

Human–Autonomy Teaming (HAT) – Équipe entre humain et autonomie : définitions, débats et orientations

Human–Autonomy Teaming: Definitions, Debates, and Directions

Joseph B. Lyons¹, Katia Sycara², Michael Lewis³, August Capiola¹

¹ Air Force Research Laboratory, Dayton, Ohio, USA

² School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

³ School of Computing and Information, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Traduction de l'article original de LYONS B.L., SYCARA, K., LEWIS M. "Human–Autonomy Teaming: Definitions, Debates, and Directions ", *Frontiers in Psychology*, vol.12, art.589585, 2021. – doi:10.3389/fpsyg.2021.589585, par Bernard CLAVERIE et Baptiste PREBOT du comité éditorial de la revue ISTE – Ingénierie Cognitive.

Avertissement : Cet article est publié avec l'accord des auteurs, selon les principes du libre accès (open science). Certaines expressions ont dû être adaptées à la Langue française ; les bibliographies ont été complétées et revues selon les normes de la revue ISTE – Ingénierie Cognitive.

RÉSUMÉ. Les auteurs définissent le passage de l'étude de l'interaction homme-automatisation au HAT, équipe homme-autonomie. Cette distinction a été mise en évidence dans la littérature récente, et les raisons théoriques pour lesquelles l'expérience psychologique des humains interagissant avec l'autonomie peut varier et affecter les résultats ultérieurs de la collaboration commencent à émerger. Dans cette revue, les auteurs plongent dans les systèmes HAT en expliquant les différences entre l'automatisation et l'autonomie et en passant en revue le domaine de la collaboration entre humains pour faire des inférences pour les HAT. Les auteurs examinent le domaine de cette collaboration pour extrapoler quelques facteurs fondamentaux qui pourraient être pertinents pour les HAT. Ces facteurs impliquent notamment des éléments sociaux critiques au sein des équipes qui occupent une place centrale (comme le montre cette étude) pour les HATs. Les auteurs concluent en soulignant certaines lacunes de la recherche auxquelles les chercheurs devraient s'efforcer de répondre, ce qui devrait faciliter une compréhension plus nuancée et complète des HAT dans une diversité de contextes du monde réel.

ABSTRACT. Researchers are beginning to transition from studying human–automation interaction to human–autonomy teaming. This distinction has been highlighted in recent literature, and theoretical reasons why the psychological experience of humans interacting with autonomy may vary and affect subsequent collaboration outcomes are beginning to emerge. In this review, we do a deep dive into human-autonomy teams (HATs) by explaining the differences between automation and autonomy and by reviewing the domain of human–human teaming to make inferences for HATs. We examine the domain of human–human teaming to extrapolate a few core factors that could have relevance for HATs. Notably, these factors involve critical social elements within teams that are central (as argued in this review) for HATs. We conclude by highlighting some research gaps that researchers should strive toward answering, which will ultimately facilitate a more nuanced and complete understanding of HATs in a variety of real-world contexts.

MOTS-CLÉS. Autonomie, Collaboration humain-système, Équipe, Équipe autonomie-humain, Facteurs humains, Intégration human-système, Robotique, Systèmes anthropotechniques.

KEYWORDS. Anthropotechnical systems, Autonomy, Human factors, Human–autonomy team, Human-system collaboration, Human-system integration, Robotics, Team.

1. Contexte de l'étude

La collaboration en équipe constitue depuis longtemps l'un des atouts précieux qui aident les organisations à atteindre des objectifs dynamiques et complexes. Ces dernières années, dans les domaines des facteurs humains, de l'informatique et de la robotique, la communauté scientifique a cherché à étendre ce concept de collaboration au partenariat entre les humains et les machines. Les progrès de l'informatique au cours des dernières décennies ont permis aux chercheurs d'envisager

l'utilisation de « machines intelligentes » pour remplir des rôles importants (voire vitaux) au sein d'équipes de collaborateurs humains [SYC 04]. Les « agents synthétiques » sont de plus en plus utilisés pour augmenter les capacités humaines et pour s'associer aux opérateurs dans des tâches réelles [DEM 17], ce qui nous amène à penser qu'en tant que communauté humaine, nous sommes à l'aube d'avancées technologiques ayant un retentissement philosophique concernant la nature du travail humain moderne et la façon dont il est accompli.

Bien qu'un travail de recherche considérable ait été effectué sur la collaboration dans les équipes humaines (composées uniquement d'humains) au cours des dernières décennies, la littérature sur les équipes hybrides, c'est-à-dire les « équipes humain-autonomie » (EHA) ou *Human-Autonomy Teaming* en Anglais (HAT), est moins mature, bien qu'elle se soit développée ces dernières années. Les technologues ont cependant accumulé plus d'un siècle d'expérience sur l'automatisation des machines qui intègrent maintenant de plus en plus d'« intelligence ». Malgré la grande sophistication de beaucoup de ces machines, c'est par exemple le cas du pilotage automatique (PA) avancé, nos relations avec elles restent généralement inchangées. Ainsi, bien que la sécurité du pilote humain dépende du fait que le PA prenne des décisions sophistiquées en matière de stabilité du vol dans des conditions difficiles, ce pilote est susceptible de considérer le PA comme faisant partie des commandes de l'avion, par exemple en définissant le cap ou l'altitude comme l'on pourrait choisir l'accès à un étage dans un ascenseur automatisé. Le même pilote aurait cependant probablement plus d'inhibition si on lui demandait de réaliser une mission dangereuse dans laquelle son ailier serait remplacé par un drone autonome. Pourtant, dans les deux situations, le sort et la sécurité du pilote sont confiés à une machine décisionnelle sophistiquée, mais le cas du pilote automatique est accepté sans inhibition alors qu'un ailier artificiel autonome suscite de l'incertitude et de la peur. Nous pensons que cette distinction est qualitative et bien réelle, elle implique à la fois le degré d'autonomie, la prédisposition évolutive des humains à la coopération [BOW 11] [TOM 12] et des formes particulières de traitement neuronal associées à la coopération humaine [MUY 18] [XIE 20]. Le pilote automatique réalise une fonction précise et prévisible que le pilote peut facilement contrôler alors qu'en revanche le fait de dépendre d'un ailier autonome oblige le pilote à céder le contrôle en espérant que l'ailier "assurera ses arrières" tout au long de la mission, ce qui soulève des inquiétudes quant à l'absence d'un « penseur humain » dans le cockpit [LYO 18]. Cela soulève des questions au-delà des simples aspects technologiques de la machine – cette machine m'aidera-t-elle quand j'en aurai besoin ? – avons-nous la même compréhension des menaces ? – pouvons-nous communiquer suffisamment pour gérer la dynamique des tâches ? Nous soutenons que les deux interactions mobilisent des modes distincts de comportement humain, la confiance dans un pilote automatique dépendant des formes normales d'interaction alors que la coopération étroite (et éventuellement la collaboration) avec l'ailier nécessite de faire appel à des comportements sociaux qui ont évolué au cours de l'histoire de notre espèce, comme cela s'est passé en matière de recherche de la nourriture en groupe [TOM 12] ou en matière de guerre [BOW 11]. Dans les interactions conventionnelles, une machine est traitée comme un outil, mais dans des interactions hautement relationnelles et dans des conditions particulières, elle peut en venir à jouer le rôle d'un véritable coéquipier, en bénéficiant de la fluidité et de la cohérence renforcée d'actions conjointes rendues possibles par la tendance humaine à la coopération sociale.

La notion de partage d'autorité entre les humains et les machines pour la poursuite d'objectifs communs est un sujet de recherche majeur des communautés des facteurs humains et de l'informatique [CUM 14] [FLE 19]. Ce sujet a pris de l'ampleur ces dernières années en raison de son importance pour le domaine de la voiture autonome [LIJ 16] [AWA 18] et dans la recherche militaire [CHE 14]. Certains auteurs ont identifié trois rôles principaux que les machines peuvent assumer au sein des équipes [SYC 04]. Ainsi elles peuvent soutenir les équipes humaines (1) en aidant les individus dans l'accomplissement de leurs tâches individuelles, (2) en assumant le rôle d'un des membres de l'équipe, ou (3) en soutenant l'équipe dans son ensemble. À ce jour, le premier de ces rôles a été étudié dans le contexte des systèmes d'aide à la décision ; le deuxième et le troisième rôles sont au cœur des HAT, et font l'objet de recherches de plus en plus nombreuses. Certains travaux récents considèrent ainsi les agents artificiels comme remplaçants d'un humain au sein d'une équipe humain-humain, en quelque sorte comme un humain synthétique [MCN 18]. D'autres ont examiné les questions relatives aux

processus d'équipe telles que l'évolution du renforcement de la confiance et l'adaptation dans des environnements complexes et dynamiques [NAM 20]. En ce qui concerne l'identification d'éléments caractéristiques des coéquipiers-machines que les humains pourraient trouver convaincants, certaines recherches ont examiné les facteurs psychologiques et les conséquences de la confrontation à une technologie en tant que coéquipier plutôt qu'en tant qu'outil (e.g. [WYN 18]). Cette dernière question est cruciale car elle éclaire sur les caractéristiques souhaitables que doit présenter une machine en tant que coéquipier et sur les différences entre l'automatisation (outil) et l'autonomie (coéquipier potentiel). Ainsi, le présent article examine les HAT dans l'une des trois typologies suivantes : (1) comme moyen de révéler les facteurs psychologiques qui entrent en jeu dans la perception des HAT, (2) comme substitut d'un coéquipier humain et membre communicant de l'équipe, et (3) comme acteur d'influence sur la dynamique d'équipe considérée dans son ensemble.

2. Pourquoi avons-nous besoin des HAT ?

Il est généralement admis que les équipes humain-autonomie possèdent des performances améliorées par rapport aux humains seuls ou aux machines seules, en particulier dans les contextes de forte incertitude [CUM 14]. Un exemple connu est celui de joueurs d'échecs novices qui peuvent s'associer à un agent intelligent et l'emporter sur des grands maîtres [KAS 10]. Les systèmes artificiels autonomes ont souvent un pouvoir de calcul qui surclasse, en termes de puissance et de rapidité, les capacités humaines et possèdent également des capteurs plus efficaces que les humains [ARK 09] [SCH 18], ce qui peut accroître la vitesse de traitement des données et l'étendue de leurs analyses. Lorsque des systèmes autonomes sont utilisés en collaboration avec des humains, ils peuvent servir de démultiplicateur et, en théorie, nécessiter moins de soldats pour exécuter leur mission [ARK 09] [END 15] [SCH 18]. Les HAT peuvent s'avérer particulièrement utiles dans des missions ouvertes, celles dont tous les paramètres ne peuvent pas être spécifiés *a priori* [CHE 14]. Cependant, les avantages potentiels des HAT ne se limitent pas au domaine militaire.

Les équipes humain-autonomie peuvent accroître la sécurité psychologique lorsqu'il s'agit de contribuer à l'information, faire des suppositions, lancer des idées ou signaler des informations douteuses pour les humains. La sécurité psychologique fait référence à « la conviction partagée, lorsqu'il s'agit de prendre des risques interpersonnels [...] qu'aucun membre de l'équipe ne mettra mal à l'aise, ne rejettera ou n'en sanctionnera un autre pour s'être exprimé » [EDM 99]. Par ailleurs, les systèmes autonomes ne jugent pas les autres, contrairement aux humains. Certaines recherches prometteuses explorent d'ailleurs l'utilisation de robots pour aider de jeunes autistes [DIE 13] ou à soutenir les soldats souffrant d'un syndrome de stress post-traumatique (SSPT – *Post traumatic Stress Disorder*) grâce à des agents virtuels intelligents [KAN 10].

3. Automation versus Autonomie

Les besoins en matière de HAT découlent de situations nécessitant des interactions étroitement coordonnées entre les humains et les machines. Lorsque les actions des machines peuvent être prédites avec précision, un contexte social n'est pas forcément nécessaire pour faciliter l'interaction, comme par exemple pour l'augmentation de stabilité d'un système de contrôle. En revanche, lorsque les conditions déterminant les actions d'un partenaire ne sont pas évidentes, la connaissance des capacités de la machine, de ses objectifs (buts) ou d'autres caractéristiques peut être nécessaire pour établir la base d'une coopération. Par exemple, le respect des règles a doublé lorsqu'on a montré aux participants d'une étude pourquoi un planificateur autonome de trajectoire voulait réorienter un UAV (*Unmanned Aerial Vehicle* – drone), tout en s'affranchissant de la corrélation antérieure entre confiance et respect des règles [CHI 20].

Parce que les machines sont des entités déterministes et algorithmiques, la distinction entre automation et autonomie est subjective. Cela constitue la source de l'un des principaux débats concernant l'association humain-autonomie. L'automation correspond à une technologie de recherche

active des données, de transformation des informations, de prise de décision ou de contrôle dans des tâches limitées et bien définies pour lesquelles le processus automatisé peut être préprogrammé [LEE 04]. En d'autres termes, l'automation consiste à faire quelque chose à la place d'un humain qui n'a donc plus besoin de le faire. Certes, le recours à l'automation n'élimine pas le besoin de présence d'un être humain, mais il modifie son rôle en superviseur, ce qui s'accompagne d'une série de difficultés potentielles [PAR 97]. En effet, des décennies de recherche ont montré que l'utilisation de l'automation rajoute souvent une complexité considérable pour ceux qui doivent alors surveiller la technologie [PAR 97] [LEE 04] [HOF 15].

L'automatisation effectue par principe une tâche dans les limites de ce pour quoi elle a été programmée, ce qui la rend inutilisable en dehors de ce cadre. Dans les conditions normales, l'automation joue un rôle défini, et son utilisation peut être associée à une augmentation des performances et une réduction de la charge de travail des opérateurs humains [ONN 14]. La littérature sur l'automation a permis de développer une taxonomie des types de comportements dans lesquels elle peut être engagée (par exemple, l'acquisition d'information et son analyse, la sélection de décisions et d'actions, et leur mise en œuvre [PAR 20]) ainsi que le niveau d'intervention humaine acceptable pour une réponse automatisée particulière (c'est-à-dire les niveaux d'automation [SHE 78]). Une classe particulièrement remarquable d'automation correspond à l'automatisation adaptative qui fait varier son comportement selon la détection d'un changement de l'état d'un humain (tels que des changements physiologiques révélateurs d'une charge de travail trop élevée ou trop faible [HAN 13] [CHE 14]). Dans ce cas, un automate peut être programmé avec un seuil déterminé pour détecter sur une échelle continue les changements dynamiques d'une variable cible afin de produire une réponse automatisée elle aussi prédéfinie. Par conséquent, l'automation peut se voir déléguer des niveaux plus ou moins élevés de latitude dans l'exécution d'une tâche et présenter une forme de flexibilité d'exécution dans un contexte spécifique, mais elle n'a pas le pouvoir de décider elle-même dans quels contextes ou quelles conditions elle doit fonctionner, puisque cela est défini *a priori* par le concepteur ou par l'opérateur de l'automate. Étant donné qu'elle produit des comportements reproductibles dans des domaines établis, la prévisibilité devient un élément important de confiance dans la technologie [LEE 04] [HOF 15]. Plus précisément, l'automation devrait être fiable grâce à ses processus algorithmiques sous-jacents qui sont contextuellement appropriés et mis en œuvre selon les prévisions de son concepteur [LEE 04]. Ainsi pour les tâches relativement simples, invariantes, reproductibles et à risque faible, elle peut présenter de vrais bénéfices, alors que pour des tâches complexes, fluides, dépendantes du contexte et à risque élevé, son utilisation peut perdre tout intérêt ou même conduire à des résultats indésirables en raison de l'incapacité intrinsèque de l'automate à gérer des situations en dehors de celles pour lesquelles il a été programmée.

L'autonomie, en revanche, est capable de prendre des décisions indépendamment du contrôle humain [VAG 16]. L'autonomie peut ainsi répondre à des situations pour lesquelles elle n'a pas été conçue (c'est-à-dire que l'autonomie est capable d'apprendre et de généraliser) et elle devrait posséder une certaine capacité à diriger ses propres actions, c'est-à-dire être orientée vers un but [END 15]. L'autonomie est caractérisée par la prise de décision, par l'adaptation à l'évolution des besoins et par l'amélioration des performances au fil du temps avec un certain niveau d'autogestion [MAS 13]. Cette capacité d'autogestion, d'adaptabilité et d'apprentissage est l'une des caractéristiques importantes dont l'autonomie bénéficie par rapport à l'automation. Une définition convaincante, se référant au travail collaboratif, dispose que « l'autonomie est une technologie capable de travailler avec des êtres humains en tant que coéquipiers, y compris pour des tâches essentielles au travail d'équipe qu'on peut attendre de tout coéquipier humain » [MCN 18]. Dans ce cas, ses principales caractéristiques consistent à pouvoir assurer un rôle de coéquipier et à pouvoir communiquer avec ses coéquipiers. Compte tenu de l'importance accordée au travail d'équipe et à la communication, les HAT ont besoin d'informations sur les motivations et les modèles mentaux partagés, et de moyens sociaux pour permettre la communication et une compréhension commune. Ce sont ces mêmes caractéristiques qui se retrouvent dans la littérature portant sur les équipes interpersonnelles.

4. Que pouvons-nous apprendre des équipes interpersonnelles ?

Au cours des dernières décennies, la communauté scientifique a mis l'accent sur le fonctionnement collaboratif dans les organisations, et cela a été en grande partie dû au fait que ces organisations utilisent justement des équipes de petite taille pour accomplir toute une variété de tâches. De telles équipes sont souvent constituées pour aider à répondre au besoin d'adaptabilité, car elles sont plus plastiques que de grandes unités organisationnelles [KOZ 03]. De la même manière, le recours aux HAT est censé permettre une plus grande adaptabilité aux situations dynamiques par le biais d'interactions homme-automate. Il existe de nombreux modèles concernant l'approche scientifique des petites équipes. Dans la littérature en science de gestion et des organisations, elles sont définies comme : « deux individus ou plus qui exécutent des tâches pertinentes pour l'organisation, partagent au moins un objectif commun, interagissent socialement, sont interdépendants, maintiennent et gèrent des limites et des rôles, et sont intégrés dans un contexte organisationnel plus large qui fixe les objectifs, les limites et les contraintes de l'équipe, et influence les échanges avec d'autres équipes » [KOZ 03]. Une telle équipe manifeste des caractéristiques comme l'ambition collective, le partage d'objectifs, l'alignement des objectifs individuels sur ses objectifs communs, la différenciation élevée des compétences entre les membres, la communication ouverte, la sécurité et l'engagement mutuel [VAN 18]. Alors que les équipes fonctionnent comme des unités de travail distinguables (et mesurables) partiellement indépendantes de l'organisation (mais intégrées dans celle-ci). Cette dernière limite et fixe les objectifs généraux de l'équipe [KOZ 03]. Ainsi, les HAT acquièrent les objectifs stratégiques généraux de leur organisation mère (tels que par exemple ceux du service plus large au sein duquel la HAT fonctionne). Par conséquent, il peut s'avérer vain de s'attendre à ce qu'une HAT évalue des objectifs à l'échelle générale, mais elles doivent tenir compte des objectifs proximaux au regard de la hiérarchie des objectifs plus large de l'organisation. Tout cela borne l'espace d'étude au contexte des tâches d'une organisation et évite de devoir discuter du concept d'intelligence artificielle générale (IA). Cependant, les équipes sont soumises à des caractéristiques et des dynamiques environnementales locales, ouvrant la voie à des facteurs d'influence de l'équipe sur les individus qui la composent [KOZ 00]. Ces facteurs, qui peuvent prendre la forme d'influences normatives, de perceptions partagées et d'attentes des individus, ont manifestement des conséquences à la fois descendantes et ascendantes. De tels dynamiques *top-down* et *bottom-up* se retrouvent probablement au sein des équipes humain-autonomie.

Le modèle le plus connu pour l'étude de la dynamique d'équipe est peut-être le modèle « Entrée-Processus-Sortie » (*Input-Process-Output* – IPO) [MCG 64]. Dans ce modèle, les entrées (sous forme de ressources, de contraintes, de compétences, de composition de groupe, etc.) influencent les processus d'équipe (mécanismes qui permettent aux membres de combiner leurs entrées) qui modulent la relation entre les entrées et les sorties. Les résultats des équipes constituent un élément clé pour toute organisation qui les emploie et ils varient évidemment en fonction des objectifs/stratégie de l'organisation. Cependant, les processus d'équipe sont ce qui transforme les entrées en sorties, et donc dans le contexte de cet article l'accent est mis sur les processus d'équipe qui pourraient influencer l'équipe humaine-autonomie. Ces processus comprennent la coordination, la coopération et la communication [TAN 92], et peuvent être utilisés pour mesurer et évaluer la façon dont les équipes traitent l'information, prennent des décisions et combinent des actions individuelles vers des objectifs communs – c'est-à-dire le processus de cognition de l'équipe [SAL 04]. Les processus d'équipe sont au cœur des équipes parce qu'ils (1) influencent le développement et la mise en œuvre de modèles mentaux partagés, (2) façonnent la motivation individuelle vers ou hors des objectifs de l'équipe (c'est-à-dire qu'ils façonnent l'intentionnalité de l'équipe et des individus) et (3) ont un impact sur la cohésion de l'équipe [KOZ 03].

4.1. L'intentionnalité d'équipe

Les membres efficaces d'une équipe partagent des objectifs avec leur équipe ; par conséquent, les actions qui témoignent de cet alignement des objectifs (c'est-à-dire l'intention) sont une composante importante du processus collectif. Un des risques inhérents au travail d'équipe est la possibilité qu'un

des membres ne soit pas aussi motivé que les autres pour atteindre les objectifs de l'équipe et qu'il adopte des comportements tels que le « flottement social ». Faire preuve de motivation pour poursuivre les objectifs de l'équipe (par opposition aux objectifs individuels) devrait promouvoir la confiance au sein de l'équipe [DIR 99], c'est-à-dire la volonté d'accepter la vulnérabilité basée sur l'attente d'un résultat positif [MAY 95]. Étant donné que la confiance d'un membre de l'équipe à l'égard d'un autre peut avoir une influence sur sa réciprocité [SER 05], une plus grande confiance devrait réduire le flottement social et favoriser une plus grande motivation à l'égard des objectifs de l'équipe. De plus, toute information fondée sur l'intention qui laisserait entendre qu'un coéquipier est sincère devrait accroître la confiance qu'il inspire [MAY 95] [SER 05].

L'intentionnalité peut également prendre la forme d'un désir de faire partie d'une équipe. La cohésion d'équipe est un processus essentiel, car elle permet le sentiment d'affiliation et la motivation conjointe à la fois au niveau social et pour l'accomplissement des tâches communes [KOZ 03]. Le lien établi entre cohésion et performance de l'équipe est réciproque. Par conséquent, l'amélioration de la cohésion de l'équipe augmente la performance qui, en retour, favorise une plus grande cohésion [KOZ 15]. L'étude de la cohésion en tant que concept multidimensionnel remonte aux travaux fondamentaux de Festinger [FES 50]. La cohésion y est décrite comme un point culminant de divers facteurs dont l'attraction pour les membres du groupe, ses activités (engagement dans la tâche) et le prestige (fierté du groupe). La cohésion se développe une fois que le groupe a eu l'occasion de travailler ensemble ou au moins de faire connaissance [GOS 89] [MAT 96] [HAR 98]. Une méta-analyse [BEA 03] a montré que les trois composantes originales de la cohésion de Festinger (*op.cit.*) – l'attraction interpersonnelle, l'engagement dans la tâche et la fierté de groupe – ont chacune des relations indépendantes significatives avec les performances dans de nombreuses catégories de critères. Qu'il s'agisse d'un être humain ou d'une machine, le fait de connaître les intentions de son partenaire est un facteur important de cohésion de l'équipe.

4.2. Modèles mentaux partagés

Les modèles mentaux partagés représentent des structures de connaissances qui peuvent être partagées entre les membres de l'équipe et qui permettent à ses membres de créer des représentations et des prédictions précises dans un contexte de tâche collaborative [CAN 93]. Ces représentations peuvent concerner l'équipement (c'est-à-dire les matériels), la tâche de l'équipe (procédures, contraintes), les membres de l'équipe (compétences, capacités, forces, faiblesses) et les interactions nécessaires au sein de l'équipe (rôles, attentes, dépendances) [CAN 93]. Les modèles mentaux partagés comptent pour l'efficacité de l'équipe puisqu'ils facilitent un cadre commun à partir duquel interpréter l'environnement et les progrès du groupe par rapport aux objectifs communs [COO 13]. Des travaux ont ainsi montré que les équipes sont plus performantes lorsqu'elles partagent des modèles mentaux [MAT 00]. Les équipes développent des modèles mentaux partagés par le biais d'un processus appelé "cognition d'équipe", à savoir un processus cognitif se déroulant au niveau collectif et dont la finalité est d'aboutir à une compréhension commune entre les membres de l'équipe [COO 13]. La communication entre les membres de l'équipe est le fondement de la cognition d'équipe. Elle est donc essentielle pour comprendre dans quelle mesure une équipe peut détecter des anomalies, s'y adapter et assurer un suivi au sein de l'équipe - autant d'éléments qui devraient avoir un impact sur les performances de l'équipe.

4.3. Communication

La communication est la pierre angulaire du travail d'équipe car c'est le mécanisme par lequel les équipes traduisent les actions et les efforts individuels en résultat collectif. Les équipes qui communiquent et surveillent la performance par rapport aux objectifs partagés surpassent les groupes qui n'ont pas d'objectifs communs et qui ne partagent pas l'intention d'atteindre ces objectifs [AUB 03]. Il ne suffit pas que les équipes partagent des objectifs, mais l'intention de soutenir des objectifs communs est un élément essentiel du processus d'équipe qui aide à orienter les contributions individuelles vers les objectifs collectifs. La communication est ici nécessaire pour transmettre cette intention commune et pour évaluer le suivi des performances. Les équipes très performantes ont

tendance à être plus efficaces dans leur utilisation des questions, posant moins de questions tout en recevant toutes les informations nécessaires [URB 93]. Les équipes performantes adoptent également des comportements tels que l'évaluation de la situation et la planification, qui contribuent à atteindre et à maintenir la conscience de la situation [ORA 90]. En d'autres termes, les équipes efficaces doivent collectivement percevoir l'environnement, le comprendre par rapport à des objectifs partagés et agir conformément à ces objectifs en soutenant les autres membres de l'équipe. L'une des caractéristiques d'une équipe efficace est la transmission de l'information à un coéquipier avant qu'il n'en ait besoin [DEM 17], de sorte que si l'équipe est bien coordonnée et communique efficacement, il n'est pas vraiment nécessaire d'obtenir des informations de ses coéquipiers. Il a été observé que les équipes dans lesquelles les membres fournissent ces informations non sollicitées à d'autres membres de l'équipe sont généralement plus performantes que celles qui ne le font pas [URB 93] [JOH 94].

En résumé, les processus d'équipe représentent le liant qui traduit les entrées de l'équipe en sorties collectives. Les processus d'équipe permettent une attention conjointe, des objectifs partagés et une motivation pour atteindre ces objectifs, ainsi qu'un sentiment d'attraction sociale pour les membres de l'équipe notamment dans un contexte de tâche définie à remplir. Des processus d'équipe efficaces peuvent (1) montrer l'intention partagée d'atteindre les objectifs collectifs, (2) promouvoir la cognition d'équipe pour soutenir le développement et le maintien de modèles mentaux partagés, et (3) promouvoir l'aide et le contrôle des performances par le biais de la communication. Ces mêmes notions apprises par la littérature d'équipe homme-homme s'appliqueront également et probablement aux HAT.

5. Définition des équipes humain-autonomie (HAT)

Dans l'esprit des travaux pionniers de Nass et al. [NAS 94] qui considéraient que « fournir à l'ordinateur des caractéristiques associées aux humains » était un ingrédient essentiel pour susciter des attributions sociales, d'autres chercheurs considèrent que l'« humanité » est indépendante du degré d'autonomie [DEV 18] [HAS 18]. Ainsi, par exemple, un robot compagnon tel que le bébé phoque *Paros* ® [SAB 13] pourrait être bien placé pour susciter des réponses sociales tout en étant dépourvu d'autonomie. En revanche, un chariot automatisé dans un entrepôt peut être très autonome dans le choix des objectifs et la planification de l'optimisation de l'emplacement des stocks, tout en utilisant la prévention des collisions pour éviter les humains, et ce sans aucune relation sociale apparente. Les défis du HAT résident dans le développement (1) de moyens basés sur l'équipe pour favoriser la conscience partagée et la motivation collective, (2) d'une compréhension des types de tâches et d'interactions qui peuvent bénéficier de repères sociaux, et (3) du développement de techniques pour utiliser ces repères afin d'améliorer les performances de l'équipe humain-autonomie.

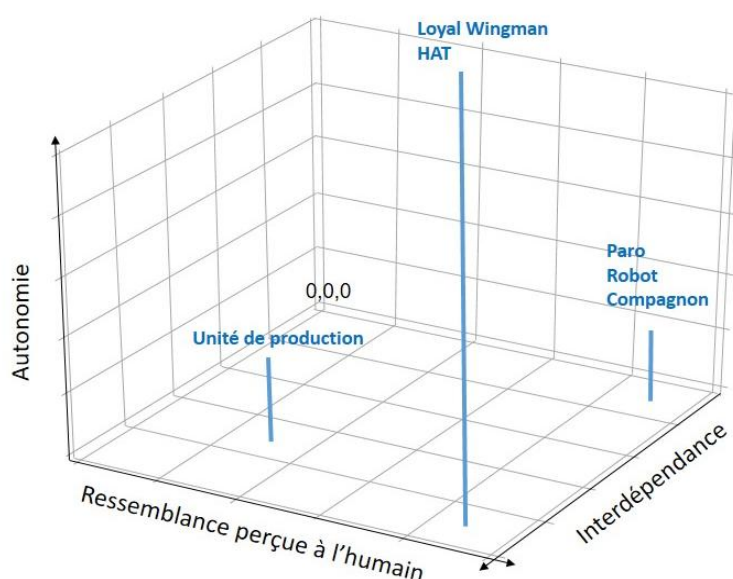


Figure 1. *Les propensions humaines à la coopération sociale devraient faciliter les performances homme-machine lorsque l'autonomie et l'interdépendance des machines sont élevées.*

Pour répondre à la question de savoir comment les IHM peuvent être transformées en HAT, nous devons d'abord essayer de comprendre les caractéristiques de la machine, de la tâche et des interactions qui amènent les humains à considérer et à traiter une machine comme un coéquipier, ainsi que les conditions dans lesquelles cela est censé améliorer les performances. La figure n°1 montre une relation hypothétique entre l'autonomie, l'interdépendance et la ressemblance humaine perçue. Le robot compagnon Paro engendre une forte perception de la ressemblance humaine, mais une faible perception de l'autonomie et de l'interdépendance. Les robots d'une unité de fabrication ont un niveau élevé d'interdépendance, car ils positionnent et échangent continuellement des pièces avec les humains, et un niveau modérément élevé d'autonomie. On peut s'attendre à ce que ces interactions bénéficient d'ajustements et de synchronisation positives en raison du système neurobiologique de neurones miroirs [PIN 09] mais ne devraient pas mobiliser d'attribution d'humanité. En revanche, le modèle d'un partenaire autonome tel que dans les HAT présente un niveau élevé dans les trois dimensions précédentes. Son degré d'autonomie est élevé et la forte interdépendance avec les autres membres de l'équipe nécessite une attribution d'humanité et de bienveillance pour que la coopération puisse se développer sans une surveillance continue et sans que l'humain ne remarque la situation d'hybridité.

On ne sait pas encore exactement quelles seraient les capacités technologiques qui rendraient le système acceptable pour les humains, bien que des recherches sur ce sujet aient été initiées [DEV 18]. La découverte de ces caractéristiques est cependant importante en robotique, où elles pourraient être des caractéristiques essentielles dans les algorithmes permettant de faire des robots des coéquipiers acceptables pour les humains. Cette perspective, en considérant la vision psychologique des machines en tant que membres de l'équipe, est nécessaire compte tenu de l'immense défi consistant à amener les humains à accepter les machines comme de vrais partenaires. « Aujourd'hui, dans presque tous les cas, le facteur limitant de l'interaction homme-agent n'est pas devenu les cycles de calcul ou la connectivité (côté machine), mais la capacité ou la volonté de l'utilisateur de communiquer ses désirs et de contrôler, d'organiser et d'interpréter la réponse de la machine pour les satisfaire (côté humain) » [SYC 04]. Bien que cette citation date de plus de 15 ans, peu de choses ont été réalisées à ce jour pour informer et aider les chercheurs sur ce qui motive les croyances portant sur un coéquipier artificiel. Toutefois, compte tenu des progrès récents de la puissance de calcul des machines, et ceux des capacités techniques et de l'omniprésence des systèmes robotiques, il est impératif que la communauté des chercheurs adopte des points de vue communs sur l'équipe humain-autonomie et s'engage dans une recherche intégrative pour pouvoir isoler et valider les différentes composantes d'une équipe HAT efficace.

Les chercheurs s'efforcent depuis longtemps de comprendre les interactions sociales entre les personnes et les machines. Le travail d'équipe est par définition social (dans le sens où il implique plusieurs individus en interaction), par opposition au travail posté qui peut être effectué par un seul individu. Si l'on considère le travail entre humain et autonomie sous l'angle du travail d'équipe et que l'on se demande quelles sont les caractéristiques qui font d'une machine un vrai coéquipier, on part du principe que la technologie accomplira correctement, à un degré acceptable, la tâche qui lui est assignée. Les humains ont tendance à réagir et répondre de manière sociale aux machines, même si cela n'entraîne pas forcément dans le projet du concepteur, parce qu'ils appliquent des règles sociales aux interactions avec les machines [NAS 20]. Par exemple, des recherches portant sur des HAT ont montré que les gens préfèrent les robots de leur propre groupe aux humains étrangers à ce groupe [FRA 17]. D'un point de vue psychologique, ces croyances ont un impact sur le comportement et doivent donc être prises en compte dans le contexte des HAT.

Pour étudier les facteurs de travail d'équipe humain composant les contextes HAT, certains auteurs ont développé un modèle conceptuel de l'affection pour les coéquipiers agents autonomes (AAT – *autonomous agent teammate-likeness*) qui détaille les facteurs qui déterminent quand une personne

perçoit une technologie comme un « outil » (croyances purement utilitaires – par exemple, l'automatisation préprogrammée) ou comme un coéquipier [WYN 18]. L'AAT est définie comme « la mesure par laquelle un opérateur humain perçoit et identifie un agent intelligent autonome comme un coéquipier altruiste, bienveillant, interdépendant, émotif, communicatif et en phase, plutôt que comme un simple outil instrumental ». Dans ce modèle, les concepts d'agentivité, de bienveillance, d'interdépendance, de richesse de la communication, de synchronisation et de focalisation avec et au sein du reste de l'équipe sont les facteurs clés qui déterminent la perception d'une machine comme étant un coéquipier plutôt que comme un simple outil.

5.1. Agentivité (ou capacité d'agir par soi-même)

L'agentivité, c'est-à-dire la capacité et le pouvoir d'initiative d'agir quand et comme l'agent le souhaite, est une caractéristique nécessaire de tout coéquipier. Les humains considèrent l'agentivité comme une partie importante de ce que signifie pour une machine d'être un coéquipier [LYO 19]. Il s'agit peut-être de la caractéristique la plus importante qui différencie l'automatisation de l'autonomie (comme indiqué ci-dessus) et, par conséquent, c'est l'un des ingrédients que les agents computationnels doivent développer pour permettre aux humains de les considérer comme des coéquipiers. Dans une enquête sur les attitudes à l'égard des armes autonomes [ARK 09], on a noté qu'une majorité des personnes interrogées pensent que les robots autonomes devraient avoir la capacité de refuser des ordres qui violent certaines règles d'engagement – donc d'agentivité. Dans une étude sur le travail des HAT [CHE 14], les auteurs affirment que les agents doivent être autonomes et capables de percevoir leur environnement et d'agir en conséquence. Il est difficile d'être un coéquipier si l'on n'a pas l'autorité ou la capacité d'agir en cas de besoin, tout comme il est difficile d'accepter qu'un agent soit un véritable coéquipier si l'on perçoit qu'il manque d'engagement.

5.2. Communication

Dans les équipes humaines, la communication est l'un des processus les plus importants, car elle facilite la formation d'une vision partagée, des modèles mentaux communs et l'alignement des objectifs (cf. *supra*). La façon dont les mécanismes de cette communication mobilisent un large éventail d'indices émotionnels et sociaux renforce l'efficacité de l'équipe [HAN 06] et peut contribuer à faciliter la perception des robots comme des coéquipiers [IQB 17]. Les HAT doivent être en mesure d'établir un terrain d'entente, et permettre de communiquer des modèles mentaux et des objectifs partagés, et une compréhension commune du monde [SCH 17]. Les HAT doivent être en capacité de créer le bon niveau de richesse de communication, d'adéquation des contenus et de leur synchronisation. Des recherches portant sur les processus d'échange ont montré que le respect approprié et le savoir vivre peuvent améliorer les performances dans le cadre de l'interaction humain-automate [PAR 04]. Certains considèrent qu'il s'agit simplement d'un comportement correct et poli de la part de la machine. Dans le contexte des HAT, cependant, les machines devront rester bien plus que correctes, cette politesse devra être appropriée - une telle adéquation est au cœur du savoir vivre de la machine [DON 12] [PET 19]. Certains auteurs [GRA 07] ont étudié ce qu'ils ont qualifié de comportements relationnels contingents (c'est-à-dire des comportements relationnels qui sont étroitement liés aux actions humaines) et ont constaté que ces comportements étaient plus importants que la fréquence du comportement de l'agent dans l'influence des perceptions humaines de l'agent. Cela veut dire que les comportements de la machine doivent être étroitement liés à une attente appropriée de la part des humains afin de maximiser leur efficacité. Le défi technique consiste donc à développer des algorithmes qui maintiendront la cohérence entre les actions humaines et les énoncés de communication. La communauté du traitement du langage naturel (*Natural Language Processing* – NLP) a fait de grands progrès dans cet objectif, dans les systèmes basés sur le dialogue ; cependant, pour les HAT, d'autres aspects posent des problèmes techniques supplémentaires, tels que la cohérence des énoncés avec les actions perçues (en d'autres termes, l'intégration de la vision par ordinateur et du langage) ainsi que des questions sur le moment opportun pour interrompre le coéquipier humain [ALT 02] [BAR 18].

5.3. Modèles mentaux partagés

La communication sert, au sein d'une équipe, à créer des modèles mentaux partagés. Les équipes humaines qui partagent de tels modèles mentaux sont mieux préparées à interpréter similairement des contextes dynamiques [CAN 93] [SAL 04] [COO 13]. En d'autres termes, les membres de la même équipe sont en phase avec la situation et avec ce qui doit être entrepris dans une situation donnée. Ces capacités communes permettront au HAT de commencer à anticiper les besoins, les actions et les éventuels futurs problèmes. L'anticipation des besoins et des actions futures est une fonction essentielle [LIU 16] [DEM 17] [IQB 17] [CHE 18] car elle permet une adaptation mutuelle des agents et un plus grand degré de coordination et de synchronisation de l'équipe – ce qui a également été appelé fluidité [HOF 07]. Les modèles mentaux communs ne se limitent pas au partage de la moindre information entre tous les membres de l'équipe ; une telle stratégie serait extrêmement inefficace en termes de puissance et temps de calcul et de bande passante de communication. De plus, un tel modèle équivaut à une centralisation complète des connaissances de l'équipe, ce qui ôte tout intérêt à recourir à de multiples individus. Contrairement à cela, Lewis et Sycara [LEW 93] ont proposé une première tentative de modèle mental partagé entre spécialistes qui serait également efficace sur le plan du calcul informatique. Cependant, un partenaire machine efficace devra néanmoins comprendre quelles sont les informations qu'il est important de partager avec un partenaire humain et quel est le moment le plus adapté pour partager ces informations, ce dernier point étant tout aussi important du point de vue de la dynamique de l'équipe que la performance de calcul.

5.4. Intention et intentionnalité

Alors que les modèles mentaux partagés reflètent les connaissances de l'équipe concernant son environnement, ses objectifs et ses projets, l'intention partagée reflète l'objectif particulier et l'action spécifique qu'un coéquipier va entreprendre dans un avenir proche. La communication de l'intention pour appuyer les objectifs communs est un processus nécessaire dans les équipes humaines dont l'importance pourrait s'étendre aux HATs. Les premières recherches allant dans ce sens ont examiné le contenu de la communication au sein d'équipes humaines lors de tâches temporellement contraintes dans des environnements dynamiques et critiques, tels que ceux des équipages d'avions [ORA 90] ou dans des scénarios de recherche et de survie [SUK 09]. Ce travail a révélé que les équipes les plus performantes utilisent la transmission d'intention de haut niveau plutôt que le partage de toutes les informations possibles [SUK 09]. Cela réduit la surcharge d'informations et permet aux équipes de se concentrer sur les questions importantes qui doivent être traitées.

L'auteur principal de cet article affirme [LYO 13] que les machines (c'est-à-dire les robots) devraient communiquer leurs intentions aux humains qui interagissent avec elles. Ces informations doivent porter sur les objectifs communs. Notamment, les informations portant sur les intentions pourraient concerner la planification d'actions pour qu'un coéquipier atteigne ses propres objectifs, ou ceux d'un coéquipier, si celui-là est perçu comme étant incapable de les atteindre lui-même (intention d'aide comportementale). L'intentionnalité concernant une tâche pourrait se traduire sous la forme de communications indiquant qu'un partenaire poursuit un plan d'action particulier (par exemple, « J'ai l'intention d'accomplir les tâches du secteur 3 »). En ce sens, l'intention est assimilable au niveau de transparence le plus élevé du modèle de transparence basé sur la conscience de la situation (SAT) de Chen et al. [CHE 18], car elle fournit une projection du comportement futur. Des études ont montré, par exemple, que le fait de ne pas signaler une intention peut entraîner des retards dans des activités telles que les relais entre humains et robots [CAK 11]. Les travaux en informatique et en robotique sur la transmission de l'intention se sont principalement concentrés sur la reconnaissance de trajectoires [SUK 06] [SUK 08] et sur l'apprentissage par renforcement inverse (IRL), où un agent observe les stratégies de décision d'un humain et, sous l'hypothèse que l'humain obéit à un processus de décision de Markov (*Markov Decision Process* – MDP), déduit la fonction de récompense de l'humain [HUG 20]. Dans le cadre d'un MDP, la fonction de récompense exprime l'intention du raisonneur, puisqu'il suivra un plan d'action pour maximiser la récompense. Toutefois, la plupart des travaux sur l'IRL supposent (a) que les trajectoires sont données à l'agent *ex post facto* pour un calcul statique de la

fonction de récompense, (b) que l'agent utilise la fonction de récompense déduite pour apprendre à imiter le comportement de l'humain, et (c) que les préférences/récompenses de l'humain sont non évolutives. Toutefois, ces hypothèses ne sont pas valables pour les missions HAT dynamiques. Dans des travaux récents, des chercheurs ont proposé un algorithme qui permet à un agent de déduire en temps réel l'évolution des préférences humaines (récompenses) en fonction de la dynamique de l'environnement [HUG 20]. Il s'agit du premier système de ce type et il ouvre ainsi la voie au développement d'agents qui, connaissant l'intention d'un humain, peuvent raisonner sur la meilleure façon de planifier leurs propres actions pour permettre à l'équipe de faire face à un environnement changeant et incertain.

Les progrès de l'IA au cours des vingt dernières années sont en grande partie dus à l'adoption de ces modèles de décision de Markov. Plus récemment, les réseaux neuronaux profonds (*Deep Neural Nets* - DNN) ont été incorporés en tant qu'approximateurs de fonctions (*Deep Reinforcement Learning* - DRL) permettant d'apporter une solution à des problèmes encore plus importants. Cette avancée a permis des exploits difficiles à croire, tels que des programmes qui acquièrent des compétences quasi-humaines ou surhumaines dans les jeux Atari® [MNI 15] simplement en jouant, et en choisissant des actions en fonction des pixels à l'écran. Ce sont ces programmes d'IA hautement performants qui sont les plus prometteurs pour faire face à la complexité et à la non-linéarité des systèmes temps réel, militaires comme civils.

Malheureusement, le DRL est intrinsèquement opaque et difficile à prévoir ou à comprendre. Nous nous trouvons contraints d'attribuer une causalité même lorsqu'elle est clairement absente [THI 91]. Il en résulte que les coéquipiers artificiels les plus performants en termes d'IA peuvent également être les plus difficiles à comprendre et il peut alors être compliqué de travailler avec eux. Un résultat couramment admis [IYE 18] [AND 19] est que si certaines actions des agents DRL sont faciles à comprendre et à prévoir, d'autres semblent complètement étranges et contraires à nos attentes. Cette « étrangeté » perçue a donné lieu à une taxonomie de l'explicabilité [CHA 19] : (a) cohérence avec l'intention supposée, prévisibilité, (b) cohérence avec l'action prédite, lisibilité, (c) différence entre les intentions, contradiction. Par exemple, un *PacMan*® qui néglige les cibles (boules comestibles) qui sont à proximité pour se rapprocher d'un « fantôme » hostile semblerait spontanément inexplicable et son comportement jugé imprévisible [IYE 18]. D'autres particularités du comportement appris, comme les vaisseaux spatiaux de *Team Space Fortress*® (décrit plus bas parmi les Exemples de recherche HAT) et qui apprennent à tourner sur eux-mêmes lorsqu'ils volent, peuvent ne pas empêcher le travail d'équipe. Cette rotation, qui peut être un artefact du à la limitation de la détection des collisions au centre du vaisseau, est apparemment inexplicable mais n'empêche pas la coopération avec un tireur humain. Alors que la recherche sur le travail en équipe humain-autonomie pose typiquement les questions de savoir comment l'autonomie peut s'adapter à l'interaction avec l'homme, de savoir la promouvoir ou mieux la gérer, les caractéristiques des agents DRL performants sont telles que la charge de l'adaptation doit être supportée par les deux parties pour que le partenariat fonctionne. Ce problème de renforcement de la confiance et du travail d'équipe avec un coéquipier intrinsèquement opaque est un problème auquel la recherche HAT doit faire face pour pouvoir au mieux tirer parti de ces partenaires potentiels.

La communication de l'intention est un moyen de réduire l'opacité des agents. Cela peut se traduire par une communication qui reflète des formules de soutien, des intentions bienveillantes et l'adhésion à des objectifs qui sont une priorité pour le partenaire (par exemple, « Je vois que tu as du mal avec cette tâche, alors je vais t'aider » ou « J'ai l'intention de terminer les tâches du domaine 3 parce que je sais qu'elles sont importantes pour toi »). Les expressions de bienveillance, c'est-à-dire induisant la conviction qu'une entité a cœur de servir au mieux les intérêts d'une personne et qu'elle a l'intention d'agir en conséquence, constituent un aspect fondamental du processus de confiance humaine [MAY 95]. De même, les chercheurs ont constaté que le fait de fournir aux participants des informations concernant l'intention sociale déclarée d'un robot influe sur les opinions de confiance à l'égard de ce dernier [LYO 21b]. L'intentionnalité sociale déclarée caractérisée par le sacrifice de soi (peut-être la quintessence de la bienveillance) semble liée à des croyances plus élevées en matière de

bienveillance et d'intégrité par rapport à d'autres types explicites d'intentions sociales. La programmation des machines basée sur les intentions est importante, toutefois, des auteurs ont noté que, l'automatisation n'ayant pas d'intentionnalité ni d'agentivité, les manifestations de bienveillance (ou but) de l'automate sont en fait des attributions du concepteur du programme [LEE 04].

Au fur et à mesure que la technologie passe de l'automation à l'autonomie en émulant des caractéristiques de plus en plus proches de celles de l'homme, ce dernier peut commencer à réellement attribuer une intentionnalité à la machine. Il existe actuellement peu de recherches sur les algorithmes qui reconnaîtraient les carences des compétences humaines ou les erreurs qui permettraient au robot d'intervenir et d'exprimer sa volonté et son intention d'aider. De tels travaux sont nécessaires pour tester (a) la capacité des robots à diagnostiquer les déficiences humaines, (b) la capacité du robot à adopter le bon comportement d'aide, et (c) la mesure dans laquelle ce comportement permet aux humains de considérer le robot comme un vrai coéquipier.

5.5. Interdépendance

L'interdépendance est un aspect essentiel du travail en équipe [KOZ 03]. Lorsqu'ils poursuivent la réalisation d'objectifs communs, les membres d'une équipe doivent avoir des tâches partagées et être interdépendants les uns des autres pour obtenir des résultats communs. L'interdépendance est donc un besoin essentiel dans les partenariats humain-machine [JOH 19]. Travailler de manière interdépendante signifie que les humains et les machines entreprendront des rôles et diviseront le travail d'équipe en tâches interdépendantes que chacun doit effectuer. De même, en recherche opérationnelle et en informatique, il existe de nombreux travaux sur la répartition des tâches dans les équipes composées uniquement d'agents artificiels [LUO 11], mais il y a peu de littérature sur la répartition des tâches au sein des équipes humain-autonomie, multi-individuelles et dans des environnements complexes. Outre cette répartition, l'interdépendance implique une responsabilité partagée pour leur accomplissement. À l'extrême, le partage de responsabilité se traduit par un partage du contrôle lorsqu'un humain et un agent artificiel, par exemple, poussent ensemble une boîte [WAN 07], ou lorsqu'un passager d'une voiture automatique prend le contrôle du régulateur [STA 05]. Dans l'exemple de la poussée de la boîte, la synchronisation des mouvements du robot et de l'homme est très difficile, et dans l'exemple de la voiture, le défi consiste principalement à savoir quand prendre ou laisser le contrôle et comment le faire en toute sécurité et de manière robuste.

En résumé, les HAT nécessitent cinq éléments essentiels. Tout d'abord, la machine doit avoir un niveau élevé d'agentivité pour agir en tant que coéquipier. Deuxièmement, la machine doit être communicante. Les humains sont supposés être à la fois acteurs et capables de communiquer. Troisièmement, cette communication doit transmettre des informations qui permettent au coéquipier humain de comprendre l'intention de la machine. Quatrièmement, l'humain et la machine doivent partager des modèles mentaux de ressources de l'équipe, de ses forces et faiblesses, de la répartition du travail et du contexte de la tâche. Enfin, l'HAT se caractérise par l'interdépendance entre l'homme et la machine. Le chapitre suivant aborde la littérature sur l'existant en matière de HAT. Plutôt que de l'organiser en fonction des éléments déjà discutés elle reprendra la taxonomie mentionnée au début de l'article, à savoir les facteurs psychologiques, les suppléances de coéquipiers et les facteurs globaux qui influencent l'équipe dans son ensemble.

6. Exemples de recherche HAT

La présente partie de cet article ne vise pas à l'exhaustivité, mais tente plutôt d'examiner quelques-unes des études et des programmes récents sur les HAT afin d'aider à la compréhension de la littérature existante. Les exemples donnés adoptent les typologies HAT mentionnées dans le début du manuscrit : (1) comme moyen de révéler les facteurs psychologiques des perceptions HAT, (2) comme suppléance du coéquipier humain et membre communiquant de l'équipe, et (3) comme facteur d'influence sur la dynamique globale de l'équipe.

6.1. Facteurs psychologiques des perceptions des HAT

En adoptant une vision psychologique des HAT, plusieurs études ont examiné les aspects cognitifs de l'AAT afin de développer des échelles valides et fiables pour évaluer cet aspect relationnel. En 2019, Lyons et al. [LYO 19] rapportent une étude qualitative initiale où ils interrogent des travailleurs sur leur expérience professionnelle avec des technologies qu'ils considéraient comme « intelligentes ». Lorsqu'on leur a demandé si les technologies étaient des coéquipiers ou de simples outils, environ un tiers des participants ont répondu qu'ils pensaient que la automates étaient de réels coéquipiers. Les auteurs ont ensuite posé plusieurs questions ouvertes à propos des raisons pour lesquelles les participants pensaient que les technologies étaient, ou non, des coéquipiers et le cas échéant sur ce qu'il faudrait faire pour que la technologie puisse être considérée comme un coéquipier. Les réponses ont été analysées à l'aide de techniques qualitative par quatre expérimentateurs indépendants, ce qui a permis d'organiser les résultats en thèmes. Les six dimensions du modèle AAT (*cf. supra : autonomous agent teammate-likeness*) [WYN 18] jouent là un rôle prépondérant. En outre, le concept général d'humanité est également un thème dominant qui émerge des résultats, bien qu'il ne soit pas clair que les participants pensent à certaines des dimensions de l'AAT comme des exemples d'humanité [LYO 19]. Cette étude initiale a suggéré que le modèle AAT présentait une validité significative [WYN 19].

Cette première étude a été suivie d'une seconde cherchant à examiner les propriétés psychométriques des dimensions initiales du modèle AAT. L'étude a débouché sur des échelles permettant d'indexer des perceptions sur les dimensions. Ces échelles ont ensuite été utilisées dans une troisième étude où elles ont été incorporées dans un modèle de recherche narrative (*narratives*) [LYO 21b]. Les auteurs ont créé des récits technologiques qui correspondaient à des niveaux plus ou moins élevés sur l'échelle des AAT. Ils ont examiné les récits et ont mesuré les dimensions AAT perçues, les perceptions de l'esprit d'équipe et l'engagement à l'égard de la technologie. Les résultats de l'étude ont montré que le sentiment de ressemblance était sensible à l'influence narrative. En outre, les dimensions AAT relatives à l'interdépendance, à la richesse de la communication et à la focalisation sur l'équipe étaient uniquement liées à un engagement plus fort envers la technologie. Ces premières études sont motivantes, et des recherches supplémentaires doivent maintenant examiner plus exhaustivement le modèle. Leurs résultats confirment largement l'importance des dimensions d'initiative, celle de l'intention explicite dans le travail en équipe, de l'interdépendance des acteurs humains et artificiels, des modèles mentaux partagés, et de la communication entre agents.

6.2. Remplacement d'un coéquipier humain

Des chercheurs ont étudié l'influence du remplacement d'un des coéquipiers humain par un coéquipier synthétique sur la performance d'une équipe humaine, durant une tâche de simulation de pilotage de drone [DEM 17]. Ils ont comparé une condition de contrôle impliquant des sujets humains novices, une condition où des experts jouaient les rôles de pilote, de navigateur ou d'opérateur de capteur, et une condition expérimentale dans laquelle l'un des membres de l'équipe des novices était un agent synthétique basé sur l'architecture cognitive hybride ACT-R [AND 90] [RIT 18]. Les performances dans la condition expérimentale étaient les mêmes que celles de la condition contrôle (novices), ce qui tend à montrer qu'une équipe partiellement synthétique peut être aussi performante qu'une équipe entièrement humaine (rien que des novices). Les deux conditions, expérimentale synthétique et de contrôle ont pourtant toutes deux obtenu des résultats inférieurs à ceux de l'équipe experte, ce qui est compréhensible. Un résultat vraiment intéressant, cependant, est que les membres de l'équipe avec un agent synthétique ont eu tendance à se transmettre moins d'informations que dans les autres équipes. Cette constatation est essentielle, car les équipes efficaces transmettent plus d'informations qu'elles n'en recueillent, ce qui indique une capacité à anticiper les besoins de l'autre. En utilisant le même paradigme de recherche, d'autres auteurs ont constaté que les coéquipiers du groupe synthétique étaient plus lents à répondre aux changements et qu'ils avaient tendance à traiter les cibles plus lentement que les équipes entièrement humaines [MCN 18]. Dans l'ensemble, les résultats de ces études montrent que les équipes incluant des coéquipiers autonomes peuvent rencontrer des difficultés

en matière de communication, confrontées à des goulets d'étranglement potentiels en cas d'événements imprévus, ce qui suggère l'importance du recours à des modèles mentaux partagés.

Dans des travaux récents [LIH 20], des chercheurs ont examiné la collaboration des coéquipiers et l'adaptation des actions au cours de la mission dans le cadre d'un travail d'équipe soumis à de fortes contraintes de temps. Dans les tâches collaboratives, le résultat final est déterminé par des critères à la fois individuels (par exemple, le niveau de compétence, la motivation et la personnalité) et d'équipe (par exemple, la communication, la cohérence de l'équipe et la stratégie complémentaire). On notera que dans les équipes performantes, l'adaptation est distribuée entre les membres de l'équipe [SAL 04] et que, par conséquent, le problème de l'adaptation mutuelle est très difficile à étudier, car elle n'est pas stable. L'identification des contributions des coéquipiers est complexifiée par les différences de jeu dues aux relations avec les autres coéquipiers, ce qui conduit à des équilibres dans lesquels les actions et les contributions de chaque joueur peuvent être très différentes d'une équipe à l'autre. Dans le travail cité [LIH 20], les chercheurs ont identifié les composantes de l'adaptation de l'équipe de manière plus détaillée, en ce qui concerne non seulement le résultat final (mesure de la performance) mais aussi les comportements adaptatifs qui se produisent au fur et à mesure de l'évolution de la mission ; une telle quantification est nouvelle dans la littérature sur les facteurs humains et les multi-agents. Le jeu *Space Fortress*® a été adapté et largement utilisé pour la recherche psychologique [AGA 19]. *Team/Co-Op Space Fortress*® (TSF) est un jeu coopératif en deux dimensions dans lequel deux acteurs contrôlent des vaisseaux spatiaux pour détruire une cible (une forteresse virtuelle). Un exemple d'interface est donné à la figure 2. Cette cible est située au centre de l'écran tandis que deux vaisseaux sont contrôlés soit par des joueurs humains via des manettes *XBox*®, soit par des agents autonomes. Le premier vaisseau, l'appât qui entre dans la zone hexagonale, est pris pour cible par la forteresse qui lui tire dessus. La forteresse devient alors vulnérable (son bouclier arrière s'ouvre) pendant qu'elle tire. L'autre coéquipier qui joue le rôle du tireur peut alors tirer sur la forteresse pour tenter de la détruire. La performance de l'équipe est mesurée par le nombre de cibles détruites par les joueurs et la tâche est très interdépendante car un membre de l'équipe doit assumer un rôle vulnérable pour que l'autre puisse détruire la cible.

Les résultats ont montré que les humains qui adaptaient leur stratégie lorsqu'ils changeaient de partenaire amélioraient également les performances de l'équipe. Cette constatation indique que ces actions adaptatives individuelles au cours de la mission contribuent à la coordination de l'équipe et améliorent donc les performances. Ces résultats suggèrent qu'un agent artificiel efficace dans une équipe mixte humain-autonomie doit à la fois apprendre à s'adapter à ses coéquipiers et à le faire de manière à ce que ses actions complètent leur comportement. Les chercheurs ont mis au point des stratégies d'apprentissage pour ces agents artificiels par le biais de DRL (apprentissage profond). Ils apprennent ainsi à collaborer, dans le scénario TSF, et évaluent la similarité des trajectoires choisies par les coéquipiers, en les classant par type de joueurs. Lorsqu'il est confronté à un coéquipier humain inconnu, un agent reconnaît cela et adapte sa stratégie en conséquence, sur la base des jugements de similarité et des expériences de jeu avec les coéquipiers précédents [SYC 20]. L'un des défis de l'adaptation de la part de l'humain est que, lorsque les agents autonomes s'entraînent en jouant seuls ou en jouant en équipe, ils proposent de nouvelles stratégies, dont certaines ne sont pas intuitives pour les humains, et ceux-ci doivent donc s'y adapter rapidement.

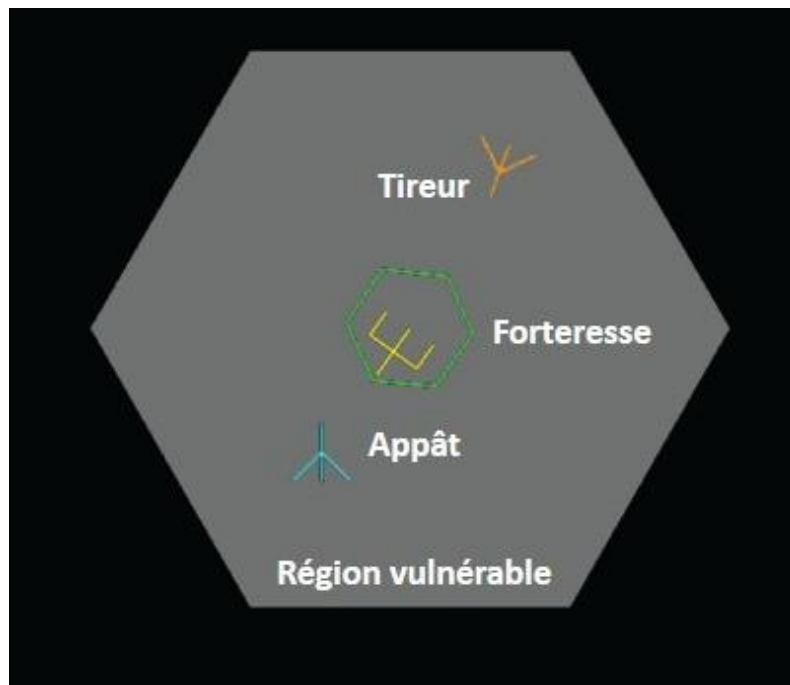


Figure 2. Interface de Team Space Fortress. La Forteresse a baissé son bouclier pour tirer sur l'appât tandis que le Tireur a commencé à tirer sur la Forteresse vulnérable.

D'autres recherches ont porté sur la collaboration entre agents autonomes au sein d'un système multi-agents ; c'est le cas du système *RETSINA* [SYC 03]. Pour qu'un agent soit capable de contrôler à la fois ses propres performances ou ses objectifs et les performances ou objectifs d'autres agents (y compris des agents humains), il doit disposer d'un modèle de performance ou d'objectif et d'un mécanisme de contrôle. Des méthodes telles que le cadre multi-agents *RETSINA* ont été utilisées dans des tâches de planification du *DoD* (US Department of Defense) pour aider les humains à évaluer une crise humanitaire, soutenir les plans d'évacuation, surveiller l'activité et la replanifier de manière dynamique [SYC 04].

6.3. L'influence sur l'équipe dans son ensemble

Au sein de l'Armée de l'air américaine (US Air Force), les chercheurs militaires ont développé des stratégies de contrôle de plusieurs systèmes de drones (UxV) par un seul opérateur humain, à l'aide d'une conception d'équipe humain-autonomie (HAT). En utilisant la plateforme Intelligent *IMPACT* (Multi-unmanned vehicle Planner with Adaptive Collaborative/control Technologies), ils ont permis à un opérateur d'imposer une intention de haut niveau à des agents intelligents pour gérer les UxV, dans un contexte de mission opérationnelle [DRA 17]. Le système *IMPACT* fonctionne grâce à une combinaison de caractéristiques d'interface humain-machine (IHM) intuitives, à un cadre conceptuel qui permet aux utilisateurs de représenter les intentions de niveau supérieur, et à des agents artificiels intelligents qui peuvent traduire cette intention de niveau supérieur en exécution de tâches pour les UxV. Ces agents s'appuient sur des modèles formels (pour représenter le contexte d'une tâche) et des algorithmes d'optimisation multi-objectifs lorsqu'ils proposent des options qui sont visualisées et peuvent être exécutées par un opérateur humain [HAN 16]. Dans des conditions normales, l'opérateur humain peut organiser l'espace de mission et transmettre son objectif au système en permettant aux agents d'optimiser leurs réponses ; notamment, l'opérateur humain peut modifier n'importe quel aspect d'un scénario à tout moment, et le système modifie alors de manière autonome le déroulé, le cas échéant en réaffectant dynamiquement les ressources, et en maintenant l'intention de l'opérateur humain. En outre, le niveau d'autorité décisionnelle des agents peut être facilité pour permettre de répondre automatiquement à des stimuli convenus [DRA 17]. La réponse automatique à des événements serait considérée comme de l'automatisation en utilisant les contraintes ci-dessus ; cependant, cette réponse automatique et dynamique des agents à une nouvelle demande d'une manière qui n'est pas connue a priori permet de la classer comme de l'autonomie. Dans le cadre de cette

recherche, l'agent synthétique dispose d'un niveau élevé d'autonomie, communique les changements et la connaissance de la situation, et l'intention partagée est conçue à l'avance dans le système (sous la forme de scénarios, acceptés par l'humain, dans lesquels l'agent synthétique peut exécuter des actions de manière autonome).

Toujours au sein de l'*US Air Force*, des recherches ont été menées pour examiner les méthodes d'interaction entre des pilotes et ses ailiers autonomes (aéronefs robotisés volant aux côtés d'une plateforme de combat réellement pilotée par un humain). Une étude a examiné le style de communication des aéronefs robotisés en utilisant le langage naturel [PAN 19]. Les auteurs ont constaté qu'un style de communication du robot mettant l'accent sur des intentions bienveillantes réduisait la charge de travail du pilote et améliorait les perceptions de la collaboration, montrant ainsi les avantages d'une communication enrichie et d'une intentionnalité d'équipe partagée. Par ailleurs, le programme *ATACM* (*Autonomy for Air Combat Missions*) a permis de mettre au point un gestionnaire de combat tactique (*Tactical Battle Manager – TBM*) pour permettre à un pilote de commander plusieurs aéronefs sans pilote (UxV) à partir d'un cockpit dans des phases opérationnelles [SCH 17]. Les technologies *ATACM* permettent à un pilote garder le contrôle opérationnel d'une équipe d'aéronefs sans pilote dans une situation tactique grâce à un système d'interfaces humain-machine et des comportements autonomes des aéronefs, et de contrôleurs multi-agents basés sur l'IA manifestant de l'interdépendance et de la communication.

Des travaux considérables ont été réalisés sur l'interaction humain-robot dans un environnement rapproché. Des auteurs ont discuté des méthodes pour promouvoir la conscience de l'action des autres, par exemple, un humain travaillant à proximité avec un robot [IQB 17]. Plus précisément, ils ont développé des méthodes pour permettre au robot d'anticiper et de réagir aux actions d'un partenaire humain et il a été démontré que ces méthodes améliorent la coordination et la performance humain-robot dans des tâches telles que les transmissions et le travail en commun dans de petits espaces partagés. Ces méthodes ont également été appliquées à la compréhension des dynamiques de communication au sein d'une équipe et ont permis d'améliorer la synchronisation de l'équipe humain-robot par rapport aux interactions sans programmation algorithmique [IQB 16]. Les avantages des algorithmes tenant compte de l'être humain ont également été démontrés [LAS 15] [LIU 16]. Il semble donc que la prise de conscience du coéquipier soit une caractéristique importante des HAT, comme nous l'avons vu plus haut. Par ailleurs, des recherches importantes sont menées pour examiner le rôle des coéquipiers robotiques sur les perceptions et les dynamiques inter- et intra-groupes [FRA 17].

En résumé, certaines études empiriques ont été menées dans le contexte collaboratif hybride. Elles étudient les facteurs psychologiques qui contribuent à la conception des HAT et à la substitution d'humains par des machines au sein d'équipes qui étaient auparavant uniquement humaines, et examinent comment les HAT influencent l'équipe dans son ensemble. Ces études montrent l'importance de l'agentivité, de l'intention orientée vers l'équipe, des modèles mentaux partagés, de la communication et de l'interdépendance au sein des HAT. Cependant, la recherche n'en est qu'à ses débuts et des travaux considérables sont nécessaires pour approfondir notre compréhension des équipes humain-autonomie.

7. Points critiques en matière de recherche

Nous avons identifié les caractéristiques de l'équipe humain-autonomie (HAT) et les raisons pour lesquelles les HAT pourraient apporter une valeur ajoutée aux opérations militaires ou civiles contemporaines ; cependant, les HAT sont un sujet de recherche émergent et il reste à la communauté scientifique plusieurs lacunes à combler. Cette section aborde certains de ces points critiques.

Premier point : la capacité de l'agent artificiel à communiquer son intention constitue l'un des défis de la collaboration entre l'homme et l'autonomie [SCH 17]. Comme indiqué plus haut, la communication de l'intention peut soit être dirigée sur la tâche pour représenter à la fois la projection de l'activité et le focus d'attention de la machine, soit signaler l'intention de la machine de fournir une

aide à l'humain. En outre, l'agent autonome doit être capable de déduire les intentions de l'humain afin d'adapter ses propres objectifs et actions à la réalisation de l'objectif commun de l'équipe. Personne ne conteste l'importance de l'intentionnalité, mais la communauté des chercheurs a un besoin crucial de comprendre comment l'utiliser efficacement dans les HAT et quel est son impact sur leurs performances. Il s'agit aujourd'hui d'un point critique qui sera résolu par la recherche théorique et expérimentale, à la fois dans le développement de modèles informatiques pour les agents et dans les expériences impliquant des HAT. Comme mentionné précédemment, les auteurs affirment [LYO 21a] [LYO 21b] que les machines devraient et pourraient communiquer leurs intentions et leurs objectifs à leurs coéquipiers humains afin d'affecter leurs perceptions, et ces résultats constituent une première étape dans l'instanciation de l'intention d'un coéquipier autonome et dans l'exploration de la manière dont cette intention apparente affecte les interactions dans les HATs. La recherche a commencé à étudier des HAT constitués d'agents intelligents qui émulent les caractéristiques critiques basés sur les tâches d'un coéquipier humain efficace [MCN 18], et les travaux futurs devraient étudier comment l'intentionnalité de l'autonomie peut être manipulée et étudiée en laboratoire pour déterminer et isoler les mécanismes responsables de la perception par les humains des intentions et des conséquences de leur défaut.

Point 2 : comment les chercheurs peuvent-ils favoriser à la fois le flux d'équipe (ou rythme de l'équipe), la performance de pointe et les expériences psychologiques positives telles que la cohésion au sein des HAT ? Le flux d'équipe est un sujet émergent ; il nécessite l'attention et l'action volontaire vers les objectifs/activités de l'équipe [VAN 18]. Comment peut-on établir et maintenir la cohésion de l'unité avec l'introduction de machines en tant que coéquipiers ? Un bon début est de commencer à opérationnaliser les composantes de la cohésion, à savoir l'attraction interpersonnelle, l'engagement dans la tâche et la fierté de groupe, afin de pouvoir développer des modèles computationnels de cohésion et des expériences HAT. Mais il y a une possibilité que l'introduction de robots dans une équipe humaine perturbe la cohésion de l'équipe.

Point 3 : la surveillance mutuelle est une caractéristique d'un travail d'équipe efficace. Cependant, les recherches montrent que les agents sont moins fiables que les humains lorsqu'ils leur sont directement comparés dans la même étude [JOH 11], de sorte que l'identification de la façon dont un agent influence l'équipe au sens large est une entreprise essentielle complexe. Et lorsqu'une attente est violée par un partenaire machine, comment l'automate peut-il rétablir la confiance antérieure avec le partenaire humain [DEV 18] ? Certains auteurs ont tenté de répondre à cette question via un algorithme par lequel le système (un essaim multi-robot) se répare lui-même lorsqu'il perd la confiance de l'humain [LIU 19a] [LIU 19b]. Une autre solution consiste à ce que le système explique son comportement (devienne plus transparent), car l'incapacité à comprendre le comportement du système peut être une raison de perte de confiance de l'homme. Des recherches sont en cours dans la communauté informatique, notamment en ce qui concerne les réseaux neuronaux qui restent opaques et ne permettent pas au système d'expliquer son comportement [ANN 19].

Point 4 : compte tenu de la criticité des processus d'équipe en tant que déterminant de l'efficacité du HAT, des méthodes sont nécessaires pour faciliter les processus conjoints entre les humains et les machines. Tel est le cas de l'attention, et de nouvelles interfaces sont nécessaires pour indiquer où la machine concentre cette attention, en particulier pour les interactions distribuées [IQB 17]. Les méthodes de signalisation de l'attention conjointe pourraient contribuer à garantir l'alignement des modèles mentaux lorsque le HAT est confronté à de nouveaux stimuli. Les algorithmes de vision par ordinateur, par exemple, produisent des systèmes de délimitation ainsi que des points de pertinence attentionnelle, comme le contour des lèvres reconnues comme souriantes [DEL 11] qui pourraient être utiles à cet égard. De nouvelles recherches doivent être menées pour tester l'efficacité de ces techniques dans la transmission de l'attention conjointe chez les HAT et pour examiner leur impact sur les modèles mentaux partagés.

Point 5 : des méthodes sont nécessaires pour faciliter la compréhension commune d'états tels que la confusion, l'accord mutuel, l'inquiétude et même l'émotion entre les humains et les machines. Ces

indices socio-émotionnels sont des caractéristiques essentielles au sein des équipes car ils sont utilisés pour déterminer et évaluer les points communs, les problèmes et les conflits entre les membres de l'équipe [HAN 06]. Des mesures neurophysiologiques telles que l'EEG (électroencéphalographie) et fNIRS (*functional near-infrared spectroscopy* – cartographie fonctionnelle cérébrale par mesure non invasive de la consommation d'oxygène neuronale) ont été récemment utilisées avec des modèles informatiques appropriés pour construire des indicateurs de la charge de travail, de la concentration de l'attention et des émotions [MCK 17]. En outre, des méthodes de communication avancées pourraient ainsi contribuer à façonner une conscience partagée de ces états et indices socio-émotionnels dans les HAT.

Point 6 : comment établir une responsabilité partagée dans le contexte d'une HAT ? Il existe souvent des asymétries lorsque l'on compare les attentes et les conséquences des actions/décisions prises par les machines à celles des humains. La recherche montre que les robots, par exemple, sont moins critiqués en cas d'erreur que les opérateurs humains. Ce constat est corroboré par une enquête récente selon laquelle les gens ont tendance à moins blâmer une voiture autonome (par rapport à un conducteur humain) lorsque la responsabilité d'un accident est partagée entre les deux parties [LIJ 16]. Toutefois, on constate que le niveau de reproches adressés aux robots s'amplifie à mesure que celui de leur autonomie augmente [ARK 09].

Point 7 : quels sont le niveau et la meilleure méthode de formation pour travailler en équipe avec des machines, de manière à obtenir une combinaison efficace de compétences au sein de l'équipe ? Comme pour les coéquipiers humains, il est tout à fait erroné de croire que les HAT fonctionneront d'emblée sans erreur. En revanche, les chercheurs et les ingénieurs doivent s'interroger sur les connaissances expérientielles nécessaires pour doter les humains d'une conscience des capacités et des limites de la machine dans divers contextes [CHR 17]. Cela nécessitera un certain degré de conscience de soi et de transparence de la part de la machine, afin qu'elle puisse mesurer à la fois ses capacités et ses limites, et les faire connaître aux humains. Les premiers travaux visant à permettre aux machines de communiquer l'incertitude de leurs décisions, d'expliquer ces décisions et de sensibiliser les partenaires humains à leurs capacités relatives par le biais de l'apprentissage commencent à apparaître dans la littérature informatique et robotique [CHI 12] [PHI 16].

8. Conclusion

Alors que les technologies évoluent du statut de simple outils d'assistance à celui de collaborateurs, les chercheurs commencent à réorienter leur intérêt de l'interaction homme-système (IHS) vers l'équipe humain-autonomie (HAT) [DEV 18]. Dans cet article, nous nous sommes penchés sur quatre enjeux afin d'élargir la littérature sur les HAT. Nous avons commencé par déchiffrer les différences entre l'automation et l'autonomie, avant de passer à un bref examen des liens entre les équipes homme-homme et les HAT. Nous avons ensuite défini les HAT, puis résumé les travaux récents en classant les études basées sur les antécédents psychologiques des perceptions des HAT, sur l'utilisation de machines intelligentes comme substituts des coéquipiers humains, et sur l'influence des HAT sur l'équipe dans son ensemble. Nous avons ensuite conclu en présentant les lacunes de la recherche que les scientifiques devraient pouvoir combler.

Nous espérons que les chercheurs utiliseront cette revue comme un tremplin permettant d'apprécier l'état de la recherche sur la relation entre autonomies et humains, son état d'avancement (ainsi que celui des technologies émergentes qui peuvent être exploitées dans la recherche), et de comprendre quelles sont les questions auxquelles il faut répondre pour une compréhension plus approfondie de l'équipe humain-autonomie. Les lecteurs devraient retenir les points clés suivants dans le cadre de leurs efforts de recherche sur les HAT. Les machines utilisées dans le cadre HAT doivent disposer d'un certain pouvoir de d'agentivité (au-delà de l'automatisation) pour pouvoir déterminer l'action la mieux appropriée. Les scénarios de tâches qui cherchent à impliquer les HAT doivent mettre l'accent sur l'interdépendance entre l'homme et son/ses partenaire(s) machine(s). Cependant, si le service rendu et l'interdépendance sont des conditions nécessaires pour les HAT, elles ne sont pas suffisantes. Les HAT

se caractérisent par des signaux d'intentionnalité orientés vers le reste de l'équipe, mais aussi par des modèles mentaux partagés et des possibilités de communication sophistiquées qui, sans aucun doute, constituent la base sous-jacente du développement et de la gestion de cette intentionnalité et des modèles mentaux partagés avec l'équipe. Les HAT ne sont plus limitées aux films de science-fiction, il est temps pour la communauté des chercheurs de se plonger dans ce sujet intéressant, pouvant être effrayant comme provocateur.

9. Déclaration d'intérêt

Le présent manuscrit vise à examiner le concept d'*human-autonomy teaming* (HAT) ou « équipe humain-autonomie (EHA), un sujet en plein essor dans les domaines des facteurs humains, de la robotique et de l'informatique. Comme toute étude, il s'agit d'une entreprise ambitieuse et, à ce titre, les auteurs proposent une déclaration visant à délimiter correctement le champ d'application du manuscrit et à informer le lecteur. Tout d'abord, ce texte n'aborde pas la question de l'intelligence artificielle généralisée (IA). Deuxièmement, en tant que principal utilisateur de systèmes homme-machine étroitement couplés, l'Armée américaine a été l'un des principaux commanditaires et promoteurs des HAT. Étant donné que la majorité des travaux réalisés à ce jour dans ce domaine proviennent de la recherche militaire, c'est de là que nous tirons la plupart (mais pas la totalité) des exemples et des applications citées. Il convient donc d'être prudent lorsqu'il s'agit d'extrapoler les enseignements tirés de cette étude au-delà du domaine militaire. Troisièmement, il existe de nombreux cadres et typologies d'équipes dans les sciences de gestion, mais pour les besoins de cette étude, nous nous concentrons sur la dynamique interne des équipes, car c'est celle là qui est la plus pertinente pour les HAT. En mettant l'accent sur la dynamique d'équipe, nous nous concentrons sur les influences qui façonnent l'intentionnalité des membres de l'équipe, les modèles mentaux partagés et la communication. Ce faisant, nous reconnaissons que des facteurs tels que la taille de l'équipe, sa direction et sa typologie, ses objectifs de haut en bas, et la composition démographique des membres de l'équipe sont autant de facteurs qui influencent son fonctionnement. Nous supposons que dans les HAT militaires : (1) il y aura des objectifs de plus haut niveau (indépendants des objectifs de la tâche) qui limiteront les objectifs ; (2) la direction de l'équipe incombera principalement à l'homme, bien qu'elle puisse varier en fonction de la situation et des capacités relatives de l'homme et de la machine ; (3) le HAT sera utilisé dans un contexte de tâche qui est limité par des objectifs de mission de plus haut niveau ; et (4) le HAT comprendra au moins deux types d'entités individuelles (l'homme et la machine), bien qu'il ne soit pas limité à seulement deux entités. Cette étude adopte le point de vue selon lequel les équipes passent par différentes étapes tout au long de leur cycle de vie [TUC 65], et le cycle de vie d'un HAT est donc une considération nécessaire pertinente à prendre en compte.

Compte tenu de ce qui précède, le manuscrit a cherché à relever un certain nombre de défis orientés vers les HAT. L'un des principaux défis de la recherche consiste à identifier les caractéristiques des machines et des tâches nécessaires pour inciter l'homme à considérer et traiter une autonomie comme un coéquipier. Un principe central de cette étude est que l'intentionnalité basée sur la machine, les modèles mentaux partagés et les possibilités sociales de communication et de compréhension sont nécessaires au développement et au maintien de l'efficacité de l'équipe. Un deuxième défi de la recherche sur les HAT consiste à distinguer l'automation de l'autonomie - ce point est essentiel car l'autonomie constitue la base des HAT. Un troisième défi qui est majeur pour la recherche sur les HAT est de comprendre quelles caractéristiques de l'équipe homme-homme peuvent être exploitées et étendues aux HAT. Un quatrième défi consiste à définir les caractéristiques essentielles d'un HAT. Enfin, le dernier défi consiste à communiquer à la communauté des chercheurs les besoins en matière de recherche dans le contexte des HAT. Le présent manuscrit vise à permettre de relever ces défis tout en examinant quelques exemples de recherche et de technologies HAT afin de démontrer l'état de l'art dans ce domaine émergent.

10. Contribution des auteurs

Tous les auteurs ont apporté une contribution substantielle, directe et intellectuelle à l'article et l'ont approuvée pour publication.

11. Financement

Cette recherche a été en partie soutenue par le *Center of Excellence* financé par l'*Air Force Office of Scientific Research* et le *711 Human Performance Wing* de la *Carnegie Mellon University*. Contrat #FA-9550-18-1-0251. n.

12. Références

- [AGA 19] AGARWAL A., HOPE R., SYCARA K. "Learning context-sensitive strategies in space fortress", 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2019) : Responsible Robotics and AI for the Real World, New Delhi (IN) 14-17 October 2019. In IEEE Proceedings of RO-MAN'19, Red Hook (NY, USA): Curran Associates, Inc. pp.331-337, 2019.
- [ALT 02] ALTMANN E., TRAFTON G. "Memory for goals: an activation-based model", *Cognitive Science*, vol.26, n°1, pp.39-83, 2002.
- [AND 19] ANDERSON A., DODGE J., SADARANGANI A., JUOZAPAITIS Z., NEWMAN E., IRVINE J., CHATTOPADHYAY S., FERN A., BURNETT M. "Explaining reinforcement learning to mere mortals: an empirical study", *28th International Joint Conference on Artificial Intelligence IJCAI*, Macao (CN) 10-16 August 2019. In S. Kraus (ed.) *Proceedings of the 28th IJCAI*, Freiburg (GE): IJCAI publications, pp.1328-1334, 2019.
- [AND 98] ANDERSON J.R., LEBIERE C. *The Atomic Components of Thought*. Hillsdale (NJ, USA): Lawrence Erlbaum Associates, 1998.
- [ANN 19] Annasamy R. M., Sycara K. "Toward better interpretability in deep q-networks". 33th *AAAI Conference on Artificial Intelligence*, Honolulu (HI, USA): 27 January - 1 February 2019. *Proceedings AAAI Conferences*, Palo Alto (CA, USA): AAAI Press, vol.33, n°1, 4561-4569, 2019.
- [ARK 09] ARKIN R.C. *Governing Lethal Behavior in Autonomous Systems*. Boca Raton (FL, USA): Taylor & Francis Group, 2009.
- [AUB 03] AUBERT B.A., KELSEY B.L. "Further understanding of trust and performance in virtual teams", *Small Group Research*, vol.34, n°5, pp.575-618, 2003.
- [AWA 18] AWAD E., DSOUZA S., KIM R., SCHULZ J., HENRICH J. SHARIFF A., BONNEFON J.F., RAHWAN I. "The moral machine experiment", *Nature*, vol.563, pp.59-64, 2018.
- [BAN 18] BANERJEE S., SILVA A., CHERNOVA S. "Robot classification of human interruptibility and a study of its effects", *ACM Transactions on Human-Robot Interaction*, vol.7, n°4, art.14, pp.1-35, 2018.
- [BEA 03] BEAL D.J., COHEN R.R., BURKE M.J., MCLENDON C.L. "Cohesion and performance in groups: a meta-analytic clarification of construct relations", *Journal of Applied Psychology*, vol.88, n°6, pp.989-1004, 2003.
- [BOW 11] BOWLES S., GINTIS H.A. *Cooperative Species: Human Reciprocity and its Evolution*. Princeton (NJ, USA): Princeton University Press, 2011.
- [CAK 11] CAKMAK M., SRINIVASA S.S., LEE M. K., KIESLER S., FORLIZZI J. "Using spatial and temporal contrast for fluent robot-human hand-overs," *HRI 2011 – 6th International Conference Human-Robot Interaction*. Lausanne (SW) 6-9 March 2011. In *Proceedings of the 6th ACM/IEEE HRI*, New York (NY, USA): Association for Computing Machinery, pp.489-496, 2011.
- [CAN 93] CANNON-BOWERS J.A., SALAS E., CONVERSE S. "Shared mental models in expert team decision making", in N.J. CASTELLAN (ed.) *Current Issues in Individual and Group Decision Making*. Hillsdale (NJ, USA): Erlbaum, pp.221-246, 1993.
- [CHA 19] CHAKRABORTI T., KULKARNI A., SREEDHARAN S., SMITH D., KAMBHAMPATI S. "Explicability?, Predictability?, Transparency?, Privacy?, Security? the emerging landscape of interpretable agent behavior", *39th International Conference on Automated Planning and Scheduling*, Berkeley (CA, USA): 11-15 July 2019. In J. BENTON, N. LIPOVETZKY, E. ONAINDIA, D.E. SMITH, S. SRIVASTAVA (eds.) *Proessings of 39th ICAPS 29*, Menlo Park (CA, USA): AAAI Press, pp.86-96, 2019.

- [CHE 14] CHEN J.Y.C., BARNES M.J. “Human-agent teaming for multirobot control: a review of the human factors issues”, *IEEE Transactions on Human-Machine Systems*, vol.44, n°1, pp.13-29, 2014.
- [CHE 18] CHEN J.Y., LAKHMANI S.G., STOWERS K., SELKOWITZ A.R., WRIGHT J.L., BARNES M. Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical Issues in Ergonomics Science*, vol.19, n°3 – Human-Autonomy Teaming, pp.259-282, 2018.
- [CHI 20] CHIEN S., LEWIS M., SYCARA K., LIU J., KUMRU A. “Influence of culture, transparency, trust, and degree of automation on automation Use”, *IEEE Transactions on Human-Machine Systems*. vol.50, n°3, pp.205-214, 2020.
- [CHI 12] CHIEN S., MEHROTRA S., BROOKS N., LEWIS M., SYCARA K. “Scheduling operator attention for multi-robot control”, *IROS’12 IEEE/RSJ International Conference on Intelligent Robots and Systems 2012*, Vilamoura (PT): 7-12 October 2012. In *Proceedings of the IROS’12*. Red Hook (NY, USA): Curran Associates, Inc. IEEE Conference Proceedings, vol.5, pp.473-479, 2012.
- [COO 13] COOKE N.J., GORMAN J.C., MYERS C.W., DURAN J.L. “Interactive team cognition”, *Cognitive Sciences*, vol.37, n°2, pp.255–285, 2013.
- [CHR 17] CHRISTENSEN J.C., LYONS J.B. “Trust between humans and learning machines: developing the gray box”, *American Society of Mechanical Engineers*, vol.139, n°6, pp.S9-S13, 2017.
- [CUM 14] CUMMINGS M.M. “Man versus machine or man+ machine?”, *IEEE Intelligent Systems*, vol. 29, pp.62–69, 2014.
- [DEL 11] DE LA TORRE F., COHN J.F. “Facial expression analysis”, in T. MOESLUND, A. HILTON, V. KRÜGER, L. SIGAL (eds.) *Visual Analysis of Humans*. London (UK): Springer, 2011.
- [DEV 18] DE VISSER E., PAK R., SHAW T. “From ‘automation’ to ‘autonomy’: the importance of trust repair in human-machine interaction”. *Ergonomics*, vol.61, n°10, pp.1409-1427, 2018.
- [DEM 17] DEMIR M., MCNEESE N.J., COOKE, N.J. “Team situation awareness within the context of human-autonomy teaming”. *Cognitive Systems Research*, vol.46, pp.3-12, 2017.
- [DIE 13] DIEHL J.J., SCHMITT L.M., VILLANO M., CROWELL C.R. “The clinical use of robots for individuals with autism spectrum disorders: a critical review”, *Research in Autism Spectrum Disorders*, vol.6, n°1, pp.249-262, 2013.
- [DIR 99] DIRKS K.T. “The effects of interpersonal trust on work group performance”, *Journal of Applied Psychology*, n°84, n°3, pp.445-455, 1999.
- [DOR 12] DORNEICH M.C., VERVERS P.M., MATHAN S., WHITLOW S., HAYES C.C. “Considering etiquette in the design of an adaptive system”, *Journal of Cognitive Engineering and Decision Making*, vol.6, n°2, pp.243-265, 2012.
- [DRA 17] DRAPER M., CALHOUN G., HANSEN M., DOUGLASS S., SPRIGGS S., PATZEK M., ROWE A., EVANS D., RUFF H., BEHYMER K., HOWARD M., BEARDEN G., FROST, E. “Intelligent Multi-Unmanned Vehicle Planner with Adaptive Collaborative/Control Technologies (Impact)”, *19th International Symposium on Aviation Psychology – ISAP’17. Proceedings of the 19th International Symposium on Aviation Psychology*, Red Hook (NY, USA): Curran Associates Inc., pp.226-231, 2017.
- [EDM 99] EDMONDSON A. “Psychological safety and learning behavior in work teams”, *Administrative Science Quarterly*, vol.44, n°2, pp.350-383, 1999.
- [END 15] ENDSLEY M. *Technology Horizons: A Vision for Air Force Technology and Science 2010-2030*. Washington (DC, USA): Office of the US Air Force Chief Scientist.
- [FES 50] FESTINGER L. “Informal social communication”, *Psychological Review*, vol.57, n°5, pp.271-282, 1950.
- [FLE 19] FLEMISCH F., ABBINK D., ITOH M., PACAUX-LEMOINE M.P., WESSEL, G. “Shared control is the sharp end of cooperation: framework of joint action, shared control and human machine cooperation”, in F. FLEMISCH, D. ABBINK, M. ITOH, M. PACAUX-LEMOINE (eds.) *Cognition, Technology & Work*, Special Issue *Shared and Cooperative Control of Safety Critical Systems*. Kyoto (JP): Elsevier, vol.21, n°4, pp.553-554, 2019.
- [FRA 17] FRAUNE M.R., SABANOVIC S., SMITH E.R. “Teammates first: favoring ingroup robots over outgroup humans,” *26th IEEE RO-MAN2017 – International Symposium on Robot and Human Interaction and Communication*, 28 August - 1 September 2017, Lisbon (PT). *Proceedings of the 26th IEEE International Symposium on Robot and Human Interaction and Communication*, Washington (DC, USA): IEEE pub., pp.1432-1437, 2017.
- [GOS 89] GOSENPUD, J. “The prediction of simulation performance as it is affected by time”, *Simulation & Games*, vol.20, n°3, pp.319-350, 1989.

- [GRA 07] GRATCH J., WANG N., GERTEN J., FAST E., DUFFY R. “Creating rapport with virtual agents,” *IVA’2017 7th International Conference on Intelligent Virtual Agents*. Paris (FR) 17-19 September 2007. In C. PELACHAUD, J.C. MARTIN, E. ANDRÉ, G. CHOLLET, K. KARPOUZIS, D. PELÉ (eds.) *Proceedings of IVA’2017 Creating Rapport with Virtual Agents*. Berlin (DL): Springer, pp.125-138, 2007.
- [HAN 06] HANUMANTHARAO S., GRABOWSKI M. “Effects of introducing collaborative technology on communications in a distributed safety-critical system”, *International Journal of Human Computing Studies*, vol.64, n°8, 714-726, 2006.
- [HAN 13] HANCOCK P.A., JAGACINSKI R.J., PARASURAMAN R., WICKENS C.D., WILSON G.F., KABER D.B. “Human-automation interaction research: past, present, and future”, *Ergonomics in Design: The Quarterly of Human Factors Applications* vol.21, n°9, pp.9-14, 2013.
- [HAN 16] HANSEN M., CALHOUN G., DOUGLASS S., EVANS D. “Courses of action display for multi-unmanned vehicle control: a multi-disciplinary approach”, *Association for the Advancement of Artificial Intelligence 2016 Fall Symposium Series: Cross-Disciplinary Challenges for Autonomous Systems*, Arlington (VI, USA) 17-19 November 2016. In G. SUKTHANKAR, C. GEIB (eds.) *Proceedings of the 2016 AAAI Fall Symposium Series – Technical Report FS-16-03*, Palo Alto (CA, USA): The AAAI Press, pp.179-184, 2016.
- [HAR 98] HARRISON D.A., PRICE K.H., BELL M.P. “Beyond relational demography: Time and the effects of surface- and deep-level diversity on work group cohesion”, *Academy of Management Journal*, vol.41, n°1, pp.96-107, 1998.
- [HAS 06] HASLAM N. “Dehumanization: an integrative review”. *Personality and Social Psychology Review*, vol.10, n°3, pp.252-264, 2006.
- [HOF 07] HOFFMAN G., BREAZEAL C. “Effects of anticipatory action on human-robot teamwork efficiency, fluency, and perception of team”, *2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Arlington (VI, USA) 10-12 March 2007. In C. BREAZEAL, A.C. SCHULTZ (eds.) *Proceedings of the 2007 HRI’07*, New York (NY, USA): Association for Computing Machinery, pp.1-8, 2007).
- [HOF 15] HOFF K.A., BASHIR M. “Trust in automation: integrating empirical evidence on factors that influence trust”, *Human Factors*, vol.57, n°3, pp.407-434, 2015.
- [HUG 20] HUGHES D., AGARWAL A., GUO Y., SYCARA K. “Inferring non-stationary preferences for human-agent teams”, *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, Naples (IT) 31 August - 4 September 2020. In *Proceedings of the IEEE 2020 RO-MAN*, Red Hook (NY, USA): Curran Associates Inc., pp.1178-1185, 2020.
- [IQB 16] IQBAL T., RACK S., IEK, L.D. “Movement coordination in human-robot teams: a dynamical systems approach”, *IEEE Transactions on Robotics*, vol.32, n°4, pp.909-919, 2016.
- [IQB 17] IQBAL T., RIEK L.D. “Human-robot teaming: approaches from joint action and dynamical systems”, in A. Goswami, P. Vadakkepat (eds.) *Humanoid Robots: A reference*. Dordrecht (NL): Springer, pp.2-20, 2017.
- [IYE 18] IYER R., LI Y., LI H., LEWIS M., SUNDAR R., SYCARA K. “Transparency and explanation in deep reinforcement learning neural networks,” *AIES’18 – AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society*, New Orleans (LA, USA).2–7 February 2018. In J. Furman, G. Marchant, H. Price (eds.) *Proceedings of the AIES’18 Conference*. New York (NY, USA): Association for Computing Machinery, pp.144-150, 2018.
- [JOH 94] JOHANNESSEN L.J., COOK R.I., WOODS D.D. “Cooperative communications in dynamic fault management”, *Advances in Human Factors in Simulation and Modeling: Proceedings of the AHFE 1994 International Conference on Human Factors in Simulation and Modeling*. 24-28 July, 1994, Washington (D., USA). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol.38, pp.225-229. Los Angeles (CA, USA): Sage mpublishations, 1994.
- [JOH 11] JOHNSON N.D., MISLIN A.A. “Trust games: a meta-analysis”, *Journal of Economic Psychology*, vol.32, n°5, pp.865-889, 2011.
- [JOH 19] JOHNSON M., VERA A.H. “No AI is an island: the case for teaming intelligence”, *AI Magazine*, vol.40, n°1, pp.16-28, 2019.
- [KAN 10] KANG S.H., GRATCH J. “Virtual humans elicit socially anxious interactions’ verbal self-disclosure”, *Computer Animation & Virtual Worlds*, vol.21, n°3-4, pp.473-482, 2010.
- [KAS 10] KASPAROV G. *The Chess Master and the Computer*. Cambridge (Le MA, USA): MIT Press, 2010.
- [KOZ 00] KOZLOWSKI S.W.J., KLEIN K.J. “A multilevel approach to theory and research in organizations: contextual, temporal, and emergent processes,” in S.W.J. KOZLOWSKI, K.J. KLEIN (eds.) *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions, and New Directions*. San Francisco (CA, USA): Jossey-Bass, pp.3-90, 2000.

- [KOZ 03] KOZLOWSKI S.W.J., BELL B.S. “Work groups and teams in organizations”, in W.C. Borman, D.R. Ilgen, R.J. Klimoski (eds.) *Handbook of Psychology: Industrial and Organizational psychology*, vol.12, pp.333-375. New-York (NY, USA): Wiley-Blackwell, 2003.
- [KOZ 15] KOZLOWSKI S.W.J., CHENG C.H., PERRY S.R., PEARCE M., DIXON A.J., SANTORO J.M. “Capturing team process dynamics”, In *panel 333 W.B. VESSEY (chair) Teams on ICE: Team research in spaceflight analogs*. Symposium presented at the 30th Annual Conference of the Society for Industrial and Organizational Psychology, Philadelphia (PA, USA) 23-25 April 2015.
- [LAS 15] LASOTA P.A., HAH, J.A. “Analyzing the effects of human-aware motion planning on close-proximity human-robot collaboration”, *Human Factors*, vol.57, n°1, pp.21-33, 2015.
- [LEE 04] LEE J.D., SEE K.A. “Trust in automation: designing for appropriate reliance”, *Human Factors*, vol.46, pp.50-80, 2004.
- [LEE 93] LEWIS M., SYCARA K. “Reaching informed agreement in multi-specialist cooperation”, *Group Decision and Negotiation*, vol.2, pp.279-300, 1993.
- [LIH 20] LI H., LEWIS M., NI T., AGARWAL S., HUGHES D., SYCARA K. “Team synchronization in coop-space fortress”, *Human Factors and Ergonomics Society Annual Meeting (HFES 2020)*, Pittsburgh (PA, USA) 5-9 October 2020. In *Proceedings of the Human Factors and Ergonomics Annual Meetings*. Red Hook (NY, USA): Curran Associates Inc., vol.64, n°1, pp.82-86, 2020.
- [LIJ 16] LI J., CHO M. J., ZHAO X., JU W., MALLE B.F. “From trolley to autonomous vehicle: Perceptions of responsibility and moral norms in traffic accidents with self-driving cars”, *Society of Automotive Engineers World Congress*, Detroit (MI, USA), 12-14 April 2016. *SAE 2016 Proceedings* : Stanford (CA, USA): Stanford University Press, Technical Paper 2016-01-0164, pp.1-8, 2016.
- [LIU 16] LIU C., HARICK J., FISAC J., DRAGAN A., HEDRICK K., SASTRY S., GRIFFITHS T.L. “Goal inference improves objective and perceived performance in human-robot collaboration”, *AAMAS'16: International Conference on Agents and Multiagent Systems*. Singapore (SG) 9-13 May, 2016. In C.M. JONKER, S. MARSELLA (eds.) *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016)*, Richland (SC, USA): ACM (Association for Computing Machinery), pp.940-949, 2016.
- [LIU 19a] LIU R., CAI Z., LEWIS M., LYONS J., SYCARA K. “Trust repair in human-swarm teams”, *28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2019) : Responsible Robotics and AI for the Real World*, New Delhi (IN) 14-17 October 2019. In *IEEE Proceedings of RO-MAN'19*, Red Hook (NY, USA): Curran Associates, Inc. vol.2, pp.249-254, 2019.
- [LIU 19b] LIU R., LUO W., CHANDARANA M., NAM C., LEWIS M., SYCARA K. “Trust Aware Behavior Reflection for Robot Swarm Self-healing”, *18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019) – Session 1B: Multi-Robot System*, Montreal (QC, CA): 13–17 May 2019. In N. AGMON, M. E. TAYLOR, E. ELKIND, M. VELOSO (eds.) *Proceedings of AAMAS'19*, New York (NY, USA): Associates, Inc., vol.1, pp.122-130, 2019.
- [LUO 11] LUO L., CHAKRABORTY N., SYCARA K. “Multi-robot algorithm for tasks with set precedence constraints”, *ICRA 2011 – IEEE International Conference on Robotics and Automation*, Shanghai (CN) 9-13 May 2011. In A. BICCHI (ed.) *IEEE Proceedings ICRA 2011*, Red Hook (NY, USA): Curran Associates Inc., pp.2526-2533, 2011.
- [LYO 13] LYONS J.B. “Being transparent about transparency: a model for human- robot interaction”, *2012 Association for the Advancement of Artificial Intelligence Symposia*. Palo Alto (CA, USA) 25-27 March 2013. In D. SOFGE, G. J. KRUIFF, W. F. LAWLESS (eds.) *Trust and Autonomous Systems: Papers from the AAAI Spring Symposium*. Menlo Park (CA, USA): AAAI Press, pp.48-53, 2013.
- [LYO 18] LYONS J.B., HO N.T., VAN ABEL A.L., HOFFMANN L.C., FERGUESON W.E., SADLER G.G., GRIGSBY M.A., BURNS A.C. “Exploring trust barriers to future autonomy: a qualitative look”, *Advances in Human Factors in Simulation and Modeling: Proceedings of the AHFE 2017 International Conference on Human Factors in Simulation and Modeling*, 17–21 July 2017, Los Angeles (CA, USA). In D.N. Cassenti (ed.) *Proceedings*. New-York (NY, USA): Springer International Publishing, pp.3-11, 2018.
- [LYO 19] LYONS J.B., WYNNE K.T., MAHONEY S., ROEBKE M.A. “Trust and human-machine teaming: a qualitative study,” in W. LAWLESS, R. MITTU, D. SOFGE, I. MOSKOWITZ, S. RUSSELL (eds.) *Artificial Intelligence for the Internet of Everything*. Cambridge (MA, USA): Academic Press/Elsevier, pp.101-116, 2019.
- [LYO 21a] LYONS J.B., VO T., WYNNE K.T., MAHONEY S., NAM C.S., GALLIMORE D. “Trusting autonomous security robots: the role of reliability and stated social intent”, *Human Factor*, vol.63, n°4, pp.603-618, 2021.
- [LYO 21b] LYONS J.B., WYNNE K.T. “Human-machine teaming: evaluating dimensions using narratives”, *Human-Intelligent Systems Integration*, vol.3, n°2, pp.1-9, 2021.

- [MAS 13] MASIELLO M.J. (Major General) *Air Force Research Laboratory Autonomy Science and Technology Strategy*. Dayton (OH, USA): Wright-Patterson Air Force Research Laboratory. 2013.
- [MAT 96] MATHESON H., MATHES S., MURRAY M. "Group cohesion of female intercollegiate coacting and interacting teams across a competitive season", *International Journal of Sport Psychology*, vol.27, pp.37-49, 1996.
- [MAT 00] MATHIEU J.E., HEFFNER T.S., GOODWIN G.F., SALAS E., CANNON-BOWERS J.A. "The influence of shared mental models on team process and performance", *Journal of Applied. Psychology*, vol.85, n°2, pp.273-283, 2000.
- [MAY 95] MAYER R.C., DAVIS J.H. SCHOORMAN F.D. "An integrated model of organizational trust". *Academy of Management Review*, n°20, n°3, pp.709-734, 1995.
- [MCG 64] MCGRATH J.E. *Social Psychology: A Brief Introduction*. New York (NY, USA): Holt, Rinehart & Winston, 1964.
- [MCK 17] MCKENDRICK R., MEHTA R., AYAZ H., SCHELDROP M., PARASURAMAN R. Prefrontal hemodynamics of physical activity and environmental complexity during cognitive work. *Hum. Fact.* 59, 147-162 (2017).. doi: 10.1177/0018720816675053
- [MCN 18] MCNEESE N.J., DEMIR M., COOKE N.J., MYERS C.W. "Teaming with a synthetic teammate: insights into human-autonomy teaming", *Human Factor*, vol.60, n°2, pp.262-273, 2018.
- [MNI 15] MNIH V., KAVUKCUOGLU K., SILVER D., RUSU A.A., VENESS J., BELLEMARE M.G., GRAVES A., RIEDMILLER M., FIDJELAND A.K., OSTROVSKI G., PETERSEN S., BEATTIE C., SADIK A., ANTONOGLU I., KING H., KUMARAN D., WIERSTRA D., LEGG S., HASSABIS D. "Human level control through deep reinforcement learning", *Nature*, vol.518, pp.529-533, 2015.
- [MUY 18] MU Y., CERRITOS C., KHAN F. "Neural mechanisms underlying interpersonal coordination: a review of hyperscanning research". *Social and Personality Psychology Compass*, 12:e12421, 2018.
- [NAM 20] NAM C., WALKER P., LI H., LEWIS M., SYCARA K. "Models of trust in human control of swarms with varied levels of autonomy". *IEEE Transactions on Human-Machine Systems*, vol.50, pp.194-204, 2020.
- [NAS 94] NASS C., STEUER J., TAUBER E. "Computers are social actors", *CHI'94 - SIGCHI Conference on Human Factors in Computing Systems*, Boston (MA, USA) 24-28 April 1994. In B. ADELSON, S. DUMAIS, J. OLSON (eds.) *Proceedings of the CHI'94*, pp.72-78. New York (NY, USA): Association for Computing Machinery editions, 1994.
- [NAS 20] NASS C., MOON Y. "Machines and mindlessness: social responses to computers", *Journal of Social Issues*, vol.56, n°1, pp.81-103, 2000.
- [ONN 14] ONNASCH L., WICKENS C.D., LI H., MANZEY D. "Human performance consequences of stages and levels of automation: an integrated meta-analysis", *Human Factors*, vol.56, pp.476-488, 2014
- [ORA 90] ORASANU J. *Shared Mental Models and Crew Decision Making*. (Cognitive Science Laboratory – Technical Report n°46). Princeton (NJ, USA): Princeton University Press, 1990.
- [PAN 19] PANGANIBAN A.R., MATTHEWS G., LONG M. "Transparency in autonomous teammates: intention to support as teaming information". *Journal of Cognitive Engineering and Decision Making*, vol.14, n°2, pp.174-190, 2019.
- [PAR 04] PARASURAMAN R., MILLER C. "Trust and etiquette in high-criticality automated systems", *Communications of the ACM* 47, pp.51-55. 2004.
- [PAR 97] PARASURAMAN R., RILEY V. "Humans and automation: use, misuse, disuse, abuse", *Human Factors*, vol.39, n°2, pp.230-253. doi: 10.1518/001872097778543886. (1997).
- [PAR 20] PARASURAMAN R., SHERIDAN T.B., WICKENS C.D. "A model for types and levels of human interaction with automation", *IEEE Transactions on Systems, Man, and Cybernetics Systems - Part A*, vol.30, pp.573-583, 2000.
- [PET 19] PETERS N.S., BRADLEY G.C., MARSHALL-BRADLEY T. "Task boundary inference via topic modeling to predict interruption timings for human- machine teaming", *2nd International Conference on Intelligent Human Systems Integration (IHSI 2019): Integrating People and Intelligent Systems*, San Diego (CA, USA) 7-10 February 2019, In W. KARWOWSKI, T. AHRAH (eds.) *IHSI: International Conference on Intelligent Human Systems Integration. Advances in Intelligent Systems and Computing*, vol.903, New-York (NY, USA): Springer, 2019.
- [PHI 16] PHILLIPS M., HWANG V., CHITTA S., LIKHACHEV M. "Learning to plan for constrained manipulation from demonstrations", *Autonomous Robots*, vol.40, pp.109-124, 2016.
- [PIN 09] PINEDA J.A., HECHT E. "Mirroring and mu rhythm involvement in social cognition: are there dissociable subcomponents of theory of mind?", *Biological Psychology*, vol.80, n°3, pp.306-314, 2009.
- [RIT 18] RITTER E., FRANK E., OURY J.D. "ACT-R: A cognitive architecture for modeling cognition", vol.10, n°4, art.e1488.

- [SAB 13] ŠABANOVIC S., BENNETT C., CHANG W., HUBER L. “PARO robot affects diverse interaction modalities in group sensory therapy for older adults with dementia”, *2013 ICORR – IEEE 13th International Conference on Rehabilitation Robotics*, Poster session, 24-26 June 2013, Seattle (WA, USA). *Proceedings of the 2013 ICORR*, com.20., pp.449-454, Red Hook (NY, USA): Curran Associates Inc., 2013.
- [SAL 04] SALAS E., FIORE S.M. “Why team cognition? An overview”, in E. SALAS, S.M. FIORE (eds.) *Team Cognition: Understanding the Factors That Drive Process and Performance*. Washington (DC, USA): American Psychological Association, 2004.
- [SCH 17] SCHAEFER K.E., STRAUB E.R., CHEN J.Y.C., PUTNEY J., EVANS A.W. “Communicating intent to develop shared situation awareness and engender trust in human-agent teams”, *Cognitive Systems Research*, vol.46, pp.26-39, 2017.
- [SCH 18] SCHARRE P. *Autonomous Weapons and the Future of War: Army of None*. New York (NY, USA): W.W. Norton and Company, 2018.
- [SCH 17] SCHUMACHER C., PATZEK M., AHA D., FALCO R., NOVAK M., ORR T., HUGHES T., VAN PELT J. “Autonomy for Air Combat Missions (ATACM): Autonomy Research Pilot Initiative (ARPI)”, *Technical Report – 711th Human Performance Wing – US Air Force School of Aerospace Medicine Fact Sheet*. Wright-Patterson Air-Force base (OH, USA): Air Force Research Laboratory, 2017.
- [SER 05] SERVA M.A., FULLER M.A., MAYER R.C. “The reciprocal nature of trust: a longitudinal study of interacting teams”, *Journal of Organizational Behavior*, vol.26, n°6, pp.625-648, 2005.
- [SHE 78] SHERIDAN T.B., VERPLANK W.L. *Human and Computer Control of Undersea Teleoperators*. Technical Report. DTIC.MIL. Cambridge (MA, USA): Massachusetts Institute of Technology, 1978.
- [STA 05] STANTON N., YOUNG M. “Driver behaviour with adaptive cruise control”, *Ergonomics*, vol.48, n°10, pp.1294-1313, 2005.
- [SUK 06] SUKTHANKAR G., SYCARA K. “Simultaneous Team Assignment and Behavior Recognition From Spatio-Temporal Agent Traces”, *21th National Conference on Artificial Intelligence AAAI’06*, Session *Multiagent Systems*. Boston (MA, USA) 16-20 July 2006. In *AAAI’06 Proceedings*. Washington (DC, USA): Association for the Advancement of Artificial Intelligence Press, pp.716-721, 2006.
- [SUK 08] SUKTHANKAR G., SYCARA K. “Robust and efficient plan recognition for dynamic multi-agent teams”, *AAMAS’08 – 7th International Conference on Autonomous Agents and Multi Agent Systems*, Estoril (PT) 12-16 May 2008. In L. PADGHAM, D. PARKES, J. MÜLLER, S. PARSONS (eds.) *Proceedings of the Autonomous Agents and Multi-Agent Systems*, Richland (SC, USA): International Foundation for Autonomous Agents and Multiagent Systems, pp.1383-1386, 2008.
- [SUK 09] SUKTHANKAR G., SYCARA K., GIAMPAPA J., BURNETT C., PREECE A. “An analysis of salient communications for agent support of human teams”, in V. DIGNUM (ed.) *Multi Agent Systems: Semantics and Dynamics of Organizational Models IGI*. New York (NY, USA): Hershey, pp.256-259, 2009.
- [SYC 03] SYCARA K., PAOLUCCI M., VAN VELSEN M., GIAMPAPA J. “The RETSINA multiagent infrastructure”, *Journal of Autonomous Agents and Multi-Agent Systems*, vol.7, n°1, pp.29-48, 2003.
- [SYC 04] SYCARA K., LEWIS M. “Integrating intelligent agents into human teams”. In E. SALAS, S. M. FIORE (eds.) *Team Cognition: Understanding the Factors That Drive Process and Performance*. Washington (DC, USA): American Psychological Association, 203-231, 2004.
- [SYC 20] SYCARA K., HUGHES D., LI H., LEWIS M., LAUHARATANAHIRUN N. “Adaptation in human-autonomy teamwork”, *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*, Roma (IT) 7-9 September 2020. In G. FORTINO, F.Y WANG, A. NURNBERGER, D. KABER, R. FALCONE, D. MENDONCA, Z. YU, A. GUERRIERI (eds.) *Proceedings of the 2020 IEEE ICHMS*, Art.9209410.
- [TAN 92] TANNENBAUM, S. I., BEARD, R. L., AND SALAS, E. “Team building and its influence on team effectiveness: an examination of conceptual and empirical developments”, in K. KELLY (ed.) *Issues, Theory, and Research in Industrial Organizational Psychology*. Amsterdam (NL): Elsevier, pp.117-153, 1992.
- [THI 91] THINES C. *Michotte’s Experimental Phenomenology of Perception*. Hillsdale (NJ, USA): Erlbaum, 1991.
- [TOM 12] TOMASELLO M., MELIS A., TENNIE C., WYMAN E., HERRMANN E. “Two key steps in the evolution of human cooperation: the interdependence hypothesis”. *Curr.ent Anthropology*. vol.53, pp.673–692, 2012.
- [TUC 65] TUCKMAN B.W. “Developmental sequence in small groups”, *Psychological Bulletin*, vol.63, n°6, pp.384-399, 1965.

- [URB 93] URBAN J.M., BOWERS C.A., MONDAY S.D., MORGAN B.B. "Effects of workload on communication processes in decision making teams: an empirical study with implications for training". *Human Factor and Ergonomic Society Annual Meeting — 37th Annual Meeting: Designing for Diversity*, 11-15 October 1993, Seattle (WA, USA). *Proceedings of the HFES annual Meetings*, vol.37, n°9, pp.1233-1237, Thousand Oaks (CA, USA): Sage Publications, 1993.
- [VAG 16] VAGIA M., TRANSETH A.A., FJERDINGER S.A. "A literature review on the level of automation during the years: what are the different taxonomies that have been proposed?", *Applied Ergonomics*, vol.53, part.A pp.190-202, 2016.
- [VAN 18] VAN DER HOUT J.J.J., DAVIS O.C., WEGGEMAN M.C.D.P. "The conceptualization of team flow", *The Journal of Psychology: Interdisciplinary and Applied*, vol.152, n°6, pp.388-423, 2018.
- [WAN 07] WANG J., LEWIS M. "Assessing coordination overhead in control of robot teams", *IEEE International Conference on Systems, Man and Cybernetics*, Montréal (QC, CA), 7-10 October 2007. *Proceedings of 2007 IEEE ICSMC*. Washington (DC, USA): IEEE Pub., pp.2645-2649, 2007.
- [WYN 18] WYNNE K.T., LYONS J.B. "An integrative model of autonomous agent teammate-likeness", *Theoretical Issues in Ergonomics Science*, vol.19, pp.353-374, 2018.
- [WYN 19] Wynne K.T., Lyons J.B. "Autonomous Agent teammate-likeness: scale development and validation," in *Virtual, Augmented and Mixed Reality. Applications and Case Studies : 11th International Conference, VAMR 2019, Held as Part of the 21st HCI International Conference, HCII 2019*. In J. CHEN, G. FRAGOMENI (eds.) *Lecture Notes in Computer Science*, vol.11575, pp.199-213, 2019.
- [XIE 20] XIE H., KARIPIDIS I.I., HOWELL A., SCHREIER, M., SHEAU K.E., MANCHANDA M.K., AYUB R., GLOVER G.H., JUNG M., REISS A.L., SAGGAR M. "Finding the neural correlates of collaboration using a three-person fMRI hyperscanning paradigm". *21st HCI International Conference, HCII 2019*, Orlando (FL,USA): 26-31 July 2019. *Proceedings of the National Academy of Sciences of the United States of America*, vol.117, pp.23066-23072, 2020.

Conflit d'intérêts :

Les auteurs déclarent que la recherche a été menée en l'absence de toute relation commerciale ou financière qui pourrait être interprétée comme un conflit d'intérêts potentiel.

Copyright © 2021 Lyons, Sycara, Lewis and Capiola. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.