyances comme celle de l'universalité de l'émotion dans l'expression faciale et la reconnaissance des expressions faciales. Sur ces fausses croyances se basent des SRÉ qui, en plus de menacer la vie privée par la récolte de renseignements sensibles (telles les données biométriques faciales), serviront d'aide à la décision dans des aspects importants de la vie des individus et peuvent conséquemment nuire à l'égalité des chances et l'égalité dans les possibilités de vie.

Notre troisième chapitre relève deux grandes limites des SRÉ qui viennent poser des défis importants aux valeurs de non-discrimination, d'efficacité et de fiabilité. Nous avons défendu l'idée que le contexte nous positionne parfois dans une situation particulière et unique vis-à-vis autrui qui nous motive parfois à agir ou exprimer certaines choses de diverses manières. Plusieurs facteurs viennent moduler le contexte dans lequel nous nous trouvons – tel que notre genre et notre statut social, mais aussi celui des autres individus impliqués

dans la situation – et ainsi, un même individu pourrait exprimer de plusieurs manières distinctes une même émotion dans la même journée, selon le contexte dans lequel il se retrouve. De même, la culture a une influence non négligeable sur le nombre d'émotions que nous avons à notre disposition ainsi que sur le champ d'action de nos émotions et c'est pourquoi nous l'avons identifiée comme la deuxième grande limite des SRÉ. La culture peut être l'un des principaux facteurs qui motive un individu à agir ou à réagir dans une situation particulière. En effet, nous avons relevé que les normes sociales, qui dictent souvent nos expressions faciales, influencent et sont issus du contexte et de notre culture d'appartenance. Ainsi, ces facteurs sont des facteurs qui modulent nos expressions faciales et viennent contrecarrer les tentatives de reconnaissance des émotions. Ces limitations posent conséquemment des enjeux éthiques importants qui limitent à la fois l'efficacité et la fiabilité des SRÉ sur divers groupes sociaux.

Notre dernier chapitre a pour objectif de poser un jugement sur la participation ou la non-participation des SRÉ à la justice sociale en procédant à une analyse éthique des enjeux et des limitations que nous avons soulevés dans les trois premiers chapitres et en les confrontant aux bénéfices liés à l'utilisation des SRÉ. Notre analyse éthique nous a permis de constater que les risques dépassaient grandement les bénéfices en ce qui a trait à l'utilisation des SRÉ. Les SRÉ ne sont pas des technologies fiables et nécessitent des bases scientifiques reconnues avant de pouvoir donner des constats justes et fiables sur lesquels il est possible de s'appuyer. Par ailleurs, les risques de discrimination semblent pour le moment inévitables et participeraient à renforcer les cercles vicieux déjà existant entre les groupes sociaux historiquement oppressés et les autorités compétentes. La discrimination des SRÉ peut potentiellement avoir des impacts négatifs majeurs sur la vie des individus ciblés tel que biaiser négativement les entrevues d'embauche, mener à un renvoi[281], se faire refuser une mise en libération conditionnelle, etc[282]. Toutefois, le domaine du *care* a attiré notre attention de par ses objectifs précis et positifs et nous a mené à une analyse de trois types différents de SRÉ qui pourraient potentiellement participer à la justice sociale. Par leurs objectifs de bienfaisance, l'utilisation de ces SRÉ s'est distinguée des autres et nous incite à croire que ces SRÉ pourraient être, contrairement aux autres types, très prochainement identifiés comme des SRÉ participant activement à la justice sociale. En ce qui concerne les autres types de SRÉ, nous croyons que, outre une base scientifique fiable, leur utilisation devrait être justifiée par des intentions de bienfaisance pour être en cohérence avec la justice sociale.

Ce projet de recherche a dû, pour des raisons d'espace, laisser de côté l'enjeu non-négligeable de la vie privée. En effet, l'utilisation des SRÉ comporte un aspect majeur d'intrusion dans la vie privée et intime des individus qu'elle analyse. Cet enjeu s'est vu laissé de côté dans ce projet de mémoire pour nous permettre d'analyser en

[281] Crawford, K. et al. « AI Now 2019 Report », *op. cit.*, p. 6.
[282] Ibid.

profondeur les limitations actuelles des SRÉ. Toutefois, il est nécessaire qu'un prochain projet de recherche se penche sur les enjeux de vie privée que soulève l'utilisation des SRÉ afin d'identifier les risques imminents et de guider les prochains encadrements. En utilisant cette technologie dans les événements, sur Internet, dans les magasins, par les services de police, etc., nous atteignons un nouveau degré de surveillance, qui dépasse la surveillance des états « externes » de l'individus – qui incluent notamment ces actes et ses paroles – à une surveillance des états « intérieurs » de l'individu, qui vise ses émotions, humeurs, croyances, traits de personnalités, etc. soit son ressenti subjectif. Par ailleurs, la question de la surveillance des états intérieurs et subjectifs de l'individu touche à la question ardemment débattue par les philosophes qui concerne les « qualia »,[283] c'est-à-dire ce qui relève du ressenti subjectif et unique auquel autrui n'a pas un accès direct[284]. Un projet pertinent pourrait donc se pencher sur la question des qualia, à savoir si une analyse fondée et fiable des états intérieurs d'un individu serait une analyse en mesure d'atteindre ces qualia.

[283] Jackson, F. « Epiphenomenal Qualia », *The Philosophical Quarterly*, 32(127), 127-136, 1982.
[284] Nagel, T. « What Is It Like to Be a Bat? », *The Philosophical Review*, 83(4), 435-450, 1974.

Bibliographie

ABC bourse. « La finance comportementale, les apports de la psychologie », *Finance comportementale*, Abc bourse, 2020, https://www.abcbourse.com/apprendre/19_finance_comportementale.html

Alcantara, C., Charest, F. et Agnostinelli, S. (Dir.), *Big Data et visibilité en ligne : un enjeu pluridisciplinaire de l'économie numérique*, Presse des Mines, Paris, 2018

Allès, E. et Robin, F. « L'impasse au Tibet », Journal OpenEdition, Perspectives Chinoises, 3, 2009, https://journals.openedition.org/perspectiveschinoises/5299#authors

APF France. « Leka, un robot pour les enfants en situation de handicap », APF France, France, 2019, https://www.apf-francehandicap.org/actualite/leka-un-robot-pour-les-enfants-en-situation-de-handicap-21349 (Page consultée le 2 mai 2021)

Aristote. *Éthique à Nicomaque* (trad. J. Tricot), Vrin, Paris, 2012

Arpaly N. *Unprincipled Virtue : An Inquiry Into Moral Agency,* Oxford University Press, New York, 2003

Article 19. « Emotional Entanglement: China's emotion recognition market and its implications for human rights », *Article 19*, U.K., 2021, https://www.article19.org/wp-content/uploads/2021/01/ER-Tech-China-Report.pdf.

Barrett, L. F. *How Emotions Are Made*, Houghton Mifflin Harcourt, New York, 2017

Barrett, L. et al. « Emotional Expressions Reconsidered : Challenges to Inferring Emotion From Human Facial Movements », *Psychological Science in the Public Interest*, 20(1), 1-68, 2019, https://journals.sagepub.com/doi/10.1177/1529100619832930

Bedford O. et Hwang K.-K. « Guilt and shame in Chinese culture: A cross-cultural framework from the perspective of morality and identity », *Journal for the Theory of Social Behaviour*, 33(2), 127–144, 2003.

Birks et al. « Robotic Seals as Therapeutic Tools in an Aged Care Facility: A Qualitative Study », *Journal of Aging Research*, vol. 2016, 2016, https://doi.org/10.1155/2016/8569602

Blue Frog Robotic. « Buddy, le premier tobot compagnon émotionnel », Blue Frog Robotics, 2018, https://buddytherobot.com/fr/buddy-le-robot-emotionnel-famille/ (Page consultée le 20 novembre 2019)

Briganti, G. « L'IA pour la détection des maladies génétiques grâce à la reconnaissance faciale », *NumeriCare*, 2019, https://www.numerikare.be/fr/actualites/e-health/l-rsquo-ia-pour-la-detection-des-maladies-genetiques-grace-a-la-reconnaissance-faciale.html (Page consultée le 3 février 2021)

Burts, C. « Brazil plans massive centralized biometric database of all citizens to improve agency data sharing », *Biometricupdate.com*, 2019, https://www.biometricupdate.com/201910/brazil-plans-massive-centralized-biometric-database-of-all-citizens-to-improve-agency-data-sharing (Page consultée le 6 mars 2021)

Briggs, J. L. *Never in Anger: Portrait of an Eskimo Family*, Harvard University Press, Harvard, 1971

Chander, A. « The Racist Algorithm? », *Michigan Law Review*, 115(6), 1023-1045, 2017, http://michiganlawreview.org/wp-content/uploads/2017/04/115MichL Rev 1023_Chander.pdf

Charbonneau, S., Cliche, D et Rocheleau-Houle, D. « Les enjeux éthiques soulevés par la reconnaissance faciale », *Commission de l'éthique en science et en technologie*, 2020, https://www.ethique.gouv.qc.ca/media/2wqngchp/cest-j_2020_reconnaissance_faciale_acc_web.pdf

Chen, C. et al. « Distinct Facial Expressions Represent Pain and Pleasure Across Cultures », *Proceedings of the National Academy of Sciences of the United States of America*, 115(43), 2018, E10013–E10021, https://www.pnas.org/content/115/43/E10013

Chen, Z. et Whitney, D. « Tracking the Affective State of Unseen Persons », *Proceedings of the National Academy of Sciences*, 2019, https://www.pnas.org/content/pnas/early/2019/02/26/1812250116.full.pdf

Chomsky, N. *Syntactic Structures*, Mouton, The Hague, Pays-Bas, 1957

Clément, C. et Cixous, H. *La jeune née*, Union générale d'Éditions, Paris, 10/18, série Féminin Futur, 1975

CNRTL. « Prosodie », *Centre national de ressources textuelles et lexicales*, s.d., https://www.cnrtl.fr/definition/prosodie/substantif (Page consultée le 4 mai 2021)

Commissariat du Canada à la protection de la vie privée. « Des données au bout des doigts : La biométrie et les défis qu'elle pose à la protection de la vie privée » Commissariat du Canada à la protection de la vie privée, 2011, https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privee/renseignements-sur-la-sante-renseignements-genetiques-et-autres-renseignements-sur-le-corps/gd_bio_201102/ (Page consultée le 14 avril 2021)

Commission européenne. « Façonner l'avenir numérique de l'Europe: la Commission présente des stratégies en matière de données et d'intelligence artificielle », *Commission Européenne*, 2020, https://ec.europa.eu/commission/presscorner/detail/fr/ip_20_273 (Page consultée le 15 mars 2021)

Cornillie, C. « Finding Artificial Intelligence Money in the Fiscal 2020 Budget », *Bloomberg government*, 2019, https://about.bgov.com/news/finding-artificial-intelligence-money-fiscal-2020-budget/ (Page consultée le 19 mars 2021)

Cour européenne des droits de l'homme. « Convention européenne des droits de l'homme », *Cour européenne des droits de l'homme*, s.d., https://www.echr.coe.int/Pages/home.aspx?p=basictexts&c=fre (Page consultée le 4 avril 2021)

Cours de droit. « Quelle différence entre droit naturel et droit positif ? », *Cour de droits, 2019,* https://cours-de-droit.net/quelle-difference-entre-droit-naturel-et-droit-positif-a121611600/ (Page consultée le 30 mars 2021)

Crawford, K. *The Atlas of AI : Power, Politics and the planetary cost of Artificial Intelligence*, Yale University Press, New Haven et Londres, 2021

Crawford, K. et al. « AI Now 2019 Report », *AI Now Institutes*, New York, 2019, https://ainowinstitute.org/AI_Now_2019_Report.html

Damasio, A. *L'erreur de Descartes*, Odile Jacob, Paris, 2010

Darpy, D. et Guillard, V. *Comportements du consommateur; concepts et outils,* Dunod (4e éd.), 2016, https://books.google.ca/books?hl=fr&lr=&id=LbMcDQAAQBAJ&oi=fnd&pg=PP3&dq=GAFA+et+%C3%A9motions&ots=nh-Kh2iHw4&sig=7zcjNYpe5ksa4mCtbqNVsjtbY9Y&redir_esc=y#v=onepage&q=%C3%A9motion&f=false

Darwin, C. « The expression of the emotions in man and animals », *University of Chicago Press*, Chicago, 1965

Déclaration de Montréal. « La Déclaration de Montréal pour un développement responsable de l'intelligence artificielle », *Déclaration de Montréal*, Principe d'inclusion de la diversité, 2018, https://www.declarationmontreal-iaresponsable.com/la-declaration

Delbrouck, P. « Les émotions humaines peuvent-elles être discrètes ? » *La Lettre du Psychiatre*, 7(1), Nouvelles technologies, 16-19, 2016, https://www.edimark.fr/Front/frontpost/getfiles/23843.pdf

Derbaix, C. et Pham, M. T. « Pour un développement des mesures de l'affectif en marketing : synthèse des prérequis ». *Recherche et Applications en Marketing*, SAGE, Association française du marketing, 4(4), 71-87, 1989, https://www-jstor-org.acces.bibl.ulaval.ca/stable/pdf/40588767.pdf?ab_segments=0%252Fbasic_search%252Fcontrol&refreqid=excelsior%3A4659bdda8ad2106fcdc2b8a4b08c2b63

Descartes, R. *Les Passions de l'âme [1649]*, Vrin, Paris, 1994

Descartes, R. *Méditations métaphysiques [1641],* Flammarion, Paris, 2011

Dreyfus, H. *Alchemy and Artificial Intelligence*, Rand Corporation, New York, 1965

Dreyfus, H. *What Computers Can't Do : A Critique of Artificial Reason*, Harper & Row, New York, 1972

Ekman, P. « An argument for basic emotions », *Cognition and Emotions*, 6(3/4), University of California, San Francisco, 1992, https://www.tandfonline.com/doi/abs/10.1080/02699939208411068

Ekman, P. « The argument and evidence about universals in facial expressions of emotion », chap. 6 in H. Wagner and A. Manstead (Eds.), *Wiley Handbook of social psychophysiology*, *Handbook of social psychophysiology*, University of California, San Francisco, 1989, https://psycnet.apa.org/record/1989-97735-006

Fehr, B. et Russell, J. « Concept of Emotion Viewed from a Prototype Perspective », *Journal of experimental psychology*, 113(3), 464-486. 1984, https://psycnet.apa.org/doi/10.1037/0096-3445.113.3.464

Forsé, M. et Parodi, M. « Justice distributive : La hiérarchie des principes selon les Européens ». *Revue de l'OFCE*, 3(3), 213-244, 2006, https://doi.org/10.3917/reof.098.0213

Gillum, J. et Kao, J. « Aggression Detectors : The Unproven, Invasive Surveillance Technology Scholls Are Using to Monitor Students », *ProPublica*, 2019, https://features.propublica.org/aggression-detector/the-unproven-invasive-surveillance-technology-schools-are-using-to-monitor-students/

Gollier, C., Hilton, D. & Raufaste, É. « Daniel Kahneman et l'analyse de la décision face au risque », *Revue d'économie politique*, 113(3), 295-307, 2003, https://www.cairn.info/revue-d-economie-politique-2003-3-page-295.htm

Gonzalez, R. J. « Hacking the citizenry? Personality profiling, big data and the election of Donald Trump », *Anthropology Today*, 33(3), 2017.

Gouvernement du Canada. « Loi sur le ministère des Femmes et de l'Égalité des genres », Gouvernement du Canada, https://laws-lois.justice.gc.ca/fra/lois/W-11.3/page-1.html (Page consultée le 4 juin 2021)

Grand View Research. « Facial Recognition Market Size, Share & Trends Analysis Report By Technology, 2021 – 2028 », *Grand View Research*, 2021, https://www.grandviewresearch.com/industry-analysis/facial-recognition-market

Haber, N., Voss, C., et Wall, D. « Making emotions transparent: Google Glass helps autistic kids understand facial expressions through augmented-reality therapy », *IEEE Spectrum*, vol. 57(4), 46-52, 2020, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9055973

Hamdi H. « Plate-forme multimodale pour la reconnaissance d'émotions via l'analyse de signaux physiologiques: application à la simulation d'entretiens d'embauche ». *Thèse de doctorat*. Université d'Angers, France, 2012

Heaven, Douglas. « Why Faces Don't Always Tell the Truth about Feelings » *Nature*, 2020. https://www.nature.com/articles/d41586-020 -00507-5

Hess, U. et Thibault, P. « Darwin and Emotion Expression », *American Psychological Association*, 64(2), 120 –128, 2009, https://psycnet.apa.org/fulltext/2009-01602-003.pdf?auth_token=a75e8192b84781dc3dc4facff1a983dedf27d78e

Holmes, S. C., Facemire, V. C. et DaFonseca, A. M. « Expanding criterion a for posttraumatic stress disorder: Considering the deleterious impact of oppression ». *Traumatology, 22*(4), 314–321, https://doi.org/10.1037/trm0000104

Hu, J., Lu, J. et Tan, Y. « Discriminative Deep Metric Learning for Face Verification in the Wild », *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR),*1875-1882, 2014, http://openaccess.thecvf.com/content_cvpr_2014/papers/Hu_Discriminative_Deep_Metric_2014_CVPR_paper.pdf

Inno3Med. « PARO », *Inno3Med*, 2018, https://www.phoque-paro.fr/

Jackson, F. « Epiphenomenal Qualia », *The Philosophical Quartely*, 32(127), 127-136, 1982

Journal Officiel de l'Union européenne. « Règlement du parlement européen et du conseil », Journal officiel de l'Union européenne, 2016, https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32016R0679 (Page consultée le 3 février 2020)

Kahneman, D., & Tversky, A. « Prospect Theory: An Analysis of Decision under Risk », *Econometrica, 47*(2), 263-291, 1979, www.jstor.org/stable/1914185

Kant, E. *Fondements de la métaphysique des mœurs [1785]* (trad. V. Delbos), Éditions Les Échos du Marquis, France, 2013

Kant, E. *Anthropology from a pragmatic point of view [1796]* (trad. M. J. Gregor), The Hague, Pays-Bas, 1974

Klucharev, V. et Sams, M. « Interaction of gaze direction and facial expressions processing: ERP study », *PubMed*, Volume 15, (4), 621–626, 2004, doi:10.1097/00001756-200403220-00010

Kosinski, M., Stillwell, D. et Graepel, T. « Private traits and attributes are predictable from digital records of human behavior », *National Academy of Sciences*, 5802-5805, 2013, https://www.pnas.org/content/pnas/110/15/5802.full.pdf?3=

Kumar, P. « Corporate Privacy Policy Changes during PRISM and the Rise of Surveillance Capitalism », *Media and Communication*, 5(1), 2017

Kuusikko, S. et al. « Emotion Recognition in Children and Adolescents with Autism Spectrum Disorders » *Journal of Autism and Developmental Disorder,* 39, 938–945, 2009, https://doi.org/10.1007/s10803-009-0700

Le Breton, D. « Sociologie des émotions : Critique de la raison darwinienne », *Recherches sociologique*, 1, 1998, https://sharepoint.uclouvain.be/sites/rsa/Articles/1998-XXIX-1_05.pdf

Leseur. A. « Les théories de la Justice », *Centre national de la recherche scientifique*, Cahier no 2005-009, École Polytechnique, Paris, 2005, https://hal.archives-ouvertes.fr/hal-00242968/document#:~:text=R%C3%A9sum%C3%A9%3A,plus%20proc%C3%A9durales%20de%20la%20justice

Lewis, T. « AI can read your emotions. Should it? », *The Gardian*, 2019, https://www.theguardian.com/technology/2019/aug/17/emotion-ai-artificial-intelligence-mood-realeyes-amazon-facebook-emotient (Page consultée le 17 mars 2021)

Li, J., Wang, L. et Fischer, K. « The organisation of Chinese shame concepts? », *Cognition and Emotion,* 18(6), 767–797, 2004, https://www.tandfonline.com/doi/abs/10.1080/02699930341000202

Light, F. « Russia Is Building One of the World's Largest Facial Recognition Networks, *The Moscow Times*, 2019, https://www.themoscowtimes.com/2019/11/12/russia-building-one-of-worlds-largest-facial-recognition-networks-a68139 (Page consultée le 19 mars 2021)

Lighthill, M. J. *Artificial Intelligence : A General Survey*. In Science Research Council, *Artificial Intelligence: A Paper Symposium*, London, SRC, 1-21, 1973.

Martinez, F. « L'individu face au risque : l'apport de Kahneman et Tversky », *Idées économiques et sociales*, 3(3), 15-23, 2010, https://doi.org/10.3917/idee.161.0015

McStay, A. « Emotional AI, Soft Biometric and the Surveillance of Emotional Life: an Unusual consensus on privacy », Big Data & Society, *Journal Sage Pub*, 2020

McStay, A. « Empathic Media: The Rise of Emotion AI », *Arts & Humanities Research Concil*, 2016.

Merlier, P. « Bienveillance, bienfaisance, bientraitance », *Philosophie et éthique en travail social. Manuel*, sous la direction de Merlier P. « Politiques et interventions sociales », Presses de

l'EHESP, France, 45-49, 2013, https://www.cairn.info/philosophie-et-ethique-en-travail-social--9782810901326-page-45.htm

Miller, D. « Justice », *Stanford Encyclopedia of Philosophy*, 2017, https://plato.stanford.edu/entries/justice/

Monino J.-L. et Sedkaoui, S. « Big Data, Open Data et valorisation des données », *ISTE Édition*, 4, Londres, 2016

Mou, X. « Artificial Intelligence: Investment Trends and Selected Industry Uses », *International Finance Corporation*, 2019, https://www.ifc.org/wps/wcm/connect/7898d957-69b5-4727-9226-277e8ae28711/EM Compass-Note-71-AI-Investment Trends.pdf?MOD=AJPERES&CVID=mR5Jvd6

Nagel, T. « What Is It Like to Be a Bat? », *The Philosophical Review*, 83(4), 435-450, 1974.

Nations Unies. « Déclaration universelle des droits de l'homme », *Nations Unies*, s.d., https://www.un.org/fr/universal-declaration-human-rights/index.html (Page consultée le 4 avril 2021)

NITRD. « Supplement to the President's FY2020 Budget », *The Networking and Information Technology Research and Development Program*, 2019, https://www.nitrd.gov/pubs/FY2020-NITRD-Supplement.pdf#page=17

O'Neil, C. *Weapons of math destruction*, Broadway Books, New York, 2017

Pennisi, P. et al. « Autism and Social Robotics: A Systematic Review », *Autism Research*, Volume 9 (2), 165-183, 2015, https://doi.org/10.1002/aur.1527

Perrigo, B. « India Has Been Collecting Eye Scans and Fingerprint Records From Every Citizen. Here's What to Know », *Time*, 2019, https://time.com/5409604/india-aadhaar-supreme-court/

Pichon, S. et Vuilleumier, P. « Imagerie et cognition, Neuro-imagerie et neuroscience des émotions », *Médecine/Sciences*, 27(8-9), 763-769, 2011, https://medweb4.unige.ch/labnic/papers/SP_PV_MedSci2011.pdf

Pu, L., Moyle, W., et Jones, C. « How people with dementia perceive a therapeutic robot called PARO in relation to their pain and mood: A qualitative study», *Journal of Clinical Nursing*, 29(3-4), 437-446, 2020, https://doi.org/10.1111/jocn.15104

Quaglioni, D. *À une déesse inconnue : La conception pré-moderne de la justice,* Chapitre 2, Le droit civil, Éditions de la Sorbonne, 23-31, 10.4000/books.psorbonne.19906

Rachael E. J. et al. « Facial expressions of emotion are not culturally universal », *Proceedings of the National Academy of Sciences of the United States of America*, 2012, https://doi.org/10.1073/pnas.1200155109

Rekacewicz, P. « Chine, une mosaïque d'ethnies », *Le monde diplomatique*, 85, 2006, https://www.monde-diplomatique.fr/cartes/chineethnies

Rhue, L. « Racial Influence on Automated Perceptions of Emotions », *SSRN*, 2018, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3281765

Roger, K. et al. « Social commitment Robots and Dementia », *Canadian Journal on Aging*, 31(1), 87-94, 2012, https://muse.jhu.edu/article/468572

Russell, J. A. « Culture and the categorization of Emotions », *Psychological Bulletin*, 110(3), 426-450, 1991, https://pubmed.ncbi.nlm.nih.gov/1758918/

Russell, J.A. « Is There Universal Recognition of Emotion from Facial Expression? A Review of the Cross-Cultural Studies », *Psychological Bulletin*, 115(1), 102-141, 1994, https://doi.org/10.1037/0033-2909.115.1.102

Russell, J. A., et Barrett, L. F. « Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant », *Journal of Personality and Social Psychology*, 76(5), 805-819, 1999, https://psycnet.apa.org/doi/10.1037/0022-3514.76.5.805

Sahin, NT. et al. « Second Version of Google Glass as a Wearable Socio-Affective Aid: Positive School Desirability, High Usability, and Theoretical Framework in a Sample of Children with Autism », *JMIR Human Factors,* 5(1), 2018, https://humanfactors.jmir.org/2018/1/e1/

Sambuc, C. et Le Coz, P. « La dignité humaine kantienne : une justification théorique des transplantations d'organes ? », *Raison publique*, 2(2), 219-238, 2012, https://doi.org/10.3917/rpub.017.0219

Santé Log. « Autisme : Pourquoi le contact visuel est difficile », *Santé Log*, 2017, https://www.santelog.com/actualites/autisme-pourquoi-le-contact-visuel-est-difficile (Page consultée le 4 mai 2021)

Sarwari, K. « You Think You Can Read Facial Expressions? You're Wrong », *Northeasteran University Media*, 2019, https://news.northeastern.edu/2019/07/19/northeastern-university-professor-says-we-cant-gauge-emotions-from-facial-expressions-alone/ (Page consultée le 4 avril 2021)

Schenk, F. « Les émotions de la raison », *Revue européenne des sciences sociales*, XLVII-144, 151-162, 2009, https://journals.openedition.org/ress/75#citedby

Schroff, F., Kalenichenko, D. et Philbin, J. « FaceNet: A Unified Embedding for Face Recognition and Clustering », *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 815-823, 2015, https://www.cv-

foundation.org/openaccess/content_cvpr_2015/html/Schroff_FaceNet_A_Unified_2015_CVPR_paper.html

Schwartz, O. « Don't Look Now: Why You Should Be Worried about Machines Reading Your Emotions » The Guardian, 2019, https://www.theguardian.com/technology/2019/mar/06/facial-recognition-software-emotional-science (Page consultée le 24 février 2021)

Taigman, Y., Yang, M., Ranzato, M. et Wolf, L. « DeepFace: Closing the Gap to Human-Level Performance in Face Verification », *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 1701-1708, 2014, https://www.cv-foundation.org/openaccess/content_cvpr_2014/papers/Taigman_DeepFace_Closing_the_2014_CVPR_paper.pdf.

Tcherkassof, A. « Le sens dessus dessous des expressions faciales d'émotions : vers un nouveau tournant paradigmatique », *HAL*, Université Grenoble Alpes, France, 2018, https://hal.archives-ouvertes.fr/tel-01868279/document

The History of Computing. « Logic Theorist – Complete History of the Logic Theorist Program », The History of Computing, 2021, https://history-computer.com/logic-theorist-complete-history-of-the-logic-theorist-program/ (Page consultee le 11 juin 2021)

Thomas, D. « The Cameras that Know if You're Happy – or a Threat », *BBC News*, 2018, https://www.bbc.com/news/business-44799239

Tracy, J. L. et Robins, R.W. « Show your pride: Evidence for a discrete emotion expression », *Psychological Science*, 15(3), 194–197, 2004 ; Ven, R. « Choose How You Feel; You Have Seven Options », *Institute of Network Cultures*, Amsterdam, 2017, https://networkcultures.org/longform/2017/01/25/choose-how-you-feel-you-have-seven-options/

Usabilis. « Définition biais cognitifs », Usabilis, 2018, https://www.usabilis.com/definition-biais-cognitifs/ (Page consultée e 1er avril 2021)

Vaak. « Automatiser les opérations avec l'IA d'analyse », VAAK, https://vaak.co/ (Page consultée le 4 avril 2021)

Vaysse, J-M. *Dictionnaire Kant*. Éditions Ellipse Marketing, France, 2007

Wierzbicka, A. « Emotions across Languages and Cultures: Diversity and Universals », Studies in Emotion and Social Interaction, *Cambridge University Press*, Cambridge, 1999, https://www.cambridge.org/core/books/emotions-across-languages-and-cultures/7C03D03C6DF34ACBD7155B6555381715

Zanin, E. « Lire pour apprendre à aimer : la littérature comme philosophie morale », *Acta fabula*, 13(3), 2012, https://www.fabula.org/revue/document6875.php

Zittrain, J. « Facebook Could Decide an Election Without Anyone Ever Finding Out ». *The New Republic*, 2014

Annexe A : Classification des affects

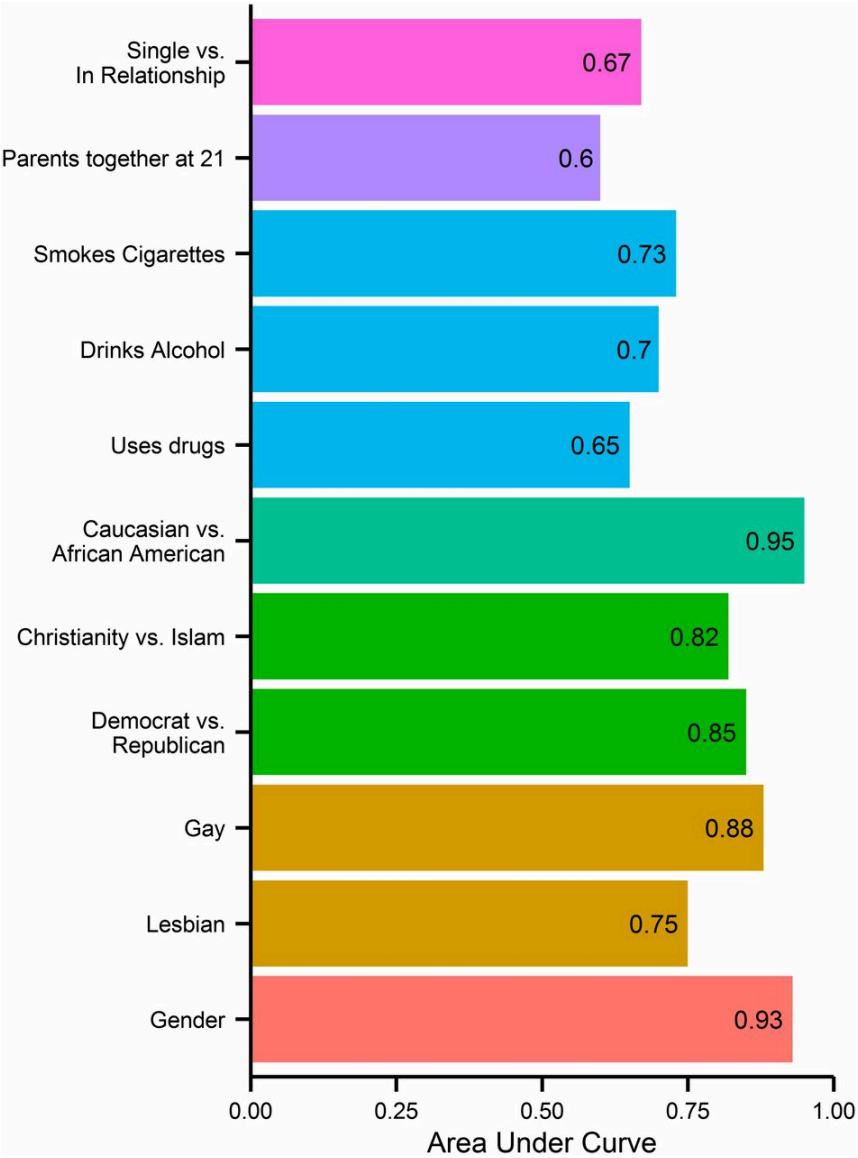
Titre du tableau[285] : Proposition d'une typologie de l'affectif

Caractéristiques / Types d'affects	Emotion (choc)	Sentiment	Humeur	Tempérament	Préférence	Attitude	Appréciation
EXEMPLES	Surprise Peur	Fierté Jalousie	Mélancolie Gaieté	Optimisme Pessimisme	Classement de stimuli	Prédispositions	Evaluation de candidats
CIBLE ou STIMULUS	Stimulus spécifique	Stimulus spécifique	Cible et stimulus non spécifiques	Cible et stimulus non spécifiques	Cible spécifique	Cible spécifique	Cible spécifique
INTENSITE SOMATIQUE (ET AUTONOMIQUE)	Forte	Moyenne	Faible à moyenne	Faible à moyenne	Moyenne (fonction des cibles)	Faible	Faible
DURABILITE	Très brève	moyenne	moyenne	longue	moyenne	moyenne	brève
FREQUENCE DES EXPERIENCES SOMATIQUES (ET AUTONOMIQUES)	Toujours	Parfois à souvent	Parfois à souvent	Souvent	Parfois	Rarement	Rarement
FREQUENCE D'EXPRESSION SOMATIQUE	Souvent	Fonction des contraintes sociales	Fonction des contraintes sociales	Souvent	Parfois	Rarement	Rarement
VOLONTE DE CONTROLE DE L'EXPRESSION (TROMPERIE)	Peut être forte surtout si les états affectifs sont négatifs	Peut être forte que les sentiments soient positifs ou negatifs	Peut être forte surtout si les états affectifs sont négatifs	Assez faible	Peut être forte en fonction des cibles	Peut être forte	Peut être forte
POSSIBILITE (FACILITE) DE CONTROLE DE L'EXPRESSION (TROMPERIE)	Faible	Assez faible à cause de la permanence[1] et/ou de l'intensité	Assez faible à cause de la permanence[1]	Assez faible à cause de la permanence[1]	Peut être faible si les cibles sont importantes et/ou si les situations sont permanentes	Elevée	Elevée
PROBABILITE D'EXPERIENCE SUBJECTIVE ELEMENTAIRE	Elevée	Assez élevée	Faible	Moyenne	Forte (conscience)	Moyenne	Forte (conscience)
IMPORTANCE DES ANTECEDENTS COGNITIFS (AMONT)	Très faible	Forte	Moyenne	Moyenne	Faible à moyenne	Forte	Très forte
PROCESSUS COGNITIFS AVAL	Parfois à souvent	Souvent	Souvent[2] (sous forme de renforcement et de justification)	Souvent[2]	Souvent	Importants en cas de dissonance cognitive	Souvent (et corrélés avec les processus cognitifs amont)

[285] Derbaix, C. et Pham, M. T. (1989). « Pour un développement des mesures de l'affectif en marketing : synthèse des prérequis ». Recherche et Applications en Marketing, SAGE, Association française du marketing, 4(4), https://www-jstor-org.acces.bibl.ulaval.ca/stable/pdf/40588767.pdf?ab_segments=0%252Fbasic_search%252Fcontrol&refreqid=excelsior%3A4659bdda8ad2106fcdc2b8a4b08c2b63, p. 78.

Annexe B : Les attributs privés

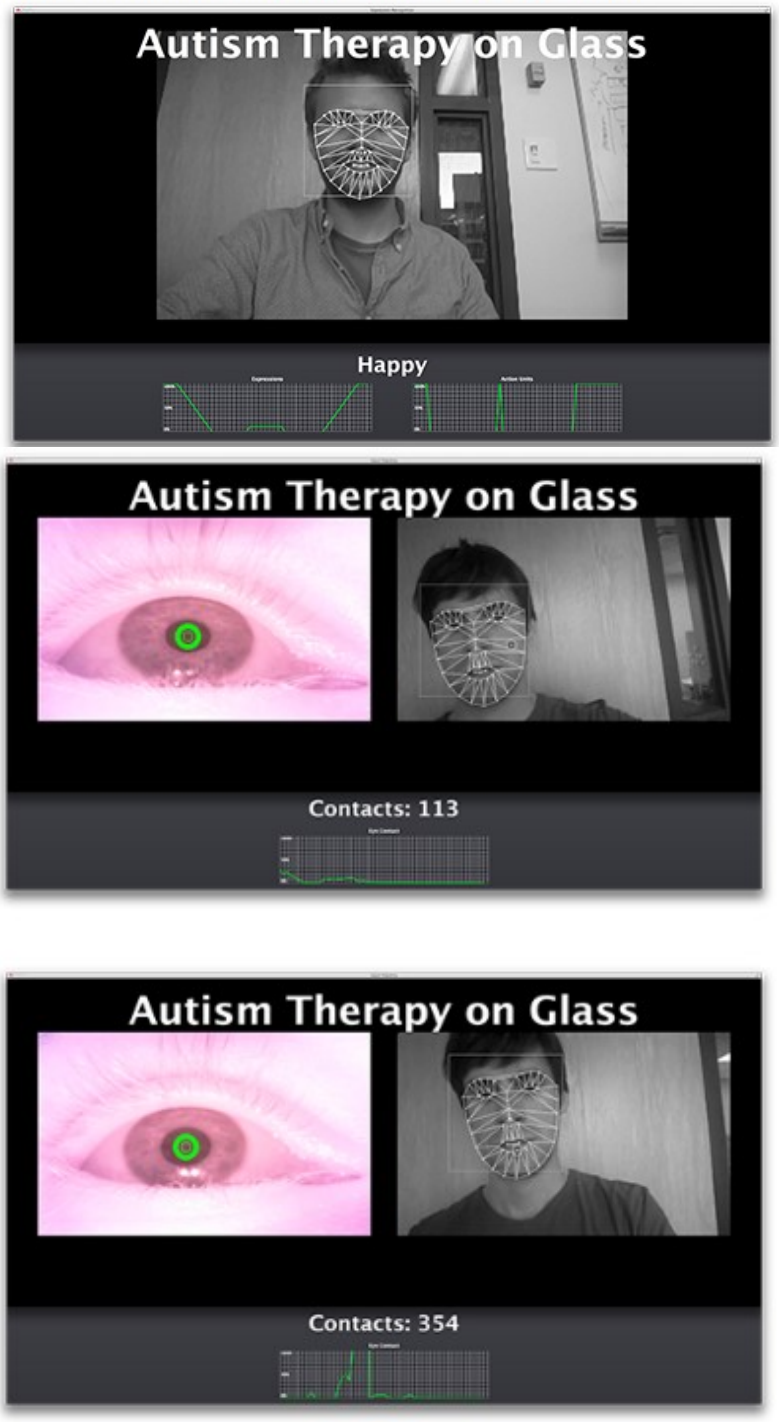
Titre du tableau[286] : Prediction accuracy of classification for dichotomous/dichotomized attributes expressed by the AUC.



[286] Kosinski, M., Stillwell, D. et Graepel,T. (2013) « Private traits and attributes are predictable from digital records of human behavior », National Academy of Sciences, pp. 5802-5805. URL : https://www.pnas.org/content/pnas/110/15/5802.full.pdf?3=.

Annexe C : Reconnaissance des expressions faciales émotionnelles[287]

Images below demonstrating the usage of the application:



[287] « Autism Therapy on Glass », https://wall-lab.stanford.edu/projects/autism-therapy-on-glass/.