



HAL
open science

Effets de la transparence sur la confiance dans la coopération humain-machine : application à l'industrie 4.0

Loïck Simon

► **To cite this version:**

Loïck Simon. Effets de la transparence sur la confiance dans la coopération humain-machine : application à l'industrie 4.0. Psychologie. Université Bretagne Sud, 2023. Français. NNT: . tel-04342683

HAL Id: tel-04342683

<https://hal.science/tel-04342683>

Submitted on 13 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

COLLEGE EDUCATION, LANGAGES

DOCTORAL INTERACTIONS, COGNITION

BRETAGNE CLINIQUE, EXPERTISE



Thèse de doctorat de

L'UNIVERSITE BRETAGNE SUD

Ecole Doctorale n° 646

Education, Langages, Interactions, Cognition, Clinique, Expertise

Spécialité : Ergonomie

Par

« **Loïck SIMON** »

« **Effets de la transparence sur la confiance dans la coopération
humain-machine : application à l'industrie 4.0** »

Thèse présentée et soutenue à « Lorient », le « 20/10/2023 »

Unité de recherche : Équipe FHOOX, Lab-STICC, UMR 6285

Thèse N° : 669

Rapporteurs avant soutenance :

Émilie, Loup-Escande Professeure des Universités, Université de Picardie Jules Verne
Julien, Cegarra Professeur des Universités, Institut National Universitaire Champollion

Composition du Jury :

Attention, en cas d'absence d'un des membres du Jury le jour de la soutenance, la composition du jury doit être revue pour s'assurer qu'elle est conforme et devra être répercutée sur la couverture de thèse

Président :	Jean-Marie Burkhardt	Directeur de recherche, Université Gustave Eiffel
Examinatrice :	Émilie, Loup-Escande	Professeure des Universités, Université de Picardie Jules Verne
Examineur :	Julien, Cegarra	Professeur des Universités, Institut National Universitaire Champollion
Examinatrice :	Marie-Pierre, Pacaux-Lemoine	IR-HDR, Université Polytechnique Hauts-de-France
Examinatrice :	Isabelle, Milleville	Chargée de recherche-HDR, Ecole Centrale de Nantes

Dir. de thèse :	Philippe, RAUFFET	Professeur des Universités, Université Bretagne Sud
Co-encadrant. de thèse :	Clément, Guerin	Maître de conférences, Université Bretagne Sud

Titre : Effets de la transparence sur la confiance dans la coopération humain-machine : application à l'industrie 4.0

Mots clés : Confiance, Coopération Humain-Machine, Industrie 4.0, Prise de décision, Systèmes Cyber-Physique, Transparence

Résumé : La transition vers l'industrie 4.0 introduit des changements sur la place des opérateurs dans l'industrie. Les nouvelles fonctionnalités des systèmes cybers-physiques (CPS) leur octroient la capacité de formuler des propositions aux opérateurs. Dès lors, les opérateurs doivent coopérer avec ces CPS et prendre la décision de leur faire confiance, ou non, en acceptant ou en refusant leur proposition. L'objectif de cette thèse est de comprendre le rôle de la confiance lors de la prise de décision de l'opérateur d'une part, et d'autre part de comprendre quelles sont les capacités de communication des CPS permettant à l'opérateur de calibrer sa confiance.

Au travers trois expérimentations nous avons testé différentes configurations de transparence des CPS et des risques des situations. La troisième expérimentation comporte une transparence adaptative.

Les résultats de ces expérimentations confirment que la confiance se structure autour de plusieurs composantes, chacune d'entre elles étant plus ou moins déterminantes de la prise de décision. Nous montrons que les informations sur les risques jouent un rôle modérateur dans l'acceptation. Nous mettons en avant le rôle important que joue la transparence dans le processus de la confiance et de l'acceptation de la proposition par l'opérateur. Enfin, cette transparence peut être utilisée de façon dynamique pour aider l'opérateur à calibrer sa confiance.

Les éléments avancés dans cette thèse ont pour objectif de remettre l'opérateur au centre des processus de conception des CPS. L'utilisation de la transparence permettra au CPS d'aider les opérateurs à calibrer leur confiance et d'améliorer la coopération humain-machine.

Title: Effects of transparency on trust in human-machine cooperation: application to industry 4.0

Keywords: Cooperation, Industry 4.0, Cyber-Physical System, Decision Making, Human-Machine, Transparency, Trust

Abstract: Transition to industry 4.0 induces many changes on operators' status in the workplace. Cyber-physical systems (CPS) have new functionalities and are able to formulate propositions to operators. Therefore, operators need to cooperate with the CPS by trusting them, or not, with the acceptance, or the denying, of their proposition. The thesis objective is to understand the role of trust in this decision making. We also want to understand what CPS' communication capacity helps operators to calibrate their trust.

We did three experiments where we tested different CPS' transparency configuration and risk situations. The third experiment used a dynamic transparency.

Experimental results confirm that trust is structured with multiple layers. Each layer predicts with more, or less precision the operators' decision making. We show that risk information is a moderator of acceptance. We also show that transparency has an effect on trust process and the acceptance of a proposition. This transparency can be used dynamically in order to help the operator calibrating his trust.

The elements put forward in this thesis aim to put the operator back at the centre of the CPS design process. The use of transparency will enable the CPS to help operators calibrate their confidence and improve human-machine cooperation.

Remerciements :

Je tenais avant toute chose à remercier mes encadrants Philippe Rauffet et Clément Guerin. Ils ont su, tout au long de ce parcours de thèse, me soutenir, aussi bien humainement, que professionnellement et scientifiquement. Ils ont su me conseiller, m'éclairer et proposer un cadre de travail épanouissant. Alors pour toutes leurs qualités et pour ces dernières années à travailler ensemble je vous suis reconnaissant.

Je tenais également à remercier Isabelle Milleville et Emilie Poirson pour m'avoir accompagné dans le cadre des CSI. Ces jalons, importants au sein d'un parcours de thèse, ont été l'occasion d'échanges scientifiques qui ont contribué à l'avancement de mes travaux.

Je voulais également remercier Emilie Loup-Escande et Julien Cegarra pour avoir accepté d'être les rapporteur.e.s de cette thèse. Je les remercie pour leurs retours sur ce manuscrit. Je remercie également Jean-Marie Burckhardt pour avoir présidé le jury ainsi que Marie-Pierre Pacaux-Lemoine, et à nouveau Isabelle Milleville, pour leur participation à la soutenance. Je tenais également à les remercier pour nos échanges lors de la soutenance.

Je remercie Christine, Laurent, Gaël, Natalie et Pascal pour leur accueil et l'intégration dans l'équipe de recherche FHOOX. Ils ont su prodiguer nombre de judicieux conseils lors de mes présentations et de nos échanges. Merci également à toute l'équipe des doctorants pour les échanges et le soutien mutuel. Merci à Ahmed, Aurelian, Ayoub, Mor, Naomi et Yvan. Je tiens tout particulièrement à remercier la désormais docteure Tabatha sans qui, littéralement, cette thèse n'aurait jamais débuté.

Merci à toutes les personnes avec qui j'ai eu la joie de travailler sur le projet Seanatic. Merci donc à Cédric, Florian, Guy, Jean-Philippe et Yohann du Lab-Sticc, Irène, Jean-Baptiste et Jérémy de l'ENSM, Goulven, Laure, Julie, Oussama et Thomas d'Azimut. Les recherches présentées dans ce manuscrit n'auraient jamais vu le jour sans nos travaux communs.

Merci également à tous les personnels administratifs et techniques de l'université et de l'école doctorale, sans qui aucun travail de thèse ne pourrait se faire. Merci particulièrement à Denis, Laurent, Marion, Thierry et Tiffany pour les moments détentes aux cafés et aux repas. Merci également à tous les enseignants de l'ENSIBS avec qui j'ai eu le loisir d'échanger durant ces quelques années à travailler à leurs côtés. Enfin, merci à tous les participants du groupe d'anglais pour les échanges que l'on a pu avoir tous ces midis. Merci à Adel, Fanny, Maddie, Melissa, Naomi, Nathalie, Perrine et Sandrine.

Je remercie également tous les participants de mes études, les étudiants de l'École Nationale Supérieure d'Ingénieurs de Bretagne Sud ainsi que les étudiants de l'École Nationale Supérieure Maritime du Havre.

Merci à ma famille qui, de près ou de loin, a contribué à la réalisation de cette thèse, à commencer pour leur transmission du goût de la science, de la compréhension et de la curiosité. Merci également pour votre soutien dans mes études, phase préalable à tout travail de thèse.

Merci à tou.te.s mes ami.e.s qui m'ont aidé, Alexia, Anaïs, Beaudic, BoumBoum, Brice, Caroline, Cédric, Céline, Châtel, Charlotte, Gerbert, Gary, Godart, Justine, Kévin D, Kévin LB, Laura, Lauranne, Léa, Léna, Lucie, Nicolas Rou et Nicolas Roy, Manuela, Marine, Marius, Maxime, Quentin C, Quentin G, Tabatha et William. Merci pour les moments de détentes, les

randonnés, les vacances ou les travaux, tous ces moments de décompression essentiels au maintien d'une motivation, quasi-sans faille, sur ces dernières années.

Enfin, je tiens particulièrement à remercier Claire. Merci de m'accompagner tous les jours et de m'avoir soutenu tout au long de ce travail de thèse. Tu as toujours été présente lorsque j'en avais besoin et tu as su m'épauler lors des moments les plus complexes. Sans toi ce manuscrit n'aurait pas vu le jour. Merci d'être là avec moi.

« I believe that scientific knowledge has fractal properties, that no matter how much we learn, whatever is left, however small it may seem, is just as infinitely complex as the whole was to start with. That, I think, is the secret of the Universe. »

Isaac Asimov

Table des matières

Introduction générale	1
Partie théorique	
4.0 : confiance de l'opérateur et transparence du CPS	3
1. Chapitre 1 : Les nouveaux défis posés par l'industrie 4.0 pour les interactions entre l'humain et la machine	4
1.1. L'Industrie 4.0 : définitions et enjeux	5
1.2. Fonctions de l'industrie 4.0 et intégration des CPS	9
1.3. Prise de décision assistée par un CPS dans l'industrie 4.0	14
1.4. Cas d'application de la prise de décision assistée par un CPS	17
1.5. Synthèse du chapitre 1	20
2. Chapitre 2 : Coopération humain-machine pour la prise de décision assistée par un CPS : la confiance comme élément central	22
2.1. Les modèles de CHM pour la prise de décision assistée par un CPS	23
2.2. Les déterminants de l'opérateur dans la prise de décision assistée par un CPS	35
2.3. Modèle de prise de décision assistée par un CPS	45
2.4. Synthèse du chapitre 2	48
3. Chapitre 3 : Transparence, confiance et prise de décision assistée par un CPS	51
3.1. La transparence comme facilitateur de coopération	52
3.2. Vers une transparence adaptative	65
3.3. Synthèse du chapitre 3	68
4. Chapitre 4 : Méthodes et outils pour l'évaluation de la confiance	72
4.1. Mesurer la confiance apprise et l'acceptation	72
4.2. Mesurer la confiance situationnelle : conscience de la situation et la perception du risque	75
4.3. Mesurer la confiance situationnelle : charge mentale	78
4.4. Synthèse du chapitre 4	83
5. Chapitre 5 : Problématique	86
5.1. Objectifs des travaux	86
5.2. Verrou 1 : Compréhension du rôle des composantes de la confiance dans la prise de décision de l'opérateur	87
5.3. Verrou 2 : Compréhension des effets des modalités de la transparence	89
5.4. Verrou 3 : Compréhension des effets d'une transparence adaptative	90
Partie empirique	
Composantes de la confiance et modalités de transparence dans la prise de décision assistée par un CPS	92
6. Chapitre 6 : Confiance et transparence dans une prise de décision opérationnelle avec un CPS de fiabilité inconnue	96
6.1. Méthode	97
6.2. Résultats	106
6.3. Discussion	112
6.4. Conclusion	117
7. Chapitre 7 : Confiance et transparence dans une prise de décision tactique avec un CPS fiable	119
7.1. Méthode	120
7.2. Résultats	129
7.3. Discussion	136
7.4. Conclusion	141
8. Chapitre 8 : Adaptation dynamique de la transparence pour calibrer la confiance	144
8.1. Méthode	144

8.2.	Résultats	152
8.3.	Discussion	157
8.4.	Conclusion	160
9.	Chapitre 9 : Discussion générale	162
9.1.	Apports théoriques, méthodologiques et pratiques de la thèse	163
9.2.	Limites et perspectives	171
	<i>Bibliographie :</i>	177
	<i>Webographie :</i>	199
	<i>Médiagraphie :</i>	200
	<i>Logiciel :</i>	200
	<i>Annexes</i>	201
	Annexe A : Consentement libre et éclairé Humanism	201
	Annexe B : Questionnaire Humanism	204
	Annexe C : Consentement libre et éclairé Seanatic 1	206
	Annexe D : Échelles de confiance dispositionnelle	209
	Annexe E : Échelles de confiance apprise	212
	Annexe F : Échelles de confiance situationnelle	214
	Annexe G : Questions socio-démographique Seanatic 1	216
	Annexe H : WDA	217
	Annexe I : ConTA et SOCA	218
	Annexe J : Atelier co-conception	221
	Annexe K : Consentement libre et éclairé Seanatic 2	223
	Annexe L : Questions socio-démographique Seanatic 2	226

Table des tableaux

Tableau 2.1 : Classification des projets Humanism et Seanatic selon différentes caractéristiques de la coopération identifiées dans la littérature	33
Tableau 2.2 : Projet Humanism et Seanatic selon leurs interférences.....	34
Tableau 2.3 : Projet Humanism et Seanatic selon les modèles de coopération.	35
Tableau 3.1 : Dimensions de transparence selon le niveau de transparence appliqué à un CPS d'aide à la fabrication de pizza	61
Tableau 4.1 : Propriétés des méthodologies de mesure	72
Tableau 4.2 : Propriétés des mesures de confiance	75
Tableau 4.3 : Propriétés des mesures de conscience de la situation et de perception du risque	78
Tableau 4.4 : Propriétés des mesures de charge mentale	83
Tableau 4.5 : Propriétés des mesures utilisées selon les composantes de confiance	84
Tableau II.1: Points communs et différences entre les projets Humanism et Seanatic	93
Tableau 6.1 : Modalités de transparence dans les six conditions expérimentales ...	100
Tableau 6.2 : Régression logistique, effet des modalités A et EO sur l'acceptation .	107
Tableau 6.3 : Régression logistique ordinaire, effet des modalités de transparence sur la confiance apprise (A. Compréhensible, B. Fiable, C. Digne de confiance)	109
Tableau 6.4 : Régression logistique ordinaire, effet des modalités A et EO sur la perception du danger	110
Tableau 6.5 : Modèle linéaire mixte de la performance des modalités de transparence sur la tâche secondaire et le temps avant la prise de décision	111
Tableau 7.1 : Modalités de criticité de la situation	121
Tableau 7.2 : Matrice d'analyse des risques.....	122
Tableau 7.3 : Modalités de transparence de l'IA utilisées dans l'expérimentation ..	122
Tableau 7.4 : Conditions expérimentales selon la criticité de la situation et la transparence de l'IA.....	123
Tableau 7.5 : Prédicativité des modèles sur les composantes de la confiance en fonction de la transparence et de la criticité de la situation	134
Tableau 8.1 : Feedback de l'IA en fonction de la justesse du diagnostic de l'IA et du choix du participant	147
Tableau 8.2 : Effet du delta transparence et du delta feedback sur le delta d'acceptation	156

Table des figures

Figure 1.1: Frise chronologique des 4 générations d'industrie (inspirée de Christoph Roser sur AllAboutLean.com)	5
Figure 1.2 : Typologie des opérateurs 4.0 (issue de Romero et al., 2016b)	8
Figure 1.3 : Cartographie des fonctions de production (inspirée de Rauffet, et al., 2018)	9
Figure 1.4 : Proposition de cadre cognitif conceptuel de l'opérateur dans une usine intelligente (selon Madonna et al. 2019).....	15
Figure 1.5 : Complémentarités des capacités des opérateurs (en haut) et des CPS (en bas) selon les catégories de prise de décision (selon Jarrahi, 2018)	15
Figure 1.6 : Photo de la plateforme SCAP utilisée dans le projet Humanism.....	18
Figure 1.7: Photo du navire utilisé pour le projet Seanatic (Julien Talec, 2013)	19
Figure 1.8 : Positionnement des fonctions explorées dans Humanism et Seanatic (inspiré de Rauffet, Guerin, Chauvin & Martin, 2018).....	19
Figure 1.9 : Positionnement de l'opérateur Humanism et Seanatic au sein de la typologie de Romero et al. (2016b)	20
Figure 2.1 : Coopération verticale (Millot, 1999 dans Kostenko, 2017)	27
Figure 2.2 : Coopération horizontale (Millot, 1999 dans Kostenko, 2017)	28
Figure 2.3 : Coopération humain-machine selon Hoc (2001).....	29
Figure 2.4 : Synthèse du KH et KHC (traduit de Millot & Pacaux-Lemoine, 2013)	31
Figure 2.5 : Liens entre les composantes de confiance (Hoff & Bashir, 2015) et les facteurs d'influence	39
Figure 2.6 : Surconfiance, confiance appropriée et sous-confiance selon les capacités perçues et réelles d'un CPS (selon Théodorou, Wortham & Bryson, 2017).....	41
Figure 2.7 : Les trois niveaux de la conscience de la situation selon Endsley (1995a)	42
Figure 2.8 : Modèle de prise de décision incluant la conscience de la situation et la perception du risque (Liu et al., 2020).....	43
Figure 2.9 : Modèle de la confiance humain-machine sur les comportements de compliance/reliance selon Lee et See (2004) et Mayer et al. (1995) (traduit et adapté de Chancey et al., 2017).....	46
Figure 2.10 : Adaptation du modèle de Chancey et al. (2017) selon les composantes de la confiance (Hoff & Bashir, 2015)	49
Figure 3.1 : Human-robot transparency model (Lyons, 2013)	55
Figure 3.2 : Situation Awareness based-agent Transparency model (Chen et al., 2014)	58
Figure 3.3 : Dynamic Situation Awareness based-agent Transparency model (Chen et al., 2018)	59
Figure 3.4 : Articulation des modèles de Lyons (2013) et Chen et al. (2018) (adapté)	60
Figure 3.5 : Événements et réactions dans une interface adaptative (adapté de Calvary & Coutaz, 2002)	66
Figure 3.6 : Dimensions de la transparence sur les composantes de la confiance dans une prise de décision assistée par un CPS	69
Figure 5.1 : Synoptique des verrous scientifiques	87

Figure 5.2 : Composantes de la confiance de l'opérateur dans le cadre de la prise de décision assistée par un CPS	88
Figure 5.3 : Compréhension des modalités de transparence dans le cadre de la prise de décision assistée par un CPS	90
Figure 5.4 : Intégration d'un mécanisme adaptatif sur la transparence dans le cadre de la prise de décision assistée par un CPS	91
Figure 6.1 : Contexte expérimental et rôle des agents humains et du cobot	98
Figure 6.2 : Interface de complétion de tâche et de proposition.....	102
Figure 6.3 : Interface des données du cobot selon les niveaux de transparence. Le point d'exclamation indique un risque de rupture de stock communiqué par le cobot	103
Figure 6.4 : Positionnement des questions de recherche dans le cadre de la prise de décision assistée par un cobot.....	105
Figure 6.5 : Probabilité d'acceptation de la proposition du cobot en fonction des modalités de transparence	108
Figure 6.6 : Effet de la transparence sur les modalités de réponses aux questionnaires de confiance.....	110
Figure 6.7 : Validation et synthèse des hypothèses initiales.....	118
Figure 7.1 : Conditions expérimentales de l'expérimentation	123
Figure 7.2 : Représentation des données lorsque la proposition est d'avancer un changement de filtre.....	126
Figure 7.3 : Interface de dialogue lorsque la proposition est d'avancer un changement de filtre.....	126
Figure 7.4 : Positionnement des hypothèses, des observables et des variables dans le cadre de la prise de décision assistée par une IA	129
Figure 7.5 : Taux d'acceptation en fonction des modalités de transparence	131
Figure 7.6 : Taux d'acceptation en fonction de la criticité de la situation	132
Figure 7.7 : Taux d'acceptation en fonction de la transparence et de la criticité de la situation	132
Figure 7.8 : Régression logistique sur l'effet de la transparence et de la situation sur l'acceptation (référence transparence « F » et criticité « Élevée »).....	133
Figure 7.9 : Régression logistique ordinaire sur l'effet de transparence et de la situation sur la confiance dans le signal (référence transparence « F » et criticité « Élevée »).....	134
Figure 7.10 : Régression logistique ordinaire sur l'effet de la transparence et de la situation sur la perception du risque (référence transparence « F » et criticité « Élevée »).....	136
Figure 7.11 : Validation et synthèse des hypothèses initiales.....	142
Figure 8.1 : Protocole expérimental	145
Figure 8.2 : Transparence en fonction du taux d'acceptation.....	146
Figure 8.3 : Installation expérimentale.....	148
Figure 8.4 : Interface de dialogue Seanatic	149
Figure 8.5 : Données utilisées par l'IA Seanatic.....	150
Figure 8.6 : Positionnement des hypothèses, des observables et des variables dans le cadre de la prise de décision assistée par une IA	151
Figure 8.7 : Codage des deltas et application à un chronogramme d'un participant.....	152
Figure 8.8 : Chronogramme moyen des participants du groupe « Constant » (haut) et « Versatile » (bas)	153
Figure 8.9 : Moyennes des groupes « versatile » et « constant » sur les échelles IMOTRIS 1 et 2	154

Figure 8.10 : Effet du delta transparence et du delta feedback sur le delta d'acceptation (Régression logistique multinomiale)	155
Figure 8.11 : Validation et synthèse de l'hypothèse initiale.....	161
Figure 9.1 : Résumé des questions de recherches validées	162
Figure 9.2 : Composantes de la confiance et déterminants de la prise de décision de l'opérateur	164

Acronymes

CA = Confiance Apprise

CAD = Confiance Apprise Dynamique

CAI = Confiance Apprise Initiale

CD = Confiance Dispositionnelle

CHM = Coopération Humain-Machine

CPS = Cyber-Physical System / Systèmes Cyber-Physique

CS = Confiance Situationnelle

CWA = Cognitive Work Analysis

GMAO = Gestion de Maintenance Assistée par Ordinateur

IA = Intelligence Artificielle

IOT = Internet des objets (Internet Of Thing)

rOFh = Robot Of Human

rTOh = Robot To Human

SA = Situation Awareness (conscience de la situation)

SAT = Situation Awareness based-agent Transparency

XAI = eXplainable Artificial Intelligence

Introduction générale

Le plan « France 2030 » est un plan de financement de 54 milliards d'euros sur 5 ans visant à l'amélioration de l'industrie française. Cette amélioration passe par le développement et l'intégration des technologies numériques pour créer des industries dites 4.0 (France 2030, 2021). En parallèle à cela, fin 2022, l'Union Européenne a lancé le programme de financement « Destination 6 : A human-centred and ethical development of digital and industrial technologies » (Horizon Europe Calls, 2022). Ce programme vise à financer les projets de recherche mettant l'humain au centre de la conception pour les industries du futur. Ces deux programmes de financement nous permettent de constater que la coopération humain-machine (CHM) joue un rôle central dans cette révolution industrielle. Or, la CHM représente des risques (physiques et/ou cognitifs) du point de vue de la santé et de la sécurité des opérateurs (Lindholm, Reiman & Väyrynen, 2020). Le challenge, pour aider à la transition vers l'industrie 4.0, est de réduire ces risques pour créer une industrie où les systèmes cyber-physiques (CPS) sont au service de l'opérateur, et non l'inverse.

Dans le cadre de cette transition, deux constats peuvent être faits. Premièrement, la part de marché liée à la vente de cobots (robots collaboratifs) est en hausse dans le milieu industriel (Couto (2021, 1er février)). Secondement, les intelligences artificielles (IA), fortes des avancées technologiques et scientifiques, deviennent des outils de calcul de plus en plus puissants et utilisés (Lai, Chen, Liao, Smith-Renner & Tan, 2021). Ces deux technologies peuvent néanmoins avoir une vision parcellaire de l'environnement et avoir accès à données limitées pour leurs algorithmes. Ainsi, si ces outils peuvent être considérés comme fiables, ils ne doivent pas être considérés comme omniscients et une marge d'erreur résiduelle existe. Or, tout comme de nombreux autres domaines (médical, militaire, juridique), l'industrie fait appel à ces outils pour aider les opérateurs dans leurs prises de décision (Schemmer, Hemmer, Nitsche, Kühl & Vössing, 2022). Le risque provient alors d'une utilisation inappropriée des propositions de ces outils, soit lié au fait que l'opérateur fasse trop confiance, ou à l'inverse, trop peu confiance. L'intégration de ces outils numériques, et leur interconnexion dans l'environnement de travail, fait d'eux des systèmes cyber-physique (CPS). Cette question de la confiance et de l'utilisation des propositions des CPS est au centre des problématiques scientifiques pour la CHM.

L'enjeu est donc ici de comprendre comment aider à la calibration de la confiance de l'opérateur dans le CPS et améliorer la prise de décision assistée par un CPS.

Comme lorsque deux opérateurs humains travaillent ensemble, la communication est un facteur primordial pour s'assurer d'une bonne coopération et de bonnes performances. Le dialogue entre l'opérateur (humain) et le CPS va aider à une compréhension mutuelle et permettre d'adopter des comportements adaptés à la situation (Hoc, 2001). Le dialogue entre l'opérateur et le CPS passe par de multiples canaux, tels que les canaux sonores ou visuels (Moulières-Seban, 2017). Ces canaux servent à transmettre les informations à l'opérateur. En fonction des capacités dont ils sont dotés, les CPS peuvent communiquer sur différentes informations. Certains CPS vont faire une proposition en explicitant ce qu'ils ont dans leur environnement, leurs algorithmes ou toutes informations pertinentes. D'autres, à l'inverse, auront des capacités de communication plus limitées, communiquant uniquement leur proposition. La communication de ces informations influence la confiance de l'opérateur dans le CPS et donc sa prise de décision. Le risque d'une compréhension erronée amenant à une mauvaise calibration de la confiance est une problématique cruciale à dépasser pour l'intégration des CPS dans le milieu industriel. D'un point de vue industriel, les futurs CPS devraient donc posséder des capacités de communication permettant à l'opérateur de calibrer sa confiance et de l'aider dans sa prise de décision.

Le premier objectif de cette thèse est de comprendre comment fonctionne le processus de confiance et son rôle dans la prise de décision d'un opérateur assistée par un CPS intelligent, aussi bien physique (cobot) que virtuel (IA). Le second objectif est de comprendre quels sont les moyens à disposition pour agir sur ces processus et les potentiels mécanismes d'adaptation permettant de favoriser la coopération humain-machine.

Nous répondrons à ces objectifs au travers deux parties. La première partie est une partie théorique et contient 5 chapitres. Cette partie nous permettra de définir les enjeux industriels actuels, la coopération humain-CPS et la transparence. La seconde partie est une partie expérimentale et contient 4 chapitres, dont 3 expérimentaux. Ces chapitres nous permettront d'apporter des réponses à nos questions de recherche, et plus globalement, des compléments de compréhension sur le rôle de la transparence dans la CHM.



Partie théorique

Prise de décision assistée par un CPS
dans l'industrie 4.0 : confiance de
l'opérateur et transparence du CPS

Chapitre 1 : Les nouveaux défis posés par l'industrie 4.0 pour les interactions entre l'humain et la machine

Ce premier chapitre a deux objectifs. Le premier est de définir le cadre dans lequel s'inscrit ce travail de thèse. Le second est de définir les besoins industriels auxquels vont répondre les travaux présentés dans ce manuscrit. Ce chapitre permet de spécifier notre classe de problème et de présenter les nouvelles interactions entre l'humain et les systèmes cyber-physiques (CPS) qui y sont associées. De ces nouvelles interactions nous définissons les différentes problématiques qui émergent.

Nous commençons par définir les évolutions qui ont conduit à l'industrie 4.0. Pour répondre aux nouveaux besoins de consommation et environnementaux, les industries parient sur l'intégration de nouvelles technologies dans les usines telles que les CPS et l'utilisation massive de données (« big data »). Nous traitons donc de l'intégration de ces CPS qui viennent créer de nouvelles interactions et modifier la façon dont les opérateurs travaillent avec le CPS (Romero et al., 2016a).

L'utilisation des CPS peut intervenir à différents niveaux et aider dans différentes fonctions de l'industrie (Rauffet, Guerin, Chauvin & Martin, 2018). Nous définissons quels sont ces niveaux et ces fonctions, et comment les CPS s'y intègrent. Ces éléments nous permettent de préciser la classe de problème qui est investiguée dans cette thèse : la prise de décision assistée par un CPS.

Au travers de l'explicitation des nouvelles capacités et compétences des CPS et des opérateurs, nous décrivons plus en détails les interactions entre l'opérateur et le CPS dans la prise de décision assistée par un CPS. Les avancées technologiques ont permis de faire évoluer les capacités des CPS, les faisant passer d'exécutants automatés à assistants. Cette évolution amène à de nouvelles problématiques à résoudre pour améliorer l'interaction humain-machine.

Nous concluons en introduisant les deux projets de recherche qui servent de support à nos expérimentations. Nous les définissons au regard des typologies d'interaction Humain-

Machine de l'industrie 4.0 et selon les niveaux et les fonctions associés à ces cas d'application. Cela nous permet de montrer les similitudes et différences entre ces deux projets.

1.1. L'Industrie 4.0 : définitions et enjeux

Selon le CNRTL (Centre national de ressources textuelles et lexicales), l'industrie est « l'ensemble des activités économiques (caractérisées par la mécanisation et l'automatisation des moyens de travail, la centralisation des moyens de production et la concentration de la propriété des moyens de production), ayant pour objet l'exploitation des sources d'énergie et des richesses minérales du sol ainsi que la production de produits fabriqués à partir de matières premières ou de matières ayant déjà subi une ou plusieurs transformations. ». Le début de l'industrialisation, sous-entendu la mise en place de l'industrie à grande échelle, est généralement placé à la fin du XVII^{ème} siècle en Angleterre et au début du XVIII^{ème} siècle dans le nord de l'Europe (Hudson, 2014). Ce début de l'industrialisation correspond à l'essor des technologies fonctionnant grâce à l'énergie produite par l'utilisation de la vapeur et de la combustion du charbon. Depuis, il est communément admis de découper l'industrie en quatre révolutions jusqu'à aujourd'hui (Frise chronologique des 4 générations d'industrie (inspirée de Christoph Roser sur AllAboutLean.com) (Figure 1.1).

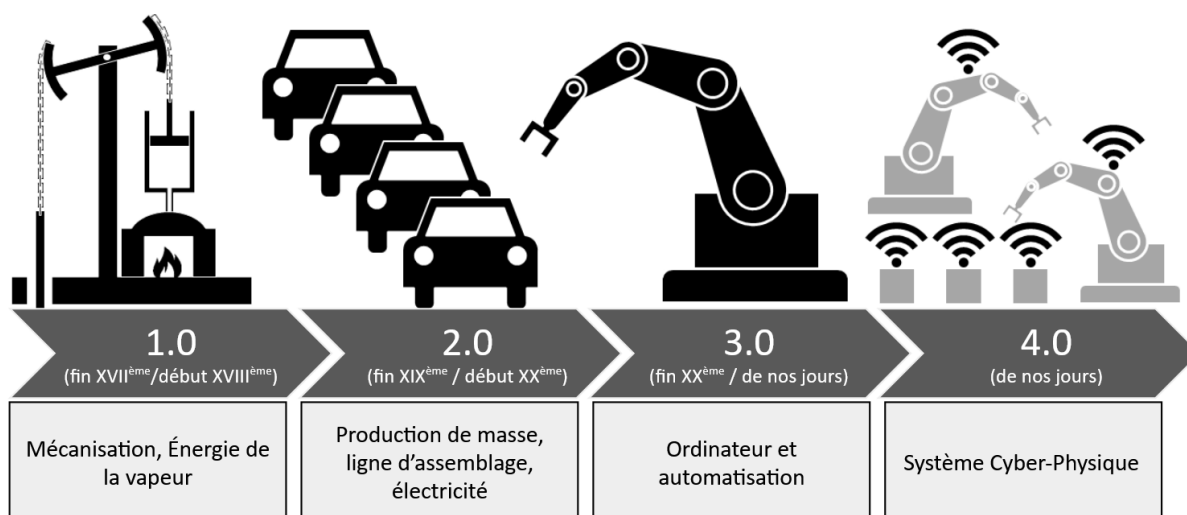


Figure 1.1: Frise chronologique des 4 générations d'industrie (inspirée de Christoph Roser sur AllAboutLean.com)

La seconde révolution de l'industrie correspond à la généralisation de la production de masse grâce aux lignes d'assemblage et à l'utilisation de l'électricité comme source d'énergie à la fin du XIX^{ème} siècle. La troisième révolution débute grâce à la démocratisation des

microprocesseurs dans les années 70. Cette démocratisation a permis la mise en place d'automates et l'utilisation des ordinateurs au centre des industries. Cette évolution de l'industrie est désormais bien implantée et une nouvelle révolution est en cours : l'industrie 4.0.

L'objectif de cette révolution est de créer des produits personnalisés et de meilleure qualité, d'utiliser des énergies renouvelables pour la transition écologique et d'améliorer les conditions de travail des employés (Rauffet et al., 2018). Cette quatrième révolution a pour spécificité la rencontre entre l'internet des objets et les systèmes cyber-physique (CPS) au sein d'usines intelligentes (Kagermann, Wahlster & Helbig, 2013). Ces nouvelles technologies vont permettre de modifier la façon dont sont gérées les usines, leur capacité de production et leur flexibilité (Kagermann et al., 2013). L'histoire de l'internet des objets remonte aux années 80 et correspond à la volonté d'interconnecter des objets via l'utilisation d'un système de communication. Cette interconnexion permet de récolter et de transmettre des informations à différents objets avec comme objectif l'adaptabilité et la flexibilité. La notion de système cyber-physique est quant à elle plus récente car apparue au milieu des années 2000 (Cardin, 2019). Il en existe de nombreuses définitions (Lee, 2006 ; Rajkumar, Lee, Sha & Stankovic, 2010 ; Baheti & Gill, 2011), mais celle de Monostori (2014) est couramment utilisée. Cardin (2019) propose de compléter cette définition en la rendant plus englobante : « *Cyber-Physical Production Systems are systems of systems of autonomous and cooperative elements connecting with each other in situation dependent ways, on and across all levels of production, from processes through machines up to production and logistics networks, enhancing decision-making processes in real-time, response to unforeseen conditions and evolution along time* » (Page 4). Cette définition met en évidence que l'utilisation de CPS permet aux opérateurs d'améliorer leurs prises de décision et de s'adapter plus facilement à de nouvelles situations, et ce à tous les niveaux de l'industrie. Les CPS dans l'industrie 4.0 peuvent prendre plusieurs formes d'agents autonomes ou automatisés (jumeaux numériques, robots, cobots, intelligences artificielles, etc.) en fonction des besoins de l'industriel et des opérateurs. Les CPS peuvent également se baser sur l'usage de différentes technologies (big data, machine learning, projection d'état, etc.) et différents moyens de communication (interfaces, réalité virtuelle, sons, etc.). L'industrie 4.0 s'articule autour de 9 piliers technologiques (Rüßmann et al., 2015) qui vont permettre de répondre à trois grands enjeux (Rauffet et al., 2018).

Le premier enjeu concerne les données et comprend la facilitation de l'accès aux informations de l'usine, l'accession à des données fiables et l'émergence de nouvelles connaissances (Hozdić, 2015). Cet enjeu repose sur l'utilisation d'IA apprenantes (machine learning) qui permettent d'analyser de nombreuses données, de réaliser des simulations ou de modéliser l'usine (Longo, Nicolletti & Padovano, 2017). Ces IA se basent sur l'utilisation des technologies de **l'internet des objets (IOT)**, de **l'intégration horizontale et verticale des systèmes informatisés** (i.e. coopération entre les machines et intégration des informations entre les niveaux), des **big datas** (i.e. accès à de nombreuses données), du **cloud computing** (i.e. stockage des données et réalisation des calculs informatiques sur des serveurs à distance) et de la **cybersécurité** (i.e. protections cyber-physiques des systèmes informatisés).

Le second enjeu concerne la gestion de la production et comprend la vérification de la production et la supervision de l'environnement. Pour répondre à cet enjeu, il est possible d'utiliser la **simulation** (i.e. virtualisation d'un objet réel) et la **réalité virtuelle ou augmentée** (i.e. utilisation d'une représentation visuelle informatisée). La simulation permet par exemple, via l'utilisation de jumeaux numériques, de prévoir les erreurs et de réduire les coûts de conception. La simulation peut aider les opérateurs grâce aux technologies de réalité virtuelle ou augmentée.

Le troisième enjeu concerne le changement des modes de production et a pour objectif la mise en place de modes de production flexibles. Un mode de production flexible permet de s'adapter aux demandes spécifiques et l'industrie passe d'une production de masse à une production de customisation (Hozdić, 2015 ; Longo et al., 2017). Les technologies d'**impression 3D** et **self-adjusting robotics** (i.e. robots capables de s'adapter à de nouvelles situations) permettront de répondre à cet enjeu de flexibilité.

Ces trois enjeux sont marqués par les changements provoqués par l'interaction entre les CPS et les humains. L'interaction humain-machine (IHM) est définie par Johanssen (2009) comme « *the interaction and communication between human users and a machine, a dynamic technical system, via a human-machine interface* » (page 132). Les problématiques liées aux nouvelles interactions, créées par l'intégration de systèmes automatiques, informatiques ou autonomes, ne sont cependant pas propres à l'industrie 4.0. Dans un contexte d'industrie 3.0, Parasuraman et Riley (1997) pointaient déjà les risques d'une mauvaise utilisation des

automates sur l'opérateur (perte de compétence, hausse de la charge de travail, mauvaise utilisation de l'automate, etc.). Fantini, Pinzone et Taisch (2018) insistent sur le fait que, dans l'industrie 4.0, il est nécessaire de comprendre ces futures interactions pour pouvoir les améliorer. Plusieurs auteurs considèrent que l'opérateur sera dans une forme de symbiose avec les technologies nouvellement implantées (Romero et al., 2016b ; Zarte, Pechmann & Nunes, 2020 ; Lu, Adrados, Chand & Wang, 2021). Dans cette optique, Romero et al. (2016a ; 2016b) proposent une typologie permettant de distinguer les différents types d'opérateurs en fonction des CPS utilisés et des interactions engendrées (Figure 1.2).



Figure 1.2 : Typologie des opérateurs 4.0 (issue de Romero et al., 2016b)

Cette typologie distingue trois grands types d'interactions. Lorsque les CPS aident à la prise de décision, ou qu'ils apportent de nouvelles données et informations à l'humain, l'interaction sera **cognitive** (exemple des « analytical operator » qui ont une compréhension de l'environnement amélioré grâce aux IOT et à l'IA). Lorsque les CPS améliorent les capacités physiques, ou soulagent le travail physique, l'interaction sera **physique** (exemple des « super-strength operator » qui voient leur travail simplifié grâce à des exosquelettes). Lorsque les CPS apportent à la fois des informations supplémentaires et que cela a un impact physique, l'interaction est à la fois **physique et cognitive** (exemple des « Healthy Operator » qui améliore la compréhension de leur fonctionnement physiologique et qui peuvent adapter leur charge de travail). Il est intéressant de noter que les interactions n'ont pas lieu

exclusivement lors de la production. En fonction du CPS, les interactions peuvent intervenir à tous les niveaux et sur toutes les fonctions de l'industrie.

1.2. Fonctions de l'industrie 4.0 et intégration des CPS

Rauffet et al. (2018) ont montré que les CPS peuvent être utilisés dans toutes les fonctions qu'un opérateur est amené à réaliser dans une industrie 4.0 (Figure 1.3). Les fonctions se classent selon trois niveaux dans l'industrie. Le niveau le plus haut est stratégique et correspond au management d'entreprise sur le long-terme. Le niveau intermédiaire est le niveau tactique. Il correspond à une gestion à moyen-terme et court-terme de la production et de la maintenance. Ce niveau survient à la fois en amont et pendant la production. Les fonctions de ce niveau serviront de soutien à la production, aussi bien dans l'optimisation de son fonctionnement (i.e. en amont) que pour être réactif à un problème (i.e. pendant la production). Le niveau le plus bas est opérationnel et est mis en œuvre pendant la production. Les fonctions opérationnelles correspondent notamment à l'activité de production et à l'activité de maintenance. Pour ces trois niveaux, et leurs différentes fonctions, des CPS peuvent aider l'opérateur à remplir ses tâches.

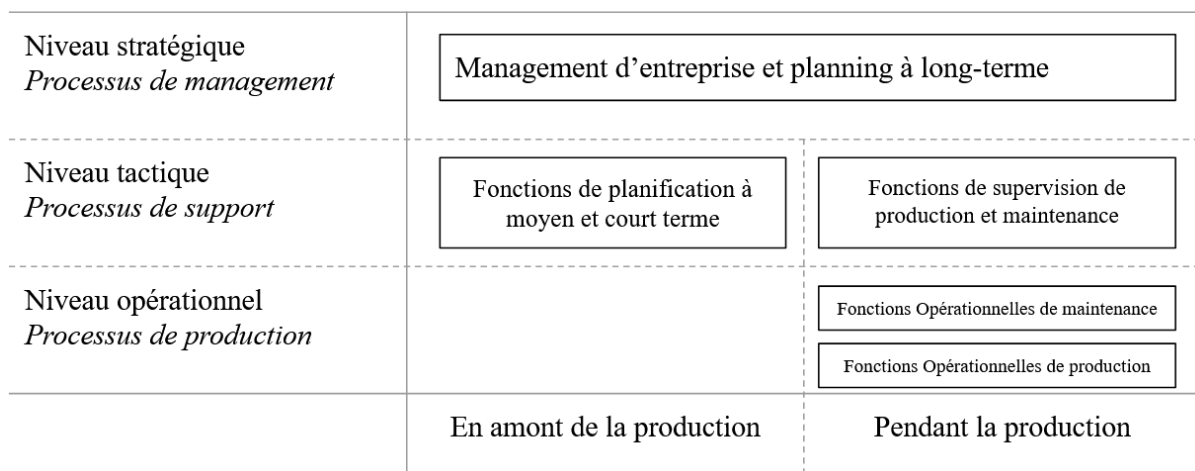


Figure 1.3 : Cartographie des fonctions de production (inspirée de Rauffet, et al., 2018)

Au niveau stratégique, l'opérateur peut être aidé par des CPS de types DSS (i.e. système d'aide à la décision) ou PGI (i.e. Progiciel de Gestion Intégré). Ces outils permettent la récolte et l'agrégation de multiples données (exemple : données de productions, de ressources humaines, etc.) en facilitant leur lisibilité. **Les CPS soutiennent l'opérateur en l'aidant dans**

sa prise de décision sur des éléments « long-terme » de l'entreprise (politique d'entreprise, gestion comptable, ressources humaines, etc.).

1.2.1. Fonctions et CPS de niveau tactique

Le niveau tactique de l'industrie se structure autour de deux fonctions dans lesquelles les CPS pourront apporter un soutien à l'opérateur.

1) *Fonction de planification.* La fonction de planification est définie comme une « méthode consistant à choisir des objectifs et à proposer des moyens pour y parvenir » (CNRTL, s.d.). Cette fonction est utilisée selon deux objets industriels d'intérêt :

- La **planification pour la production** renvoie à l'anticipation des événements et permet d'organiser la production. Cette planification permet de supporter la fonction de production par une gestion appropriée des agents ou des stocks en fonction des commandes à venir. Les opérateurs utilisent notamment des CPS de type FAO (i.e. Fabrication Assistée par Ordinateur) ou des HRIS (i.e. systèmes d'information de gestion des ressources humaines). Ces CPS communiquent des contraintes aux opérateurs : contraintes entre les automates pour la FAO et contraintes de disponibilité pour les HRIS. **Ces CPS soutiennent l'opérateur dans sa prise de décision sur la gestion tactique de la ligne en amont de la production.**
- La **planification pour la maintenance.** Il est possible de catégoriser la maintenance en trois types : 1) la maintenance **réactive** visant à réparer les équipements après un incident, 2) la maintenance **systématique** visant à réparer les équipements selon un calendrier (horaire ou à date), 3) la maintenance **préventive** visant à réparer les équipements selon leur état d'usure réel. La maintenance préventive permet de planifier les opérations de maintenance au moment le plus opportun en prenant en compte des éléments externes (visites obligatoires, disponibilités des équipes, etc.) et l'état de l'équipement (réel ou projeté) (NF EN 13306:V2018). Cette planification sert de support à la production en limitant les risques de casses et les temps d'arrêt. La fonction planification est assistée par des CPS de type GMAO (i.e. « Gestion de Maintenance Assistée par Ordinateur ») pour de la maintenance systématique et des IOT et des IA pour la maintenance préventive. La GMAO fournit des éléments de

contraintes calendaires aux opérateurs. Les IOT et les IA fournissent de multiples informations, allant des données remontées par les capteurs à de la projection d'état. Les IA les plus développées proposent également de nouvelles dates de maintenance plus adaptées. **Ces CPS soutiennent l'opérateur dans sa prise de décision de niveau tactique sur la gestion du calendrier de maintenance.**

2) *Fonction de supervision.* La fonction de supervision est une technique industrielle qui consiste à suivre de façon permanente l'état des agents et les étapes de production. Elle est définie comme un « *contrôle exercé par la personne qui supervise* » (CNRTL, s.d.). C'est par l'acquisition des données (état de l'équipement, stocks, etc.) que cette supervision est effectuée. Cette supervision est aidée par l'utilisation de logiciels. L'activité de supervision permet d'améliorer la connaissance de son environnement de travail. Cette fonction est considérée comme se situant à un niveau tactique car elle va, par exemple, permettre la résolution de conflits (Milot & Mandiau, 1995). La fonction supervision de production est supportée par des CPS de type SCADA (i.e. Système de contrôle et d'acquisition de données) ou des logiciels de pilotage de la production. Ces CPS permettent à l'opérateur d'avoir accès à des données complémentaires sur l'état et le fonctionnement des équipements. **Ces CPS soutiennent l'opérateur dans ses prises de décision de niveau tactique pour la gestion de la ligne et de la maintenance pendant la production.**

1.2.2. Fonctions et CPS de niveau opérationnel

Le niveau opérationnel de l'industrie se structure autour de deux fonctions dans lesquelles les CPS apportent un soutien à l'opérateur.

1) *Fonction de production.* La fonction de production représente toutes les activités qu'est amené à réaliser un opérateur afin de transformer une matière première. Cette fonction peut être définie comme « *l'ensemble des activités qui permettent à l'homme de créer et de s'approprier les produits de l'agriculture, de l'industrie, ou d'assurer les services permettant de satisfaire les besoins de la société* » (CNRTL, s.d.). Afin de pouvoir utiliser les matières premières, l'opérateur est également amené à déployer des savoir-faire. Pour l'aider dans sa fonction de production, l'opérateur est accompagné de CPS de type cobot. Ces CPS permettent à l'opérateur d'augmenter ses capacités de production et d'avoir accès à de

nouvelles informations. **Ces CPS apportent trois types d'aide : augmentation des capacités de production de l'opérateur, aide à la surveillance locale de la ligne, soutien de l'opérateur dans ses prises de décision de niveau opérationnel. Ces trois aides améliorent la gestion de la ligne pendant la production.**

2) *Fonction de maintenance.* La norme NF EN 13306:V2018 définit la maintenance comme : « *l'ensemble de toutes les actions techniques, administratives et de management durant le cycle de vie d'un bien, destinées à le maintenir ou à le rétablir dans un état dans lequel il peut accomplir la fonction requise.* ». La fonction de maintenance est la mise en œuvre de la planification de maintenance par l'utilisation des compétences techniques des opérateurs. Les opérateurs sont amenés à agir directement sur les équipements pour les entretenir, les réparer ou les améliorer. Pour la réalisation de la maintenance, les CPS de type VR (casque ou lunette de réalité virtuelle ou augmentée) sont implémentés. Ces CPS permettent à l'opérateur d'avoir des informations complémentaires sur les procédures à réaliser sur l'équipement pendant qu'il réalise son opération. **Ces CPS soutiennent l'opérateur dans sa compréhension du système et dans le suivi des procédures pendant la production.**

Dans l'industrie 4.0 les CPS soutiennent le travail des opérateurs à différents niveaux et de diverses façons. Les nouvelles interactions que les opérateurs ont avec ces CPS amènent à une transformation du travail. Les opérateurs 4.0 font preuve de plus de flexibilité et d'une capacité à s'adapter à un environnement complexe et dynamique (Hirsch-Kreinsen, 2014 ; Longo et al., 2017). Le rôle des opérateurs évolue et les opérateurs passent d'opérateurs peu qualifiés à des opérateurs qualifiés et avec des responsabilités (Hirsch-Kreinsen, 2014 ; Di Nardo et al., 2020). Pour Inkulu, Bahubalendruni, Dara et Sankaranarayana (2022), les challenges liés à ces interactions sont multiples. On y retrouve le besoin de reconnaître, pour l'opérateur comme pour le CPS, les intentions de l'autre. L'opérateur et le CPS doivent être en capacité de comprendre ce que fait l'autre pour s'adapter. L'identification des tâches en cours doit être faite de façon continue pour amener à une optimisation de la production. Le risque est cependant lié aux erreurs que les CPS peuvent commettre et dégrader la compréhension des opérateurs, leurs compétences, et au final la performance dans la tâche

(Bainbridge, 1983 ; Baxter, Rooksby, Wang & Khajeh, 2012 ; Jiang, Karran, Coursaris, Léger & Berringer, 2022).

Dans le cas de l'usage d'une IA pour la maintenance prévisionnelle, le challenge est avant tout cognitif. Les IAs ont la possibilité d'acquérir des données sur ce qu'il se passe dans leur environnement et de réaliser des projections grâce à l'utilisation d'algorithmes (Seguin, Rioual, Diguët & Gogniat, 2022). Un opérateur devra alors comprendre les données générées par l'IA tout comme les données concernant les autres contraintes (calendrier de production par exemple) (Simon, Guerin, Rauffet & Lassalle, 2021a ; Wellsandt et al., 2021). Ce travail de décision nécessite à l'opérateur de faire appel à de nouvelles compétences comme l'analyse de données. Pour l'aider dans cette tâche, l'interaction humain-machine doit être simplifiée.

Dans le cas de l'interaction avec un cobot, les challenges sont à la fois physiques et cognitifs. Physique d'une part, car il est nécessaire à l'opérateur de comprendre les mouvements que fait le cobot pour s'assurer d'être en sécurité (Kadir, Broberg & Souza da Conceição, 2018). Cognitif d'autre part, car l'opérateur a besoin de comprendre la répartition des tâches au sein de cette équipe humain-machine (Kadir et al., 2018 ; Weiss, Wortmeier & Kubicek., 2021) et de comprendre le fonctionnement du cobot (Weiss et al., 2021).

Les CPS dépassent le cadre de l'outil et peuvent être assimilés à des coéquipiers (Weiss, et al., 2021). Ces nouvelles interactions se rapprochent de ce que Hollnagel définit comme les « Joint Cognitive Systems » (2005). Les CPS et les humains sont en interaction pour atteindre des buts communs et ont une performance de groupe. Le CPS fait partie intégrante de l'équipe comme un autre opérateur et doit échanger avec l'humain pour améliorer la performance (Weiss, et al., 2021). Pour Di Nardo et al. (2020) ces changements sont notamment apportés par les IA, utilisant et analysant les big datas, et les robots de production (type cobot). Pour ces auteurs, la distinction principale entre l'opérateur traditionnel et l'opérateur 4.0 se situe dans ses compétences et ses connaissances. L'opérateur 4.0 est un opérateur qui est expérimenté (Hirsch-Kreinsen, 2014), dans une position de supervision et de gestion des CPS (Di Nardo et al., 2020) et qui doit résoudre des problèmes (Gorecky, Schmitt, Loskyll & Zühlke, 2014). En plus de ses tâches initiales (de production par exemple), l'opérateur est amené à superviser les CPS pour résoudre leurs dysfonctionnements. Pour aider les opérateurs, les CPS donnent accès à des données ou peuvent faire des propositions

(Peres, Jia, Lee, Sun, Colombo & Barata, 2020 ; Rosin, Forget, Lamouri & Pellerin, 2022). Les propositions des CPS peuvent aider à la prise de décision pour des fonctions de gestion de la planification d'interventions ou pour la résolution de problèmes survenant sur une ligne de production par exemple.

1.3. Prise de décision assistée par un CPS dans l'industrie 4.0

Les opérateurs ont de plus en plus recours aux CPS pour résoudre des conflits dans l'industrie 4.0 (Rosin et al., 2022) et les nouvelles demandes de l'industrie 4.0 complexifient les activités quotidiennes de l'opérateur (Longo et al. 2017). Pour aider les opérateurs, les CPS possèdent des capacités et des compétences complémentaires (Kagerman & Wahlster, 2022). Dans la prise de décision assistée par un CPS ce sont notamment les capacités d'acquisition, d'analyse et d'affichage de nombreuses données qui viennent aider l'opérateur (Jarrahi, 2018). Ces capacités renvoient au risque plus général identifié par Woods, Patterson et Roth (2002) sur l'afflux massif de données. Ce risque est très présent dans l'industrie 4.0 avec l'accès aux données en temps réel (Habib, Pacaux-Lemoine, Berdal & Trentesaux, 2021) et est lié au fait que ces données doivent être interprétées par l'opérateur (Madonna, Monica, Anastasi & Di Nardo 2019). Ces données amènent l'opérateur à avoir de multiples éléments à investiguer en parallèle et ceci dans un temps réduit. La conséquence pour les opérateurs est une difficulté à identifier les informations les plus importantes (Woods et al., 2002). Madonna et al. (2019) montrent qu'une opération de maintenance utilisant la réalité augmentée (RA) nécessite plus d'activités cognitives avec la RA que sans la RA. Ils montrent qu'une opération de maintenance avec RA conduit à plus d'activités d'interprétation et d'observation pour l'opérateur. Pour comprendre les changements de l'opérateur 4.0 au niveau cognitif, Madonna et al. (ibid) proposent une conceptualisation de son travail cognitif (Figure 1.4). Contrairement à un opérateur sans CPS, l'opérateur qui travaille avec un CPS peut percevoir la situation à travers deux visions. La première vision est sa propre perception de l'environnement. La seconde vision est celle proposée par le CPS. Cette vision est basée sur les données acquises et analysées par le CPS, et transmises à l'opérateur par l'interface humain-machine.

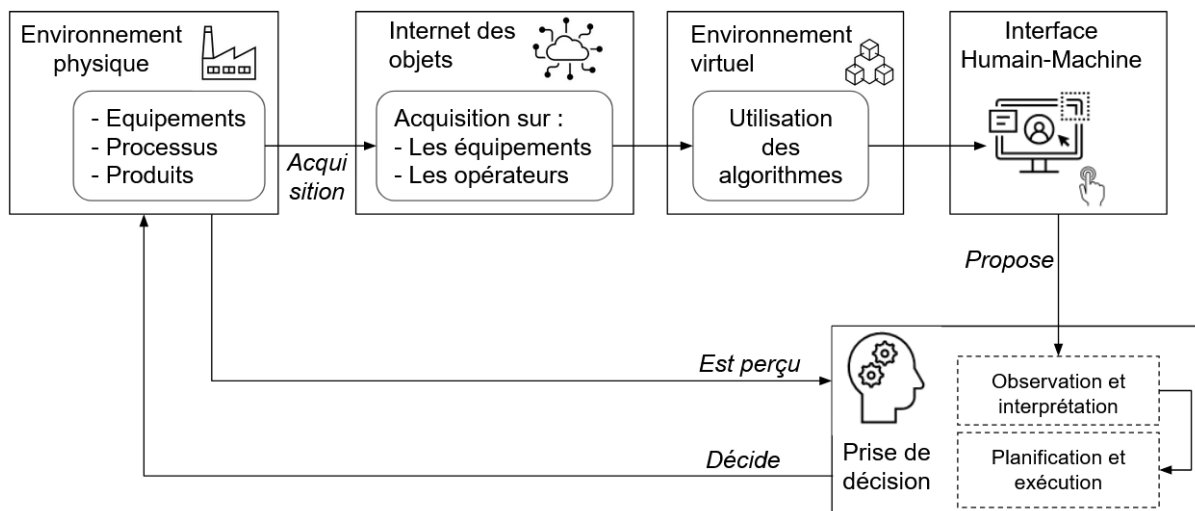


Figure 1.4 : Proposition de cadre cognitif conceptuel de l'opérateur dans une usine intelligente (selon Madonna et al. 2019)

Pour la prise de décision assistée par un CPS, Jarrahi (2018) met en avant que les CPS de type intelligence artificielle améliore la prise de décision de l'opérateur en apportant une réflexion analytique (basée sur les données et les algorithmes). La réflexion analytique du CPS complète la vision holistique de l'opérateur (basée sur de l'instinct). Jarrahi distingue trois caractéristique de la prise de décision avec l'incertitude, la complexité et l'équivocité. Pour chacun de ces caractéristiques, l'opérateur et le CPS se complètent selon leurs différentes capacités (Figure 1.5).

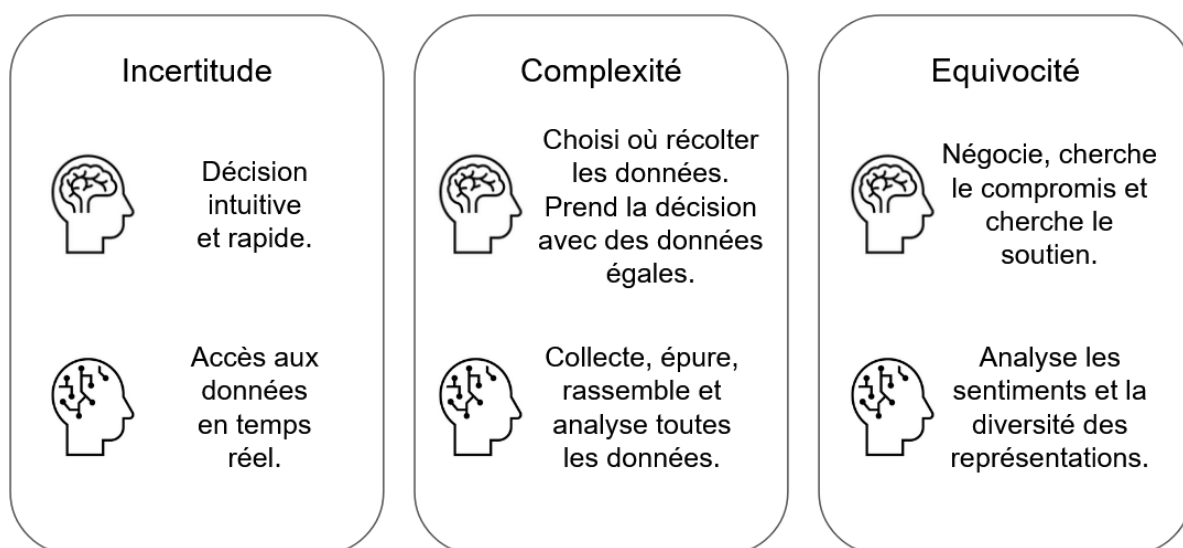


Figure 1.5 : Complémentarités des capacités des opérateurs (en haut) et des CPS (en bas) selon les catégories de prise de décision (selon Jarrahi, 2018)

Par rapport à l'incertitude, le CPS permet l'accès à des données en temps réel pour aider à la prise de décision. Par exemple, dans le cadre de la surveillance aérienne, le contrôleur aérien est amené à dérouter des avions. Ici, le CPS vient aider la prise de décision de l'opérateur en lui indiquant des données telles que la vitesse et la trajectoire de l'avion concerné. Par rapport à la complexité, le CPS collecte et analyse les données. Cette caractéristique de la prise de décision intervient, par exemple, pour décaler la maintenance d'un équipement. Pour aider à la prise de décision, le CPS fournit des projections d'état de l'équipement sur la base des données récoltées. Ces projections sont issues d'algorithmes ayant une part d'incertitude inhérente à leur fonctionnement. Les CPS fournissent donc des **probabilités** d'états futurs ou de survenue d'événements. Par rapport à l'équivocité de la prise de décision, le CPS aide à analyser les sentiments et à la représentation des différents points de vue des parties prenantes. Pour réaliser un changement organisationnel, le CPS aide à la prise de décision en fournissant un « panoptique¹ » des sentiments et des représentations des personnels impactés. Jarrahi met en avant que pour la prise de décision, les CPS apportent une vision complémentaire à celle de l'opérateur. Un CPS et un opérateur agissant de concert sont plus performants que leurs performances individuelles (Steyvers, Tejeda, Kerrigan & Smyth, 2022).

Lors d'une prise de décision assistée par un CPS, l'instant où l'assistance intervient peut différer (Steyvers & Kumar, 2022). Elle est dite « concourante » si l'aide intervient à la survenue du problème, « sur demande » lorsque l'opérateur fait expressément appel au CPS, « séquentielle » lorsque l'opérateur prend une première décision et que l'assistance du CPS vient ensuite, et enfin de type « temporisation » quand le CPS attend avant d'apporter son assistance. La connaissance du CPS par l'opérateur joue également un rôle dans la prise de décision (Steyvers & Kumar, *ibid*). Un opérateur ne connaissant pas le CPS utilise moins l'assistance qu'un opérateur connaissant le CPS. Plusieurs auteurs identifient un besoin d'information pour l'opérateur (Lai et al., 2021 ; Steyvers & Kumar, 2022) : les informations sur la fiabilité du CPS et sur son raisonnement. L'accès à ces informations aide l'opérateur à prendre sa décision et à utiliser le CPS correctement.

¹ Un panoptique est un terme usité en architecture pour parler, dans un bâtiment, d'un pont d'observation permettant d'embrasser du regard tout l'environnement (exemple : panoptique dans un pénitencier). Le terme est utilisé ici pour parler d'une méthode de présentation permettant d'avoir toutes les informations au même endroit.

Le besoin d'information est en lien avec la nécessité de comprendre ce que fait l'autre pour que « tous les membres de l'orchestre » jouent ensemble (Habib et al., 2021). Se pose alors la question du dialogue entre l'humain et le CPS pour qu'ils puissent se coordonner. Pour de nombreux auteurs, il est primordial de mettre l'humain au centre de l'industrie 4.0 afin d'aider à l'intégration de ces nouveaux CPS (Gorecky et al., 2014 ; Romero et al., 2016a ; Guerin, Rauffet, Chauvin & Martin, 2019 ; Fantini et al., 2020 ; Habib et al., 2021 ; Kagerman & Wahlster, 2022). S'assurer d'une bonne intégration de l'humain permet d'améliorer les performances tout en favorisant des conditions de travail (physique, physiologique et mental) de qualité (Roméro et al. (2016b).

Nous avons vu que la prise de décision assistée par un CPS génère de nouvelles interactions pour l'opérateur. Pour mieux comprendre comment améliorer ces nouvelles interactions, nous nous appuyons sur des cas d'application. Ces cas d'applications permettent de faire le lien entre le champ théorique et le champ pratique.

1.4. Cas d'application de la prise de décision assistée par un CPS

Dans le cadre de cette thèse, nous investiguons la prise de décision assistée par un CPS sur des cas d'application issus de deux projets. Ces projets s'inscrivent tous les deux dans le contexte de l'industrie 4.0, mais traitent de différentes facettes de la prise de décision assistée par un CPS.

Projet Humanism. Le projet Humanism est financé par l'Agence Nationale de la Recherche (N° ANR-17-CE10-0009) et est porté par trois partenaires universitaires : le Lab-STICC UMR CNRS 6285, le CReSTIC EA 3804, et le LAMIH UMR CNRS 8201. Un des objectifs de ce projet est d'améliorer le dialogue et les interfaces humain-machine dans l'industrie 4.0. Dans ce projet, l'opérateur réalise des fonctions de supervision, de gestion et de production sur la ligne (Figure 1.6). L'interaction est de niveau tactique pendant la production (Figure 1.8 pastille orange). L'interaction se fait avec un cobot possédant des compétences de diagnostic et de supervision de stock. Ce cobot a comme activité la réalisation de commandes. L'opérateur est dans une interaction cognitive avec le cobot car il supervise le déroulé de la tâche de celui-ci, sans intervenir directement dessus (Figure 1.9 pastille orange). Dans ce

projet, la prise de décision assistée par le CPS est « concourante », car le cobot propose son assistance au même instant que la survenue du problème.



Figure 1.6 : Photo de la plateforme SCAP utilisée dans le projet Humanism

Projet Seanatic. Le projet Seanatic, financé par l'ADEME, est porté par cinq partenaires : le Lab-STICC (UMR CNRS 6285), et les entreprises Azimut (spécialisée dans l'interface), IoT.bzh (spécialisée dans les capteurs l'internet des objets), Thalos (spécialisée dans les communications satellites) et Piriou (spécialisée dans la construction navale). L'objectif du projet est de proposer une solution complète (IOT, IA, communications et interface humain-machine) de maintenance prévisionnelle pour le milieu maritime (Figure 1.7). La maintenance prévisionnelle nécessite de coopérer avec un CPS, qui prend ici la forme d'une IA. L'interaction avec cet agent est de niveau tactique en amont de la production (Figure 1.8, pastille bleue). Cette IA a comme activité le pronostic d'état des équipements du navire. Elle possède des compétences d'analyse, de diagnostic et de calcul. Le fait que l'opérateur soit en interaction avec l'IA permet de la classer comme relevant d'une interaction cognitive

(Figure 1.9, pastille bleue). Dans ce projet, la prise de décision assistée par le CPS est « concourante », car l'IA propose son assistance au même instant que la survenue du problème.



Figure 1.7: Photo du navire utilisé pour le projet Seanatic (Julien Talec, 2013)

- Thématique explorée grâce à Humanism
- Thématique explorée grâce à Seanatic

<p>Niveau stratégique <i>Processus de management</i></p>	<div style="border: 1px solid black; padding: 5px;">Management d'entreprise et planning à long-terme</div>	
<p>Niveau tactique <i>Processus de support</i></p>	<div style="border: 1px solid black; padding: 5px;">● Fonctions de planification à moyen et court terme</div>	<div style="border: 1px solid black; padding: 5px;">● Fonctions de supervision de production et maintenance</div>
<p>Niveau opérationnel <i>Processus de production</i></p>		<div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">Fonctions opérationnelles de maintenance</div> <div style="border: 1px solid black; padding: 5px;">Fonctions opérationnelles de production</div>
	En amont de la production	Pendant la production

Figure 1.8 : Positionnement des fonctions explorées dans Humanism et Seanatic (inspiré de Rauffet, Guerin, Chauvin & Martin, 2018)

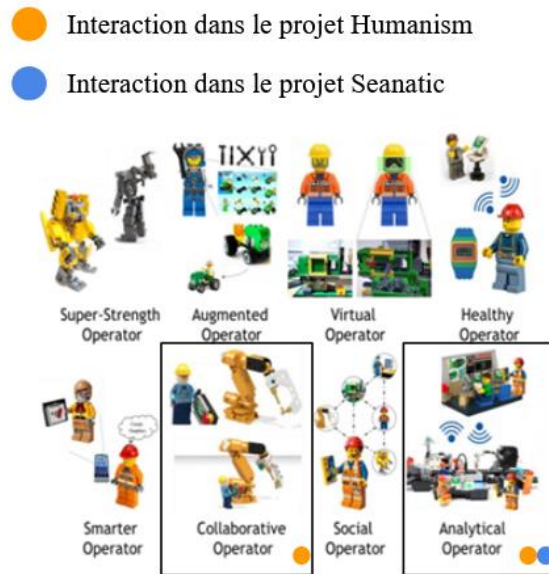


Figure 1.9 : Positionnement de l'opérateur Humanism et Seanatic au sein de la typologie de Romero et al. (2016b)

1.5. Synthèse du chapitre 1

Nous vivons actuellement un tournant industriel avec l'intégration massive de CPS dans les industries. Les industries 4.0 visent à répondre aux changements de consommation et sociétaux (personnalisation et flexibilisation des moyens de production) (Kohler et Weiss, 2016) et aux enjeux climatiques (industries moins polluantes) (Kagerman et Wahlster, 2022). Les CPS, grâce aux évolutions technologiques telles que les IOT ou les algorithmes d'apprentissage des IA, viennent aider les opérateurs à tous les niveaux et pour toutes les fonctions de l'industrie (Rauffet et al., 2018). Ces nouvelles aides apportent leur lot de dangers et de challenges à résoudre. L'accès à de nombreuses données modifie le travail de l'opérateur, le rendant plus complexe d'un point de vue cognitif (Madonna et al., 2019). L'intégration de ces CPS crée de nouvelles interactions humain-machine qu'il est nécessaire de comprendre et d'étudier pour aider à la transition vers l'industrie 4.0.

Les CPS et les IA sont de plus en plus utilisés comme aide à la décision dans de multiples domaines (Lai et al., 2021). Nous avons des exemples dans le domaine juridique (Kleinberg, Lakkaraju, Leskovec, Ludwig & Mullainathan, 2018), le médical (Wu et al., 2019 ; Mckinney et al., 2020), la finance (Day, Cheng & Li, 2018) ou le militaire (Rasch, Kott & Forbus, 2003). L'industrie n'y échappe pas (Jarrahi, 2018 ; Stauder & Kühn, 2021 ; Ahmed, Jeon & Piccialli, 2022 ; Senoner, Netland & Feuerriegel, 2022) grâce à la capacité des CPS d'acquérir,

d'agréger, d'analyser et de restituer des informations. L'opérateur, en sa qualité de superviseur, utilise les propositions formulées par les CPS pour l'aider dans ses prises de décision. Cependant, Kagerman et Wahlster (2022) considèrent que les mauvaises interactions et les mauvaises actions de l'opérateur sont des sources typiques d'erreurs dans l'industrie 4.0. Ces mauvaises interactions proviennent d'une non-détection d'anomalie ou de défaut de production. Les opérateurs peuvent adopter, quant à eux, des solutions non adaptées pour résoudre le problème et la situation. Reiman, Kaivo-oja, Parviainen, Takala et Lauraeus (2021) pointent un manque de maturité du domaine des facteurs humains ou de l'ergonomie dans l'industrie 4.0. Pour répondre à ces deux problèmes, cette thèse propose d'améliorer la compréhension de la prise de décision assistée par un CPS en mettant l'humain au centre de la réflexion.

Chapitre 2 : Coopération humain-machine pour la prise de décision assistée par un CPS : la confiance comme élément central

Dans ce chapitre nous abordons la prise de décision assistée par un système cyber-physique (CPS) sous l'angle des facteurs humains. Cette interaction pose la difficulté de coordonner les informations et les propositions fournies par le CPS avec l'expertise de l'opérateur. L'opérateur étant superviseur, le choix final, et la responsabilité qui en découle, lui reviennent. Il est donc nécessaire de comprendre la nature de l'interaction et le processus de décision de l'opérateur.

Ce type d'interaction est à considérer comme une coopération entre l'opérateur et le CPS. La coopération implique que les actions entreprises par les agents puissent entrer en interférence (Hoc, 2001). Pour aider à l'amélioration de la coopération, nous pouvons la catégoriser selon trois angles. 1) Quelles interférences sont générées ? 2) Comment ces interférences peuvent-elle être résolues. 3) Quels sont les rôles et les relations entre l'opérateur et le CPS. Comprendre la coopération sous ces trois angles permet de préciser ce qu'est une prise de décision assistée par un CPS avec un regard facteurs humains.

La prise de décision assistée par un CPS est également le fruit de plusieurs déterminants liés à l'opérateur tels que la confiance, la conscience de la situation et la charge mentale (Parasuraman, Sheridan & Wickens, 2008 ; Kintz, Banerjee, Zhang, Anderson & Clark, 2022). La définition de ces déterminants nous permet de comprendre quels sont leurs effets sur la prise de décision de l'opérateur. Ces déterminants permettent également d'explicitier la perception de l'assistance et du CPS par l'opérateur. La compréhension de ces déterminants nous permet d'identifier les leviers sur lesquels agir pour aider à la coopération.

Pour comprendre le lien entre les déterminants et la prise de décision de l'opérateur assistée par un CPS nous proposons un nouveau modèle pour représenter cette classe de problème. Ce modèle est basé sur les modèles de la prise de décision proposé par Chancey et al. (2017) et sur celui de la confiance de Hoff & Bashir (2015). L'utilisation de ce modèle nous permet de comprendre comment agir sur les déterminants et leur lien avec la prise de décision de l'opérateur.

2.1. Les modèles de CHM pour la prise de décision assistée par un CPS

Nous avons montré dans le chapitre précédent que l'intégration des nouveaux CPS dans l'industrie 4.0 amène à de nouvelles interactions entre l'opérateur et le CPS. Ces interactions interviennent à des niveaux tactiques et opérationnels et pour des fonctions de production et de maintenance. Les compétences de ces CPS avec l'utilisation du big data, et leurs capacités de calcul, leur permettent notamment de formuler des propositions aux opérateurs (Kagerman & Wahlster, 2022). Ces nouvelles interactions dépassent le cadre de l'utilisation du CPS comme un simple outil mais plutôt comme un coéquipier pour atteindre un but commun (Weiss et al., 2021).

2.1.1. Définition de la coopération humain-machine

L'interaction est un terme englobant traitant de toutes les formes d'actions réciproques que peuvent exercer deux agents l'un sur l'autre. Il existe cependant plusieurs types d'interaction (Rogalski, 1994). La coopération est le type d'interaction sur lequel nous mettons notre focus. A la différence de la collaboration, qui est l'atteinte d'un objectif par une tâche commune aux agents, la coopération est l'atteinte d'un objectif par des sous-tâches différentes entre les agents (Weistroffer, 2014). Comme vu dans le chapitre précédent, l'opérateur et le CPS n'ont pas les mêmes capacités et n'exercent pas les mêmes tâches afin d'arriver au but commun. Le type d'interaction que nous étudions s'inscrit donc dans le cadre de la coopération.

Hoc (1996) relève cinq aspects récurrents dans la littérature permettant de définir la coopération « - l'existence d'un objectif commun entre les agents et d'un plan qu'ils partagent ; - la reconnaissance ou l'inférence par l'un des agents des intentions d'action de l'autre agent ; - le partage dynamique des tâches ou l'évolution des rôles ; - la facilitation de l'atteinte des buts de l'autre ; - la résolution de conflits » (page 170). Hoc (2001) précise également qu'une interaction est de type coopération si les agents ont au minimum :

- Des buts amenant à des **interférences**. Ces interférences peuvent porter sur un niveau stratégique (exemple : la définition des buts diffère), tactique (exemple : les processus et la planification pour atteindre ces buts diffèrent) ou opérationnel (exemple : conflit sur le partage de ressources ou sur la réalisation d'une tâche). Ces interférences

peuvent être positives comme négatives (respectivement l'ajout ou le retrait d'une ressource par exemple).

- La possibilité pour l'opérateur, comme pour le CPS, de **gérer ces interférences** pour faciliter ses tâches individuelles ou pour aider à la réalisation du but commun. Il n'est pas obligatoire que l'humain et l'agent soient tous deux en capacité de gérer ces interférences pour être en coopération. Si seulement l'un des deux est en capacité de gérer les interférences, alors ils seront considérés en situation de coopération.

Dans la coopération, les activités de l'opérateur et du CPS ne sont pas indépendantes, c'est ce qui va créer les interférences. Hoc (ibid) et Loiselet et Hoc (2001) distinguent quatre catégories d'interférence :

- **L'interférence de condition préalable** apparaît dans les situations où la tâche d'un agent humain (ou d'un CPS) ne peut être réalisée qu'après la réalisation de la tâche d'un autre agent humain (ou d'un CPS). Il n'est pas nécessaire que le premier agent communique sur l'accomplissement de sa tâche car le second peut inférer sur cet accomplissement en fonction du contexte. Cette interférence crée de la dépendance entre les deux agents et impacte la coordination. Par exemple sur une ligne de production, un agent devra attendre de recevoir la pièce produite par son collègue en amont avant de pouvoir la peindre.
- **L'interférence bilatérale (dite d'interaction selon Hoc, 2001)** apparaît dans la situation où les agents peuvent se gêner l'un l'autre dans l'accomplissement de leurs objectifs individuels ou communs. Pour Loiselet et Hoc (2001) « *deux activités sont en interférences d'interaction si les ressources, les procédures d'atteintes, les conditions de réalisation ou les résultats de ces activités s'affectent mutuellement, mettant en péril la réalisation de l'un ou des deux activités* » (p. 173). Dans le cadre d'un objectif individuel, un changement d'action en amont de la ligne, généré par un agent, va venir gêner la réalisation de la pièce du second agent. Dans le cadre d'un objectif commun, si deux agents n'ont qu'un seul crayon pour écrire leurs idées lors d'un processus d'idéation, ils se gêneront l'un l'autre lorsqu'ils veulent poser des idées à l'écrit. Cette interférence viendra gêner l'objectif commun qui est de trouver un maximum d'idées.

- **L'interférence de contrôle mutuel** apparaît dans une situation où un agent contrôle la tâche (par exemple en la simulant mentalement) qu'un second agent réalise réellement. Si l'évaluation est positive (i.e. il existe un accord entre les deux agents) alors l'interférence est résolue directement. En revanche, lorsqu'il y a un désaccord entre les deux agents, un compromis doit être trouvé. Par exemple, un agent sur une ligne de production remplit sa commande, et le superviseur va mentalement réaliser la même tâche selon la procédure. S'il apparaît un écart entre ce qu'il imaginait et ce que l'agent fait, alors un compromis devra être trouvé pour résoudre ce conflit. Dans l'aviation, ce type d'interférence peut se constater à un niveau tactique lorsque le contrôleur aérien contrôle mentalement la planification de trajectoire réalisée par le pilote.
- **L'interférence de redondance** apparaît dans des situations non planifiées et planifiées. Dans une situation non planifiée, la tâche ne peut pas être attribuée à l'opérateur ou au CPS en amont dû à une absence de connaissances sur les performances potentielles des agents. C'est l'agent avec les caractéristiques les plus appropriées (emplacement, charge mentale, ou autre) au moment où la tâche doit être réalisée qui devient chargé de sa réalisation. Par exemple, l'opérateur et le CPS peuvent tous deux accomplir une tâche de port de charge (ex. soulever un sac), mais c'est l'agent qui est le plus près qui réalisera cette tâche. Dans une situation planifiée, les deux agents réalisent la tâche en même temps jusqu'à ce que l'un des deux atteigne l'objectif. Imaginons une ligne de production où il y a une alerte sur une pièce défectueuse. Les deux agents vont faire la tâche de supervision en simultanée jusqu'à ce qu'un des deux agents repère ladite pièce. De la même façon, deux agents peuvent imaginer une nouvelle planification (à un niveau tactique) ou de nouveaux buts (à un niveau stratégique) en même temps, et seule une des deux propositions sera sélectionnée.

La coopération se distingue par la présence de diverses interférences entre l'opérateur et le CPS. La façon dont ces interférences sont résolues peuvent être comprises au travers différentes modélisations de la coopération humain-machine (CHM). Ces modélisations permettent de comprendre quel est le rôle de l'opérateur et du CPS dans la coopération pour ensuite comprendre quels sont les moyens à disposition pour les résoudre.

2.1.2. Gestion des interférences dans la coopération : approche structurelle et fonctionnelle

Les deux modèles présentés ici vont permettre de comprendre quel est le rôle du CPS et quel est le rôle de l'opérateur au sein de la coopération. En fonction de leur position hiérarchique, si l'un est supérieur de l'autre, la façon de résoudre l'interférence sera différente. Le premier modèle auquel nous nous intéressons est celui proposé par Schmidt, Rasmussen, Brehmer et Leplat (1991). Cette modélisation considère trois types de coopération :

- **La coopération augmentative** : afin d'accroître les capacités d'un agent seul (capacité cognitive ou physique) pour réaliser sa tâche, il est possible d'ajouter un second agent. Ce second agent a des compétences identiques et permet, notamment, de pallier une augmentation de la charge de travail si celle-ci n'est pas gérable par un seul agent. Par exemple, sur une chaîne production, l'ajout d'un CPS faisant la même activité qu'un autre opérateur sera considéré comme de la coopération augmentative.
- **La coopération intégrative** : pour réaliser une tâche, il est parfois nécessaire de s'appuyer sur des compétences qu'un seul agent ne possède pas. La coopération intégrative s'appuie sur les différentes compétences de plusieurs agents pour permettre de réaliser la tâche. Ici l'exemple serait qu'un opérateur soit le seul à avoir une licence pour conduire un chariot télescopique, et qu'il aide un autre opérateur à déplacer des caisses. Un autre exemple est celui d'un opérateur qui ne pourra pas calculer les probabilités de survenue d'un événement en utilisant de nombreuses données, mais qu'un CPS pourra effectuer.
- **La coopération débative (ou confrontative (Hoc, 1996))** : les capacités de raisonnement d'un agent sont limitées par ses biais cognitifs. Pour pallier ce problème, il est possible de confronter les résultats de différents agents. Cette confrontation va permettre de débattre autour de différentes solutions et d'en dégager une plus appropriée que les autres. Cette vérification et ce contrôle des propositions par l'un et l'autre des agents permet notamment de réduire les erreurs. Par exemple, deux agents vont opposer leurs idées, comme augmenter la cadence ou retarder des commandes, et leur débat pourra faire émerger une solution commune.

Le second modèle est celui proposé par Millot et Mandiau en 1995. Ce modèle a été pensé sur la base d'observations de situations de coopération dans le domaine de l'aviation et de la gestion aéroportuaire. Les auteurs ont distingué deux types de CHM. Le premier type de coopération est **vertical** (Figure 2.1). L'opérateur est le seul à prendre la décision. Pour cela il est aidé par le CPS qui peut lui prodiguer des conseils ou des informations. Ici, l'opérateur a un rôle décisionnaire et le CPS un rôle d'aide à la décision. Ce type de coopération nécessite de fait une compréhension et un dialogue entre les deux agents. Outre la nécessité de prendre en compte les limites cognitives de l'opérateur (ce qui peut être mis en parallèle avec l'analyse de Woods et al. (2002) sur l'utilisation de nombreuses données issues des nouvelles technologies et le risque que cela représente pour l'opérateur), ce dialogue entre les deux agents soulève plusieurs éléments pouvant être problématiques :

- La représentation des actions des agents et leurs effets dans l'environnement
- La représentation des croyances
- La représentation de la prise en compte des intentions des agents
- La représentation du plan d'action

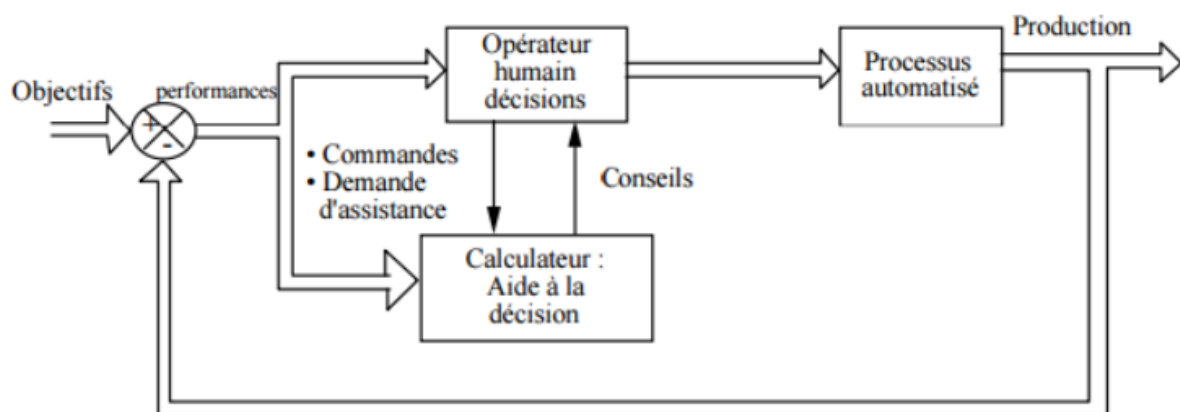


Figure 2.1 : Coopération verticale (Millot, 1999 dans Kostenko, 2017)

Le second type de coopération est **horizontal** (i.e. les agents sont au même niveau hiérarchique et se partagent les tâches) (Figure 2.2). Ici, les CPS, au même titre que les opérateurs, sont capables de superviser et de prendre des décisions. Ce type de coopération fait émerger la possibilité d'une réallocation dynamique des fonctions entre le CPS et l'opérateur en fonction des tâches à réaliser, des compétences des agents ou de la charge de travail. Cette réallocation des fonctions peut être implicite (sur base des états de l'opérateur

tels que le stress, le taux de charge, etc.) ou explicite (en fonction d'une nouvelle stratégie ou d'une nouvelle organisation par exemple). Il est important de noter que ce type de coopération pose la difficulté de communiquer sur l'allocation des fonctions entre les agents, ceux-ci étant changeants. Millot (1988) indique que par essence, lié aux spécificités de la CHM (Hoc, 1996), la coopération horizontale aura toujours une part de verticalité car l'opérateur reste, in fine, superviseur du CPS.

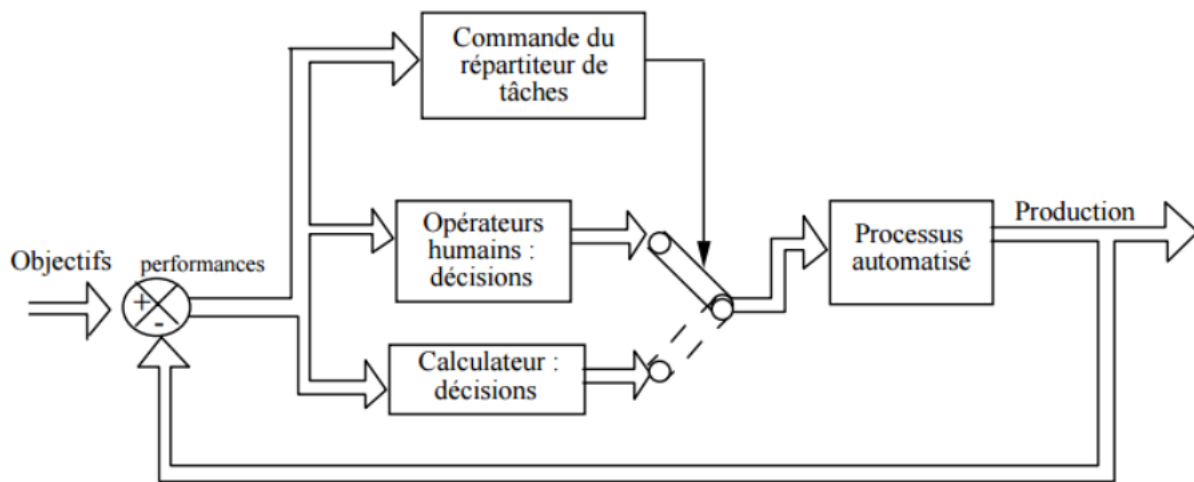


Figure 2.2 : Coopération horizontale (Millot, 1999 dans Kostenko, 2017)

Enfin, Millot et Mandiau (1995) n'excluent pas le fait de mixer ces deux types de coopération dans une perspective d'organisation multi-niveaux. Cette organisation se compose d'un opérateur aidé d'un CPS (coopération verticale) qui prend la décision de l'allocation des fonctions entre un second opérateur et un second CPS (coopération horizontale).

Ces deux modèles présentés permettent de mieux appréhender la façon dont peut s'organiser une coopération entre un opérateur et un CPS. Ces modèles permettent également de comprendre qu'il existe un lien hiérarchique puisque l'opérateur reste superviseur du CPS. L'opérateur étant décisionnaire, la résolution de l'interférence se fait avec sa prise de décision. Il reste cependant à comprendre comment et par quels moyens l'opérateur et le CPS peuvent communiquer pour s'accorder et améliorer la gestion des interférences.

2.1.3. Maintien et actualisation du COFOR pour la résolution d'interférences

Hoc (2001) propose un modèle cognitif de la coopération en situation dynamique afin d'explicitier la gestion des interférences. Cette approche se base sur l'idée que la CHM est semblable à une coopération humain-humain. Le fait que les CPS soient autonomes et capables de prendre des décisions en fait d'elles plus que de simples outils ou assistants. Pour cet auteur, il existe cependant deux spécificités dans la CHM. La première spécificité est que les capacités de l'opérateur ne peuvent pas être transférées dans le CPS, limitant une réplification stricto-sensu des modèles de coopération humain-humain. La deuxième spécificité est que l'opérateur reste superviseur du CPS, le rendant responsable du comportement du CPS. Il existera donc toujours une relation « chef/subordonné » dans le cadre d'une CHM. Une coopération réussie, selon Hoc, sera basée sur l'utilisation d'un référentiel commun (Common Frame Of Reference ou COFOR). Ce COFOR correspond à des connaissances partagées, des croyances et des représentations (Hoc, 2001). Ce référentiel commun, permettant d'améliorer la coopération entre un opérateur et un CPS (Hoc, 2001), est composé de trois niveaux (Figure 2.3).

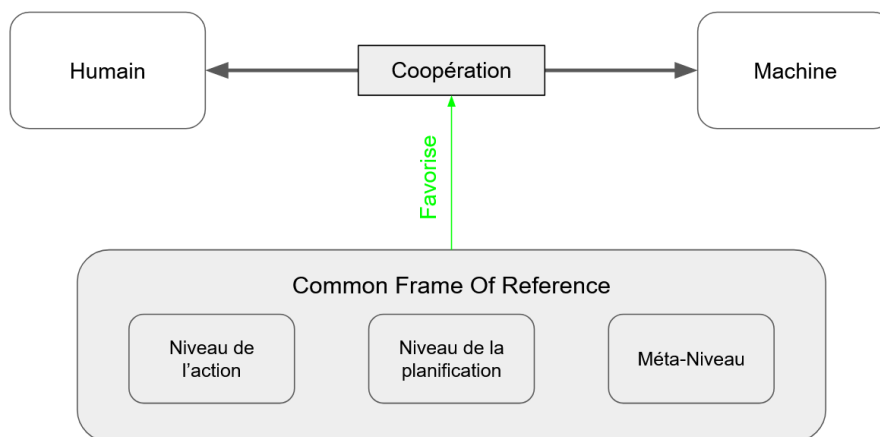


Figure 2.3 : Coopération humain-machine selon Hoc (2001)

Le premier niveau correspond aux communications sur les actions entreprises par le CPS (i.e. coopération dans l'action). Ce niveau contient les informations sur les activités en cours et à court terme de même que les interférences locales (leur création, détection, anticipation et résolution). Le second niveau correspond au partage des représentations permettant de guider l'activité (i.e. coopération dans la planification). Ce niveau du COFOR contient notamment les buts communs, les plans et l'allocation des fonctions (i.e. comment sont

réparties les tâches entre les coopérants). Le troisième niveau correspond aux méta-connaissances (i.e. méta-coopération). Ces méta-connaissances sont les connaissances que l'opérateur (ou le CPS) a sur ses propres activités et sur celles des autres. Ce méta-niveau, qui passe par l'utilisation de codes communs, aide à prévenir les risques et facilite la prise de décision en termes de coûts/bénéfices.

Ce modèle de la coopération centré autour d'un COFOR permet d'appliquer des principes de conception des systèmes humain-machine (Hoc, 2001). Pour favoriser la coopération au niveau de l'action, l'un des principes est de tendre vers l'identification en amont des interférences qui vont survenir, ainsi que leur mise à jour régulière. Afin de faciliter la coopération au niveau de la planification, des principes de conception existent pour permettre la création d'un COFOR, pour faciliter l'actualisation d'un COFOR existant ou bien de communiquer sur l'allocation des fonctions entre les agents coopérants. Enfin, le méta-niveau de la coopération sera lui favorisé par l'utilisation de codes communs spécifiques au domaine, ou grâce à l'intégration des informations concernant l'autre agent. Cela permet de prédire les comportements de l'autre, ce qui aura pour effet d'aider à la coopération.

Millot et Pacaux-Lemoine (2013) proposent une synthèse des travaux portant sur la coopération et notamment en se basant sur les trois modèles précédemment présentés. Selon leur proposition, chaque agent (opérateur ou CPS) possède deux dimensions, chacune composée d'aptitudes internes et externes (Figure 2.4).

La première dimension est le Know-How (KH), « **savoir-faire** ». Cette dimension porte sur les compétences de l'agent (en vue de réaliser une tâche). Les aptitudes internes de cette dimension sont liées aux processus et compétences que l'agent a pour résoudre le problème. Les aptitudes externes, quant à elles, concernent l'acquisition d'informations pertinentes dans l'environnement de l'agent.

La seconde dimension est le Know-How-to-Cooperate (KHC), « **savoir-coopérer** ». Cette dimension porte sur les capacités de l'agent à coopérer avec un autre agent sur une tâche commune (Pacaux-Lemoine, 2020). La dimension KHC permet la gestion des interférences en les détectant, en les créant et en les résolvant (de manière à coordonner les agents). La création d'interférence facilitatrice (i.e. positive) va aider l'autre agent à atteindre

ses buts. Les aptitudes internes vont notamment participer à la création d'un COFOR. Les aptitudes externes sont là pour aider à communiquer avec l'autre agent. Ces communications passent par trois principales sources que sont l'observation directe (mouvements, émotions, etc.), le dialogue et l'observation des effets de l'agent sur la tâche (Pacaux-Lemoine, 2020).

Savoir-Faire (KH) :

- **Aptitudes internes :**
 - liées à la résolution de problème (en fonction du processus)
 - liées aux compétences (SRK, Rasmussen, 1983)
 - liées aux capacités de traitement (notamment des inférences)
- **Aptitude externes** pour :
 - accéder aux informations (des processus et de l'environnement)
 - agir (sur le processus)

Savoir-Coopérer (KHC) :

- **Aptitudes internes** pour :
 - construire un modèle des autres agents (sur le KH/KHC) et créer un COFOR
 - déduire les intentions des autres agents
 - analyser la tâche et comprendre la structure et la forme de la coopération
 - produire un plan d'action commun et le partager
 - se coordonner avec les autres agents
 - accepter de répondre aux requêtes des autres agents
- **Aptitudes externes** pour :
 - communiquer avec les autres agents

Figure 2.4 : Synthèse du KH et KHC (traduit de Millot & Pacaux-Lemoine, 2013)

Il existe trois façons de gérer les interférences (Pacaux-Lemoine & Derbernard, 2002). Il est possible de gérer l'interférence en négociant, ici les agents réduisent la différence entre le COFOR et leur représentation individuelle (i.e. leur cadre de référence). Pour ce faire, soit les agents modifient le COFOR, soit ils modifient directement leur cadre de référence sur la base d'explications entre eux, et de l'activation des aptitudes du KHC. Pour gérer les interférences il est également possible d'accepter (exemple : accepter la proposition d'une machine). Ici l'agent va mettre à jour son cadre de référence pour le faire correspondre au COFOR. Selon les auteurs, cette modalité de gestion de l'interférence intervient lorsque la négociation représente un coût trop élevé ou que l'un des agents souhaite faciliter le travail d'un autre. La troisième façon de gérer une interférence est le fait d'imposer. Ici l'un des agents va imposer son cadre de référence au cadre de référence de l'autre agent.

2.1.4. La prise de décision d'un opérateur assistée par un CPS

La prise de décision d'un opérateur assistée par un CPS crée des interférences pouvant appartenir aux quatre types identifiés par Hoc. Le moment de la survenue de l'assistance du CPS permet d'identifier si la proposition crée une interférence de condition préalable (exemple : l'opérateur doit attendre la proposition pour réaliser son activité), bilatérale (exemple : la proposition interrompt l'opérateur dans son activité), de contrôle mutuel (exemple : l'opérateur doit surveiller que la proposition du CPS soit formulée) ou de redondance (exemple : l'opérateur et le CPS formulent une proposition en parallèle). Si l'opérateur reste superviseur, la résolution de cette interférence peut passer par l'acceptation de l'opérateur, ce qui modifie son cadre de référence, ou par l'imposition, ce qui modifie le COFOR et par extension le cadre de référence du CPS. Cette coopération est verticale car l'opérateur prend la décision finale et supervise le CPS. Dû à l'hétérogénéité des compétences entre l'opérateur et le CPS, cette coopération est également intégrative : le CPS possède des informations et des capacités de calcul que l'opérateur n'a pas. Enfin, cette coopération est débative car la proposition du CPS est confrontée au plan initial de l'opérateur, il y a donc une confrontation entre les deux propositions.

Les deux projets (Humanism et Seanatic) que nous avons présentés dans le premier chapitre n'ont pas les mêmes contextes et ne présentent donc pas les mêmes niveaux d'interférences et situations de coopérations. Le Tableau 2.1 caractérise ces interférences selon les cadres théoriques et conceptuels que nous venons de présenter.

Tableau 2.1 : Classification des projets Humanism et Seanatic selon différentes caractéristiques de la coopération identifiées dans la littérature

		Humanism	Seanatic
Type d'interférence (Hoc, 2001)	<i>Condition préalable</i>	X	X
	<i>Bilatérale (dite d'interaction)</i>	X	X
	<i>Contrôle mutuel</i>	X	X
	<i>Redondance</i>		
Niveau de l'interférence (Hoc, 2001)	<i>Stratégique</i>		
	<i>Tactique</i>		X
	<i>Opérationnel</i>	X	
Gestion des interférences (Pacaux-Lemoine & Debernard, 2002)	<i>Négociation</i>		
	<i>Acceptation</i>	X	X
	<i>Imposition</i>	X	X
Structure de coopération (Millot & Mandiau, 1995)	<i>Verticale</i>	X	X
	<i>Horizontale</i>		
Types de coopération (Schmidt et al., 1991)	<i>Augmentative</i>		
	<i>Intégrative</i>	X	X
	<i>Débatrice</i>	X	X

Les travaux de Hoc (2001) ou de Pacaux-Lemoine et Debernard (2002) permettent d'analyser le type et la gestion des interférences. Nos deux cas particuliers de coopération font intervenir des interférences de type et de niveaux différents. En revanche, les possibilités de résolution de ces interférences sont les mêmes. Le Tableau 2.2 permet d'explicitier les types des interférences, et leurs niveaux, au regard des situations rencontrées dans les projets Humanism et Seanatic. Dans ces deux projets, l'opérateur a une activité de **gestionnaire**. Dans le projet Humanism l'opérateur gère la production de la ligne, et dans Seanatic il gère le calendrier de maintenance. Pour rappel, dans le projet Humanism, le CPS a une activité de réalisation de commande et émet des **propositions pour être aidé dans cette activité**. L'opérateur supervise que ces propositions soient en accord avec la gestion de la production.

Dans le projet Seanatic, le CPS a une activité de pronostic et émet des **propositions pour partager ses pronostics** avec l'opérateur. L'opérateur supervise que ces propositions soient en accord avec la gestion du calendrier de maintenance.

Tableau 2.2 : Projet Humanism et Seanatic selon leurs interférences

	Humanism	Seanatic
Types des interférences rencontrées (Hoc, 2001)	<p>Condition préalable : pour gérer la production l'opérateur a besoin de la proposition du cobot sur son besoin en ressource.</p> <p>Bilatérale (interaction) : la proposition du cobot crée un conflit sur les résultats des activités. D'un côté le résultat escompté est la réalisation des commandes local (<i>cobot</i>) et de l'autre le résultat escompté est la gestion de la production globale (<i>opérateur</i>).</p> <p>Contrôle mutuel : pour gérer la production, l'opérateur contrôle la proposition du cobot.</p>	<p>Condition préalable : pour gérer le calendrier de maintenance, l'opérateur a besoin de la proposition de l'IA.</p> <p>Bilatérale (interaction) : la proposition de l'IA crée un conflit sur les procédures d'atteintes pour gérer le calendrier de maintenance. La proposition d'avancer ou reculer (<i>IA</i>) peut mettre en péril la gestion des dates de maintenances (<i>opérateur</i>).</p> <p>Contrôle mutuel : pour gérer le calendrier de maintenance, l'opérateur contrôle la proposition de l'IA.</p>
Niveaux des interférences rencontrées (Hoc, 2001)	Niveau opérationnel : l'interférence porte sur les ressources.	Niveau tactique : l'interférence porte sur les processus de planification de maintenance.
Gestion des interférences (Pacaux-Lemoine & Debernard, 2002)	Acceptation ou imposition : L'opérateur peut choisir d'accepter la proposition (acceptation : il modifie son cadre de référence) ou de la refuser (imposition : il modifie le COFOR).	

Que ce soit avec l'aide d'un CPS de type cobot (Humanism) ou avec l'aide d'un CPS de type IA (Seanatic), nous constatons que ces situations de coopération sont verticales (Millot & Mandiau, 1995). Ces situations de coopération sont à la fois débatives et intégratives (Schmidt et al., 1991). Si les deux projets ont les mêmes typologies de coopérations, leur expression au travers des projets diffère. Le Tableau 2.3 permet d'explicitier la forme que prennent ces situations de coopération dans les projets Humanism et Seanatic.

Tableau 2.3 : Projet Humanism et Seanatic selon les modèles de coopération.

		Humanism	Seanatic
Typologie des coopérations rencontrées	Verticales (Millot & Mandiau, 1995)	L'opérateur va décider de suivre, ou non, la proposition du cobot	L'opérateur va décider de suivre, ou non, la proposition de l'IA
	Intégratives (Schmidt et al., 1991)	Le cobot a accès à des informations inconnues de l'opérateur	L'IA a accès à des informations inconnues de l'opérateur. L'IA a des capacités de calcul plus puissantes que l'opérateur
	Débatives (Schmidt et al., 1991)	Avec l'aide des informations du cobot, l'opérateur peut proposer plusieurs plans d'action pour résoudre l'interférence	Avec l'aide des informations de l'IA, l'opérateur a une flexibilité sur le choix de la date de maintenance.

La prise de décision assistée par CPS se caractérise par de multiples interférences dont la gestion passe, soit par l'acceptation, soit par l'imposition. La gestion de ces interférences se structure de façon verticale, où l'opérateur est le supérieur hiérarchique du CPS et prend la décision, et où le CPS apporte des informations supplémentaires. Afin de communiquer et de coopérer, l'opérateur et le CPS partagent un espace de connaissance commun : le COFOR.

Pour comprendre comment l'opérateur gère ces interférences, il faut s'intéresser à l'impact qu'elles ont sur lui. Dans la CHM, trois déterminants sont mis en avant chez l'opérateur : la confiance, la conscience de la situation et la charge mentale (Parasuraman et al., 2008). Ces déterminants permettent de comprendre **pourquoi** l'opérateur prend telle ou telle décision pour gérer une interférence.

2.2. Les déterminants de l'opérateur dans la prise de décision assistée par un CPS

Le premier déterminant de l'opérateur dans la prise de décision assistée par un CPS est celui de la « **confiance** ». Ce déterminant est considéré comme un des principaux à prendre en compte dans une CHM (Parasuraman & Riley, 1997). Dû à un environnement numérique potentiellement perturbé, le déterminant confiance est d'ailleurs considéré comme pivot pour l'industrie 4.0 (Koh, Orzes & Jia, 2019). Le second déterminant de

l'opérateur est la compréhension de la situation. Cette compréhension de la situation peut être analysée selon deux éléments. Premièrement, la « **conscience de la situation** » qui est directement liée à la prise de décision (Endsley, 1995b ; Stanners & French, 2005). Secondement, la « **perception du risque** » qui joue un rôle dans la coopération avec la machine (Parasuraman & Mouloua, 1996 ; Chancey et al., 2017). Le troisième déterminant est la « **charge mentale** ». Cette charge mentale joue un double rôle à la fois dans l'interaction avec le CPS (Parasuraman & Riley, 1997) et dans le fait qu'elle sous-tende la conscience de la situation (Endsley, 1995b). La charge mentale est également liée à la performance (Young, Brookhuis, Wickens & Hancock, 2015) et à la prise de décision (Soria-Oliver, Lopez & Torrano, 2017). La confiance, la conscience de la situation et la charge mentale sont notamment mises en avant par Parasuraman et al. (2008) comme étant primordiales pour la compréhension et l'amélioration de la performance d'un couple humain-machine dans un système complexe.

2.2.1. La confiance : un déterminant à calibrer

La notion de confiance dans la CHM est importante car c'est un indicateur du futur usage d'une machine (Marsh & Dibben, 2003 ; Sanders, Kaplan, Koch, Schwartz & Hancock, 2019 ; Märtins, Kehl, Westmattelmann & Schewe, 2022) et de son acceptation (de Visser, Pak & Shaw, 2018). La confiance est au centre de multiples recherches et ce depuis de nombreuses années. Il en existe de nombreuses définitions, chaque auteur ou presque ayant la sienne. La définition usuelle du dictionnaire du Larousse est la suivante : « *Sentiment de quelqu'un qui se fie entièrement à quelqu'un d'autre, à quelque chose* ». Cette première définition met en avant la confiance comme le ressenti d'une personne envers une autre personne ou un objet. Pour comprendre cette relation de confiance entre l'opérateur et le CPS, il est possible de se baser sur ce que l'on connaît de la confiance « humain-humain » (de Visser et al., 2018 ; Davis, 2019 ; Stuck, Tomlinson & Walker, 2021 ; Alarcon, Gibson, Jessup & Capiola, 2021). La définition proposée par Mayer, Davis et Schoorman (1995) est la suivante : « *The willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party* » (p. 712). Cette définition précise que la confiance se base sur l'attente d'un comportement spécifique de l'autre, ce qui nécessite de connaître les capacités de l'autre afin de lui faire confiance. Cette définition

indique également que pour faire confiance, il n'est pas nécessaire que la personne soit en position de surveiller ou de contrôler ce que fait l'autre.

Pour Hoff et Bashir (2015), trois éléments sont consistants dans la majorité des définitions de la confiance, et ce à travers les différents domaines qui étudient cette thématique (psychologie, sociologie, philosophie, etc.). Pour qu'une relation de confiance puisse exister entre deux agents il faut trois prérequis :

1. Un triptyque « agent qui fait confiance (« trustor ») - agent à qui l'on fait confiance (« trustee ») – situation »
2. L'agent (humain ou technologique) en qui l'ont fait confiance a un intérêt dans la tâche et la relation. Pour les auteurs, dans le cadre d'une technologie, cet intérêt est celui de l'équipe de conception derrière la technologie
3. Cet agent en qui l'on fait confiance peut échouer ou se tromper.

Comme le précise Hoff et Bashir (ibid), une différence fondamentale entre la confiance interpersonnelle et la confiance envers un CPS réside dans le fait que la confiance interpersonnelle se base essentiellement sur la bienveillance du « trustee » (i.e. vouloir aider l'autre). Là où la confiance entre un humain et le CPS se basera sur les caractéristiques du CPS (ses buts, ses processus, et sa performance). Faire confiance implique que le CPS puisse se tromper ou échouer, ce qui crée de l'incertitude (Stuck et al., 2021). Tout l'intérêt de l'étude de la confiance est de savoir si l'opérateur qui fait confiance dans le CPS a raison de le faire. Dans les travaux sur les interactions humain-machine, la définition de la confiance de Lee et See (2004) est couramment utilisée (4654 citations sur google scholar au 14/06/2023) : « *the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability* » (p. 54). Pour ces auteurs, la confiance est une attitude, une prédisposition mentale, envers autrui. Se basant sur la théorie du comportement (Ajzen et Fishbein, 1977), cette attitude explique la probabilité d'adopter un comportement de confiance. Dans notre contexte, ce comportement de confiance est d'accepter la proposition. Cette définition permet de montrer qu'il y a un rapport d'aide entre deux agents, l'un venant aider le second à réaliser sa tâche. Cette définition met également en avant le fait que dans l'interaction, l'opérateur est en situation d'incertitude et de vulnérabilité. Ces deux éléments sont relatifs à la définition de Mayer et al. (1995) où l'incertitude et la vulnérabilité

proviennent du fait d'accepter de faire confiance et de ne pas pouvoir contrôler ce que fait autrui.

L'attitude de confiance (nous utiliserons le terme **confiance** par la suite) se construit grâce à plusieurs facteurs (Hancock et al., 2011 ; Hoff et Bashir, 2015 ; Davis, 2019 ; Khavas, 2021 ; Yang, Karakaya, Dominioni, Kawabe & Bengler, 2021 ; Kaplan, Kessler, Brill & Hancock, 2023). Ces facteurs peuvent être classés sous trois grandes catégories selon qu'ils soient liés à l'opérateur, au CPS ou à l'environnement :

- **Relatifs à l'opérateur** (« trustor ») :
 - Capacité de l'opérateur : capacité attentionnelle, expertise, compétence, charge cognitive de l'opérateur, compréhension du CPS, expériences précédentes et conscience de la situation
 - Caractéristiques de l'opérateur : démographie, personnalité, attitude envers les CPS, aisance avec les CPS, confiance en soi et facilité à faire confiance
- **Relatifs au CPS** (« trustee ») :
 - Performance du CPS : le comportement, l'indépendance, la fiabilité, la prédictibilité, le niveau d'automatisation, le taux d'échec, le nombre de fausses alarmes et la transparence
 - Attributs du CPS : proximité physique, dialogue, la personnalité du CPS, l'adaptabilité, le type de CPS, le genre et l'anthropomorphisme
- **Relatifs à l'environnement** (« situation ») :
 - Collaboration de l'équipe : l'appartenance au groupe, la culture de l'équipe, le temps d'interaction, les communications et le partage de modèles mentaux communs
 - Tâche : type de tâche, complexité de la tâche, risques liés à la tâche, présence de multi-tâche et le type d'environnement physique

Pour Hoff et Bashir (2015), la confiance se construit à partir d'un processus dynamique et qui est structuré autour de trois composantes interdépendantes (Figure 2.5).

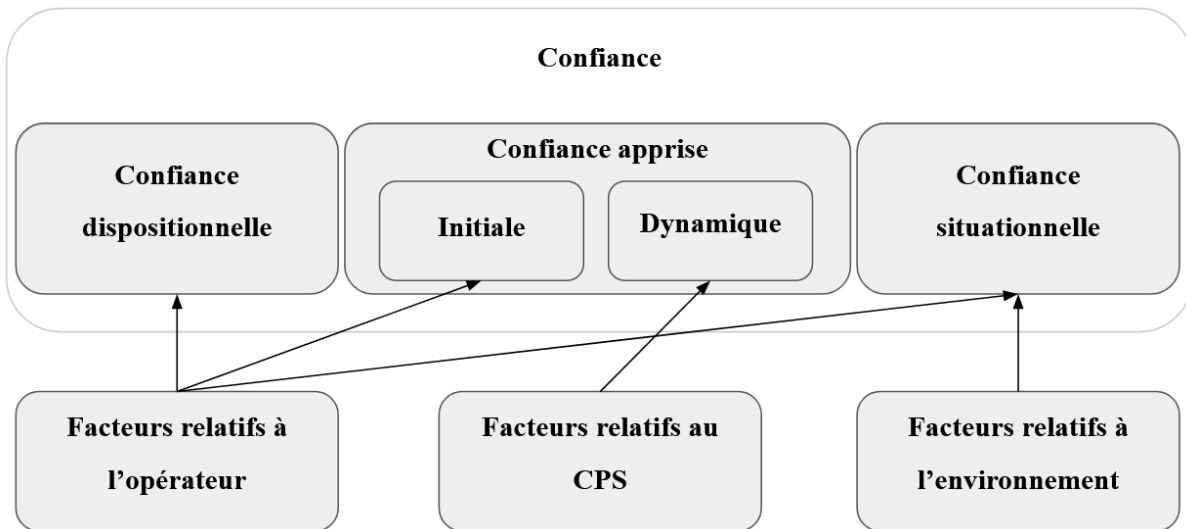


Figure 2.5 : Liens entre les composantes de confiance (Hoff & Bashir, 2015) et les facteurs d'influence
 Les trois grandes catégories de facteurs que nous venons de présenter, et qui permettent la création de la confiance, vont venir influencer sur les différentes composantes de confiance :

- **La confiance dispositionnelle (CD)** est la confiance préalable qu'éprouve l'opérateur envers les CPS et qui provient des facteurs relatifs à l'opérateur. Nous y retrouvons ses caractéristiques sociales (âge, genre, culture, etc.) et ses traits de personnalité (attitude envers la technologie, propension à faire confiance, etc.). Elle ne s'applique pas à un CPS ou à une situation particulière et cette confiance n'est pas dépendante du contexte.
- **La confiance apprise** est la confiance envers un système particulier et provient des facteurs relatifs à l'opérateur et des facteurs relatifs au CPS. Cette confiance est « initiale » lorsqu'elle se situe avant une interaction, ou « dynamique » dès lors que l'opérateur est en interaction avec le CPS. La **confiance apprise initiale (CAI)** est le produit des expériences passées (attitudes, expériences vécues, réputation et compréhension du système). La **confiance apprise dynamique (CAD)** provient des caractéristiques du CPS (anthropomorphisme, facilité d'usage, communication, transparence et niveau de contrôle) et de sa performance (fiabilité, prédictibilité, fausses alarmes, etc.).
- **La confiance situationnelle (CS)** est la variation de la confiance accordée au CPS selon la situation et provient des facteurs relatifs à l'environnement et de ceux relatifs à l'opérateur. On y retrouve plusieurs éléments externes (exemple : type de système,

complexité de la tâche, risques et bénéfices de la tâche, situation organisationnelle) et des éléments internes à l'opérateur liés à une situation donnée (confiance en soi, expertise, humeur, et capacité attentionnelle). Contrairement à la confiance dispositionnelle, la confiance situationnelle est dépendante du contexte de l'interaction.

Pour qu'une CHM soit performante il serait tentant de penser que la confiance doit être « haute », c'est-à-dire que l'opérateur fasse hautement confiance au CPS. Au même titre que ne pas faire confiance à la machine recèle des dangers (exemple : ne pas prendre en considération les alertes que le CPS transmet à l'opérateur), faire trop confiance au point de ne jamais mettre en doute le CPS est risqué (exemple : se baser exclusivement sur les informations de la machine) (Parasuraman & Riley, 1997). Dans le cas d'une surconfiance ou d'une sous-confiance, le risque est que l'opérateur n'interagisse pas avec le CPS de façon appropriée (« Misuse » ou « Disuse » selon Parasuraman & Riley, *ibid*). Un opérateur qui est en sous-confiance dans la machine peut se retrouver à négliger l'aide que le CPS peut lui proposer, et ce même si cela lui apporte des bénéfices. A l'inverse, être en surconfiance peut amener un opérateur à utiliser l'aide que le CPS va proposer, même lorsque cette aide n'est pas appropriée ou qu'elle dépasse les compétences réelles du CPS. Cette confiance doit donc être calibrée afin que l'interaction et la coopération avec le CPS soit appropriée (de Visser et al., 2020). Cette calibration de la confiance est définie par Théodorou, Wortham et Bryson (2017) comme le moment où l'opérateur a réussi à se créer un modèle mental du système comprenant ses capacités et ses limites (Figure 2.6). C'est l'adéquation entre ce que l'opérateur pense que le CPS est capable de faire (i.e. les capacités perçues) et ce que le CPS peut réellement faire (i.e. les capacités réelles).

L'objectif est donc d'avoir une confiance envers un CPS qui soit appropriée ou calibrée. Fort d'une confiance calibrée, l'opérateur est en capacité de discerner les situations où il peut coopérer avec le CPS, en acceptant une proposition par exemple, et les moments où il est préférable de ne pas se fier au CPS, en refusant la proposition.

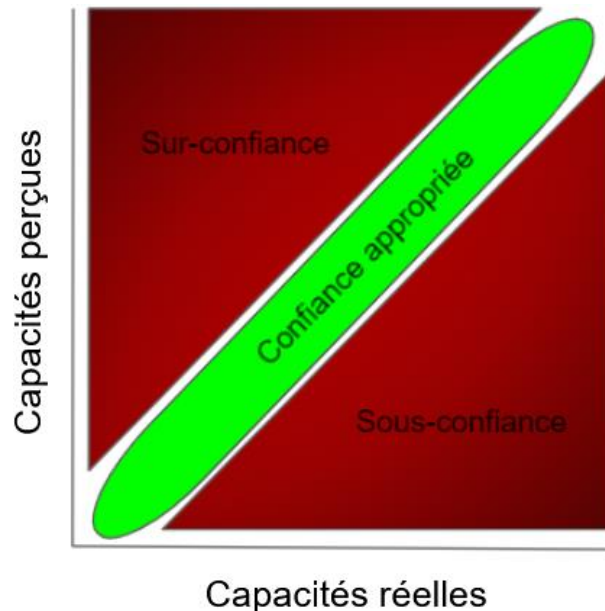


Figure 2.6 : Surconfiance, confiance appropriée et sous-confiance selon les capacités perçues et réelles d'un CPS (selon Théodorou, Wortham & Bryson, 2017)

2.2.2. La conscience de la situation et la perception du risque : des déterminants à maintenir

La conscience de la situation (SA ou « Situation Awareness » en anglais) est définie par Endsley (2006) comme « *the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future* » (p. 529). L'un des objectifs d'une bonne conscience de la situation est de permettre une prise de décision pertinente grâce à un modèle mental adapté à la situation. Endsley (1995a) propose une formalisation de la conscience de la situation divisible en trois niveaux correspondant à la perception, à la compréhension et à la projection (Figure 2.7).

Le **premier niveau** de la conscience de la situation est la perception des éléments de l'environnement et de la situation actuelle. Un opérateur sur une ligne de production percevra son environnement direct (exemple : sa ligne de production, les automates, les CPS, etc.). Le **second niveau** de la conscience de la situation est la compréhension de la situation. L'opérateur va mettre en lien les différentes données qu'il perçoit pour avoir une compréhension de la situation (exemple : la ligne et les automates sont arrêtés, signifiant que la production est stoppée). Le **troisième niveau** de la conscience de la situation est la projection des états futurs de la situation. La compréhension de la situation, des enjeux et de

leur dynamique va permettre à l'opérateur de prévoir ce qu'il va se passer (exemple : la ligne ne va pas redémarrer sans l'aide de l'opérateur).

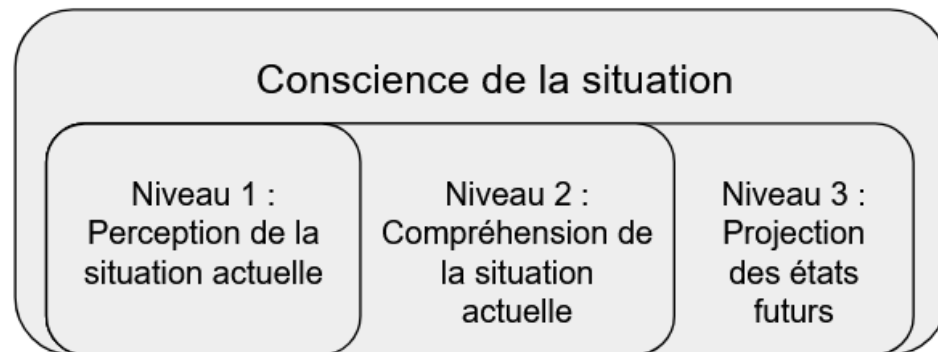


Figure 2.7 : Les trois niveaux de la conscience de la situation selon Endsley (1995a)

Endsley précise en 2015 que la conscience de la situation n'est pas composée de niveaux linéaires. Le modèle prévoit en effet que son utilisation puisse être dirigée par le but, et pas seulement par les données. Une personne ayant une SA de niveau 2 ou 3 pourra chercher des informations de niveau 1 pour améliorer sa conscience de la situation 3. L'idée est que la construction de la conscience de la situation est itérative. Elle se base sur un processus continu d'acquisition d'informations sur la situation et de mise à jour du modèle mental créé.

La conscience de la situation est supposée être un prédicteur de la confiance (Hancock et al., 2011). Les résultats montrent qu'avoir une meilleure conscience de la situation permet d'améliorer la confiance d'un individu dans un CPS (Sonoda & Wada, 2017 ; Petersen, Robert, Yang & Tilbury, 2019). Une meilleure conscience de la situation amène également à une meilleure performance (Petersen et al., 2019). Dans les études s'intéressant à un CPS comme support d'aide à la décision, la SA n'est que rarement mesurée (van de Merwe, Mallam & Nazir, 2022, Table II). L'étude de Roth, Schulte, Schmitt et Brand (2020) permet toutefois de voir que les conditions expérimentales ayant la meilleure performance sont également celles dans lesquelles l'humain a une meilleure conscience de la situation. Les auteurs n'ont pas effectué de corrélation entre la performance et la conscience de la situation, ce lien est donc purement descriptif. Sanders, Oleson, Billings, Chen et Hancock (2011) différencient la conscience de la situation et la perception du risque. Pour eux ce sont deux éléments importants pour comprendre la coopération humaine-machine dans des contextes à risques.

La conscience de la situation est un concept holistique de compréhension de l'environnement (Kinateder, Kuligowski, Reneke & Peacock 2015), ce qui inclut la perception du danger. Au premier niveau de la conscience de la situation, des éléments sont repérés dans l'environnement (exemple : de la fumée, une diode rouge sur le four). Au second niveau l'opérateur met en lien ces deux éléments (exemple : le four prend feu). Le troisième niveau amène à l'évaluation du risque (exemple : il existe une probabilité d'être exposé au feu et de se brûler). Cette évaluation du risque (probabilité d'exposition à un danger et gravité de la conséquence) est à différencier de la perception subjective du risque, qui est une attitude envers un événement. Cette attitude est basée sur les expériences personnelles et les croyances de l'opérateur envers ce danger (Slovic, 2000). Liu, Hirayama, Saiki et Murase (2020) proposent un modèle combinant la conscience de la situation selon Endsley et la perception subjective du risque en tant qu'attitude. Dans leur modèle de prise de décision, la conscience de la situation précède la perception du risque (Figure 2.8).

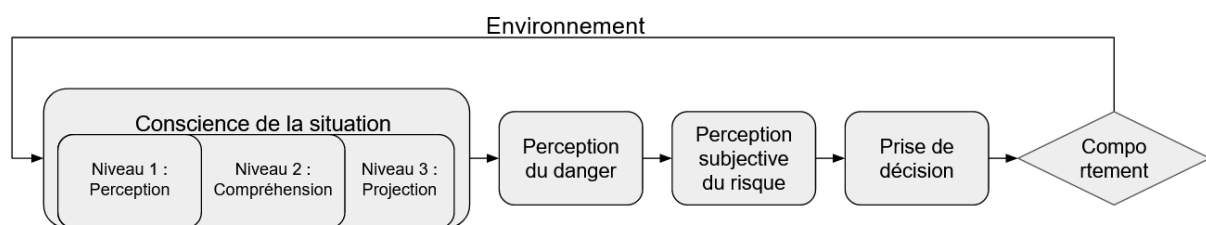


Figure 2.8 : Modèle de prise de décision incluant la conscience de la situation et la perception du risque (Liu et al., 2020)

La conscience de la situation est une première étape permettant de réunir les informations sur l'environnement. Une fois ces informations réunies, l'opérateur définit si cela constitue un danger par rapport à ses expériences passées et ses connaissances (exemple : est-ce qu'il est normal que de la fumée s'échappe du four ?). La perception du risque est différente selon le sentiment de menace ressenti par l'individu. Cette perception du risque est ensuite médiatrice entre la perception du danger et la prise de décision. Dans la situation où de la fumée s'échappe du four, un opérateur peut se dire que cette situation ne représente que peu de risque et ne nécessite pas d'action à mettre en place.

De ce fait, si le risque est quantifiable, sa perception diffère selon les opérateurs (Slovic, 2000 ; Siegrist & Árvai, 2020). Une perception du risque sous ou surévaluée est le reflet de biais cognitif, et d'un modèle mental (i.e. la conscience de la situation) inadapté à la

situation, ce qui amène à des décisions non pertinentes (Orasanu, Dismukes & Fischer, 1993 ; Williams & Noye, 2004 ; Blaauwgeers, Dubois & Ryckaert, 2013). La perception du risque est, selon Chancey et al. (2017) et Parasuraman et Riley (1997), un modérateur entre la confiance et l'acceptation. Hoesterey et Onnasch (2022) ont observé qu'une perception élevée du risque réduit le comportement de confiance envers le CPS. Cependant, la perception du risque reste peu étudiée dans la CHM (Atchley et al., 2023).

La compréhension d'une situation, intégrant la conscience de la situation et perception du risque associé à cette situation, est un construit cognitif se basant sur l'acquisition et le traitement de données. Chez l'humain, cette capacité de traitement est limitée (Miller, 1956 ; Baddeley & Logie, 1999). Dépasser cette limite risque de conduire à un état de surcharge mentale.

2.2.3. La charge mentale : un déterminant à réguler

Il n'existe pas de définition de la charge mentale permettant un consensus au sein de la communauté scientifique (Young, et al., 2015). Il est cependant possible de considérer la charge mentale comme une construction hypothétique stipulant qu'un opérateur ressent une charge issue du coût mental nécessaire pour atteindre un certain niveau de performance (Hart & Staveland, 1988). Pour Hart et Staveland (1988), comme pour Meshkati (1988), la charge mentale résulte de l'interaction entre les caractéristiques de l'opérateur, les caractéristiques de la tâche et des circonstances dans lesquelles se déroule la tâche. La norme ISO 10075 décrit les risques liés à la charge mentale : « *Une contrainte mentale dépassant les capacités humaines de traitement de l'information conduira directement à des erreurs dans les performances humaines. Dans de tels cas, la contrainte mentale atteint un niveau rendant en principe impossible l'accomplissement de la tâche requise. De telles exigences dysfonctionnelles doivent donc être évitées par une conception adéquate des tâches, des outils et des interfaces homme-machine/ordinateur.* ». La charge mentale représente un risque lorsqu'elle est trop élevée.

Lors d'une CHM, cette charge mentale va avoir un impact sur la performance (Roundtree, Goodrich & Adams, 2019) et sur le succès de cette interaction (Prewett, Johnson, Saboe, Elliott & Covert, 2010 ; Ötting, Masjutin, Steil & Maier, 2020). Il semble exister de

nombreux moyens pour permettre de gérer cette charge cognitive (visualisation, type d'interface, niveau d'automatisation) (Prewett et al., 2010), mais les résultats ne sont pas assez constants pour permettre de proposer des recommandations dans la CHM.

Hudon, Demazure, Karran, Léger et Sénécal (2021) ont mis en évidence un lien entre la charge mentale et la confiance dans une intelligence artificielle. Les modes de présentation qui provoquaient chez l'opérateur moins de charge mentale étaient corrélés avec les mesures de confiance les plus élevées. Karpinsky, Chancey, Palmer et Yamani (2018) ont observé un résultat similaire. Dans les conditions où la charge mentale était plus élevée, la confiance ressentie était plus faible. Cependant, une réplication de cette étude par Sato, Yamani, Liechty et Chancey (2020) n'a pas permis de retrouver cette corrélation. L'explication de ce résultat divergeant serait dû à une absence de variation de charge mentale, contrairement aux études précédentes.

2.3. Modèle de prise de décision assistée par un CPS

Les cas de coopération investigués dans cette thèse portent sur des prises de décision (accepter ou refuser) assistée par un CPS (cobot ou IA). Cette assistance prend la forme d'une proposition de transfert de ressources pour préparer des commandes ou de changement de planning de maintenance. Cette question de la prise de décision est une facette précise de la coopération. La prise de décision est un vaste champ de recherche qui a été étudié aussi bien en philosophie (Szaniawski, 1980), en management (Al-Tarawneh, 2011) ou en informatique (Moore & Whinston, 1986). Une simple recherche de ce terme dans le moteur de recherche « Thèse.fr » amène à 24164 réponses tous domaines confondus (au 14/06/2023). Rien que dans le champ de la psychologie cette question est traitée depuis plus de 70 ans, les premiers articles remontant aux années 50 (Ewards, 1954). Cependant, la définition même de la prise de décision au travers ces différents champs, et malgré les époques, reste sensiblement identique puisqu'elle est définie comme le fait de choisir une solution afin de répondre à un problème. Efendic (2017) définit de façon plus précise la prise de décision comme « *how people (including other organisms and machines) combine desires (utilities, personal values, goals, ends, etc.) and beliefs (expectations, knowledge, means etc.), to choose a certain course of action* » (page 23). Nous retrouvons dans cette définition les dimensions présentées dans la section 2.2 (« Les déterminants de l'opérateur dans la prise de décision assistée par un

CPS ») avec notamment les « beliefs » (ou croyances) qui font référence aussi bien à la confiance dans le CPS qu'à la conscience de la situation. Si les modèles de la coopération présentés dans la partie 2.1.2 prennent en compte la prise de décision, elle n'est pas leur finalité. En effet, ces modèles peuvent également être appliqués pour comprendre l'exécution d'une tâche. Ajoutons que les modèles de coopération auxquels nous avons fait référence ne permettent pas de comprendre les processus sous-jacents à la prise de décision de l'opérateur lorsqu'il reçoit une proposition d'un CPS. En revanche, le modèle proposé par Chancey et al. (2017) a pour finalité la compréhension de la prise de décision pendant une CHM. De plus, ce modèle met l'accent sur la dimension de la confiance et sur la perception du risque (Figure 2.9).

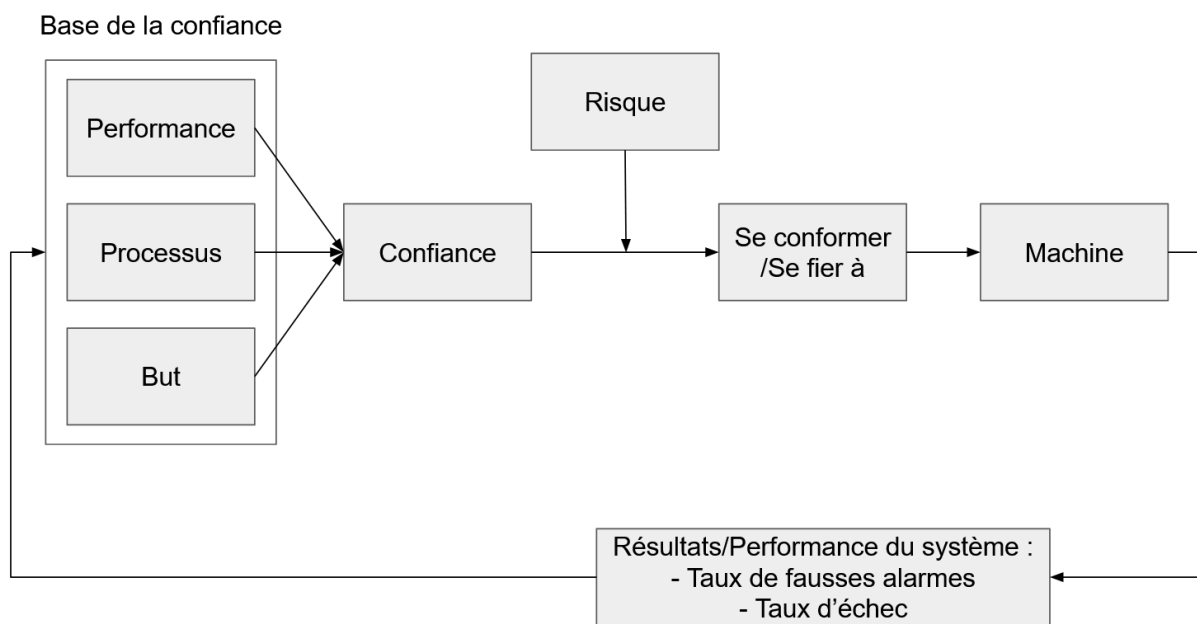


Figure 2.9 : Modèle de la confiance humain-machine sur les comportements de compliance/reliance selon Lee et See (2004) et Mayer et al. (1995) (traduit et adapté de Chancey et al., 2017)

Le modèle proposé par Chancey et al. (2017) se base sur la définition de la confiance de Lee et See (2004), présentée précédemment, et sur le paradigme « **compliance-reliance** ». La « compliance » (« **accepter** ») et la « reliance » (« **se fier à** ») sont des manifestations comportementales de la confiance (Lee & See, 2004). L'acceptation correspond au moment où un opérateur répond favorablement à un signal émanant d'un CPS. L'acceptation se dégrade lorsque le CPS envoie une alerte alors que la situation est normale (i.e. **fausses alarmes**). L'action de « se fier à » correspond au moment où un opérateur n'opère pas de réponse et n'engage pas d'opération spécifique face à un CPS silencieux ou dans un état

« normal ». L'action de se fier se dégrade lorsque le CPS n'envoie pas de signal alors que la situation se dégrade (i.e. **échec**). Cet effet des fausses alarmes et des échecs est inclus par Hoff et Bashir (2015) dans la Confiance Apprise Dynamique (CAD). Chancey et al. (2017) utilisent plus particulièrement la notion de confiance dans le signal. C'est entre la confiance et sa manifestation comportementale (« accepter » ou « se fier à ») que le risque vient être modérateur (Fig. 2.9). La définition du risque privilégiée par Chancey et ses collègues est celle de Sitkin et Pablo (1992) « *the extent to which there is uncertainty about whether potentially significant and/or disappointing outcomes of decisions will be realized* » (p. 10). Le risque est lié à l'incertitude et à la probabilité que les résultats d'une prise de décision soient négatifs. Plus l'incertitude est élevée et/ou plus la probabilité de résultats négatifs est forte, plus le risque est élevé. Mais comme le soulignent Mayer et al. (1995), ce n'est pas la confiance qui représente un risque, mais le comportement. L'attitude en soit (i.e. **avoir** confiance) n'implique pas de conséquences. En revanche, le comportement de se rendre vulnérable aux potentiels échecs de l'autre (i.e. **faire** confiance) engendre une décision et des conséquences. Le risque tel que défini ici se rapproche du niveau 3 de la conscience de situation. Pour rappel, ce niveau intègre la connaissance de la probabilité, et l'incertitude qui en découle, des résultats attendus après un événement. Le modèle de Chancey et al. (2017) permet de mettre en avant le côté itératif et cyclique de la CHM. La confiance de l'opérateur est le résultat des performances de la machine, qui eux-mêmes sont liés à la prise de décision de l'opérateur (i.e. le comportement de confiance).

L'objectif de Chancey et al. (ibid) était de comprendre le rôle de la confiance dans le paradigme de la « compliance/reliance » en fonction des fausses alarmes, du taux d'échec et du risque perçu. Les résultats des expérimentations de Chancey et al. (ibid) n'ont cependant pas permis de valider entièrement leurs hypothèses. En effet, si l'effet du risque, et de sa perception, vont dans le sens attendu, il n'est pas significatif. La confiance, elle, joue un rôle significatif dans l'acceptation d'une proposition.

Dans le cadre de notre problématique, c'est l'acceptation qui nous intéresse. Comme relaté au début de ce chapitre, nos cas d'étude portent sur un opérateur qui est en situation d'acceptation d'une proposition (i.e. un signal provenant d'un CPS). Les résultats de Chancey et al. (2017) et de Hoesterey et Onnasch (2022) sont des premiers éléments permettant de

comprendre le rôle de la perception du risque sur le comportement de confiance. Toutefois, ces résultats nécessitent encore d'être approfondis. Il est important de comprendre le rôle des déterminants que sont la confiance et la compréhension de la situation (conscience de la situation et risque perçu) dans le comportement de confiance. La charge mentale, comme expliqué plus en amont, a déjà des effets bien compris sur la performance, mais il est nécessaire de toujours la contrôler pour s'assurer qu'elle ne soit pas à un niveau trop élevé.

2.4. Synthèse du chapitre 2

Ce chapitre a permis de montrer que la CHM repose sur une dynamique complexe entre un humain et un CPS. Ces situations de coopération sont caractérisées par des interférences que les agents doivent gérer. Ces situations de coopération sont régies par plusieurs déterminants de l'opérateur qui améliorent ou détériorent (au sens de la performance) la coopération. Le cadre de cette thèse porte sur une facette particulière de la coopération qui est la prise de décision assistée par un CPS. En reprenant la définition de la confiance de Lee et See (2004) nous pouvons voir que dans les deux projets, l'opérateur est dans une situation d'incertitude et de vulnérabilité. Dans le projet Humanism, l'opérateur est vulnérable face aux problèmes de ressources intervenant sur la ligne et il peut utiliser l'aide du CPS pour pouvoir les résoudre. Dans le cadre du projet Seanatic, l'opérateur est vulnérable face à la gestion d'interférence sur le calendrier de la maintenance (les processus) et sur les probabilités de casse des équipements (les ressources). Le CPS lui permet alors d'optimiser le calendrier de maintenance et d'éviter la survenue d'événements problématiques. Dans les deux cas d'étude, le CPS vient aider avec des informations et une proposition.

Afin de guider l'ensemble des travaux menés dans cette thèse, nous proposons le modèle suivant sur la base des travaux de Hoff et Bashir (2015) et de Chancey et al. (2017) (Figure 2.10). Ce modèle fait figurer le rôle de la confiance dans le cadre d'une prise de décision assistée par CPS. La confiance de l'opérateur envers le CPS joue un rôle central. La **confiance** (i.e. l'attitude) préfigure l'**acceptation** (i.e. le comportement). Nous reprenons la catégorisation de la confiance en trois composantes proposées par Hoff et Bashir (2015). Ces trois composantes sont la **confiance dispositionnelle** (issue des facteurs individuels), la **confiance apprise (initiale)** provenant des attitudes et attentes ; **dynamique** provenant de la performance de l'agent) et la **confiance situationnelle** (provenant de la charge mentale, de la

conscience de la situation et de la perception du risque). Une différence notable avec le modèle proposé par Hoff et Bashir (ibid) est l'ajout de la confiance situationnelle comme modératrice du lien entre la confiance apprise et l'acceptation.

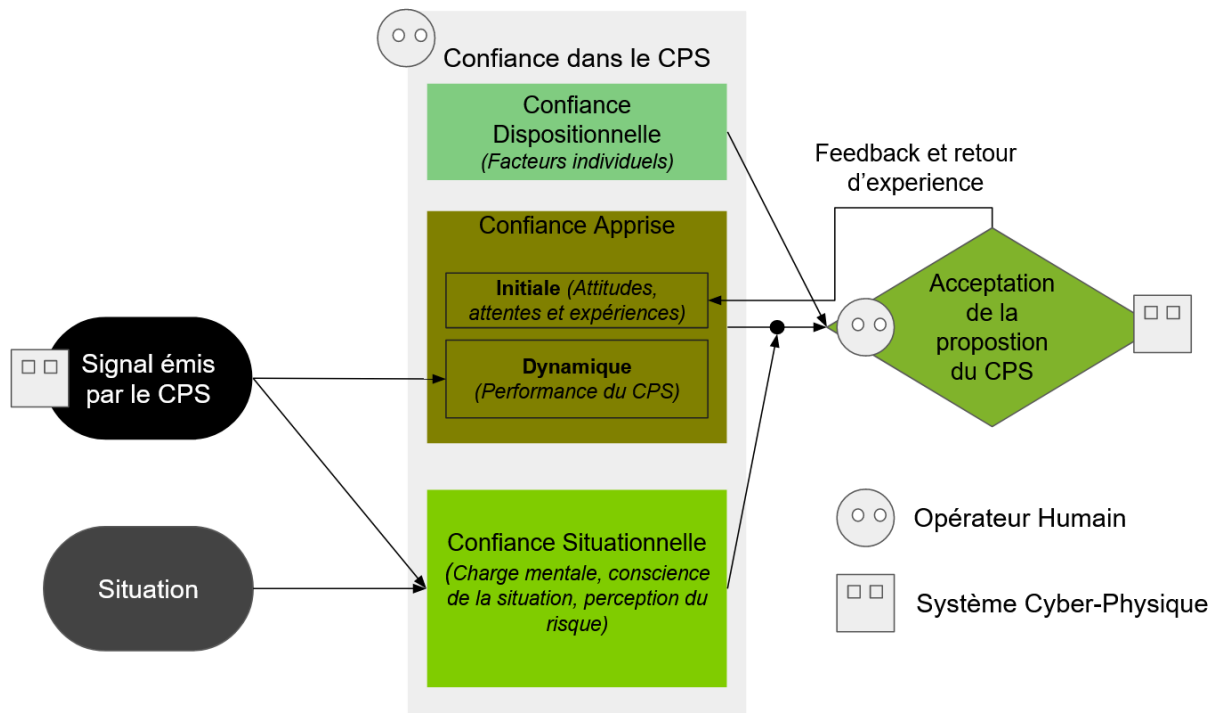


Figure 2.10 : Adaptation du modèle de Chancey et al. (2017) selon les composantes de la confiance (Hoff & Bashir, 2015)

En effet, selon Chancey et al. (2017), le risque vient modérer le lien entre la confiance et l'acceptation. Au début d'une situation de coopération, le CPS émet un signal (i.e. une proposition) qui vient modifier la confiance apprise dynamique. C'est par la teneur des informations et de ce que l'opérateur en comprend, en termes de performance de l'agent (sa fiabilité, ses compétences ou sa prévisibilité par exemple), que le signal influence la confiance apprise dynamique. Lorsque le CPS communique sur la situation alors le signal émis modifie également la confiance situationnelle (en communiquant sur les dangers existants par exemple). La façon dont ce signal est présenté vient également impacter la confiance situationnelle par son effet sur la charge mentale. Enfin, la situation en elle-même vient modifier la confiance situationnelle selon que cette situation représente un danger ou qu'elle requiert de l'investissement cognitif de la part de l'opérateur. La confiance apprise dans le CPS est modérée par cette confiance situationnelle et amène l'opérateur à accepter, ou refuser, la proposition du CPS. Une fois la situation de coopération terminée, l'opérateur

intègre ses retours d'expérience, et feedbacks éventuels, pour faire évoluer sa confiance apprise initiale. Celle-ci peut augmenter si la situation de coopération a réussi, et diminuer si la situation de coopération a échoué ou si elle s'est mal passée.

Nous cherchons désormais à comprendre quels éléments dans la CHM peuvent-être utilisés pour calibrer, maintenir et réguler la confiance apprise et la confiance situationnelle de l'opérateur dans sa prise de décision assistée par un CPS, et in fine, l'acceptation de la proposition. La proposition émise est l'élément central car il influence directement la CAD et la CS, et indirectement la CAI. Le concept de transparence est la capacité du CPS à transmettre des informations pour être compréhensible et plusieurs auteurs l'ont identifié comme élément susceptible d'aider à la calibration de la confiance (Hoff & Bashir, 2015 ; de Visser et al., 2020 ; Naiseh, Al-Thani, Jiang & Ali, 2021). Cette transparence est également mise en lien avec la conscience de la situation d'un opérateur sans impliquer de hausse de la charge mentale (van de Merwe et al., 2022). Plus généralement, la transparence est considérée comme un élément primordial à prendre en compte pour l'amélioration de la CHM (Chen, 2021), notamment dans l'industrie 4.0 (Guerin et al., 2019). Le prochain chapitre vise à expliciter le concept de transparence, son rôle et ses utilisations dans la CHM.

Chapitre 3 : Transparence, confiance et prise de décision assistée par un CPS

Ce chapitre a pour objectif de montrer comment la transparence peut venir aider pour une prise de décision assistée par un système cyber-physique (CPS). La transparence peut être définie comme la capacité d'un CPS à transmettre des informations à l'opérateur, et ce pour aider à la coopération. Dans le cadre de l'industrie 4.0, la transparence est considérée comme un critère de performance et permet d'améliorer la coopération (Hozdić. 2015 ; Lu, Morris & Frechette, 2016 ; Wang, Pynadath & Hill, 2016). De plus, la transparence des CPS fait partie des challenges majeurs pour la coopération humain-machine (CHM) (Correia & Lindley, 2022). La transparence représente également un intérêt car cette capacité va venir influencer les composantes de la confiance et aider à la prise de décision (Chen, 2021).

Pour ce faire nous commençons par expliciter le concept de transparence au travers trois modèles. Ces modèles sont prescriptifs et sont usuellement utilisés dans la littérature (Rajabiyazdi et Jamieson, 2020). Ces trois modèles permettent de définir la transparence d'un CPS selon des dimensions (Lyons, 2013) ou selon des niveaux (Chen et al., 2014 ; 2018). Ces modèles ne sont pas à opposer, et leur articulation permet de comprendre et de modéliser la transparence de façon plus fine.

Pour comprendre le lien entre la transparence, la confiance apprise et la confiance situationnelle (conscience de la situation, perception du risque et charge mentale) nous faisons une revue des différentes expérimentations existantes, avec un focus sur celles utilisant un paradigme de prise de décision assistée par un CPS. L'utilisation de la transparence permet, par la modification des compétences de communication du CPS (son « savoir coopérer » Pacaux-Lemoine, 2021) de maintenir et de modifier le COFOR. Par la revue des résultats existants nous montrons qu'il existe encore des challenges dans la compréhension des effets de la transparence sur les composantes de la confiance dans la prise de décision assistée par un CPS

La coopération entre un humain et un CPS ne se résume pas à une seule interaction, et les composantes de la confiance évoluent avec le temps (Hoff & Bashir, 2015). Nous revenons donc sur le principe d'interface adaptative, et en quoi la transparence peut être

intégrée dans ce courant de recherche. La transparence peut être un levier pour qu'un CPS avec des capacités de transparence adaptative puisse réagir en fonction d'événements. Ce champ encore récent permettrait d'aider à calibrer la confiance lorsque l'on estime qu'elle est trop élevée, où à l'inverse, trop faible.

Nous concluons ce chapitre en identifiant les liens entre les dimensions et les niveaux de la transparence et le modèle de la prise de décision formulé dans le second chapitre. Cette mise en perspective permet d'explicitier les éléments que nous investiguons.

3.1. La transparence comme facilitateur de coopération

Nous retrouvons le concept de transparence à travers de multiples champs de recherche tel qu'en physique (Wei et al., 2021), en science organisationnelle (Maas & Yin, 2021) ou en économie (Nurunnabi, 2021). De ces trois exemples nous constatons la double interprétation de la transparence. La première interprétation est au sens de « voir au travers » (« seeing through »). Ce « voir au travers » est la propriété d'un matériau le rendant translucide et permettant de voir ce qu'il y a derrière (exemple d'une vitre). La seconde interprétation est au sens de « voir dans » (« seeing into »). Ce « voir dans » est la capacité, par exemple d'une organisation, à communiquer sur ce qu'il se passe en interne pour que les éléments externes en comprennent le fonctionnement. Dans le cadre de la coopération avec un système cyber-physique (CPS), c'est la transparence comme une propriété permettant à l'opérateur de « voir dans » le CPS qui est à utiliser (Ososky, Sanders, Jentsch, Hancock & Chen, 2014). Pour Ososky et al. (ibid) « voir dans » le CPS, et donc le comprendre, permet le maintien et la correction du COFOR en vue d'une confiance calibrée.

Les concepteurs d'intelligence artificielle (IA) s'intéressent à la façon dont ils peuvent permettre à l'opérateur de « voir dans » l'IA afin d'améliorer la coopération. La problématique est de pouvoir faire comprendre aux opérateurs le fonctionnement de l'IA, son explicabilité et les éléments discriminants pris en compte par l'algorithme (Hudon et al., 2021). En d'autres termes, comment l'IA peut expliquer ses décisions afin d'être comprise par l'opérateur. Doit-elle détailler tout son algorithme, seulement une partie, ou bien être opaque ? Une récente étude a par exemple montré que l'IA ne devait pas expliquer tout son algorithme, sous peine de réduire l'efficacité de l'opérateur (Guesmia et al., 2021). Guesmia

et al. (ibid) observent qu'il est préférable d'avoir accès à une explication intermédiaire, plutôt qu'une explication sommaire ou à une explication complète. Cet entre-deux permet à l'opérateur de comprendre le fonctionnement de l'IA sans avoir à interpréter de nombreuses données ou l'algorithme dans son entièreté. Cependant, le courant XAI (eXplainable Artificial Intelligence) s'intéresse principalement à l'explicabilité de l'algorithme d'une IA en occultant les autres informations que pourrait communiquer l'IA (Meuwissen & Bollen, 2021). Selon Chen et al. (2014) la transparence d'un agent est « *the descriptive quality of an interface pertaining to its abilities to afford an operator's comprehension about an intelligent agent's intent, performance, future plans, and reasoning process* » (page 2). Il en va de même avec la définition proposée par Wortham et Théodorou (2017) pour qui la transparence est la capacité d'un CPS à rendre accessible à l'individu ses états internes et ses processus de prise de décision. Ces deux définitions ont comme point commun que la transparence peut être définie comme les informations transmises par un CPS afin d'améliorer la CHM par la création et le maintien du COFOR. La transparence peut de ce fait être considérée comme une compétence du « savoir-coopérer » (voir Figure 2.4 : **Synthèse** du KH et KHC (traduit de Millot & Pacaux-Lemoine, 2013)) car c'est la compétence de savoir communiquer avec autrui (l'une des aptitudes externes du KHC). Les informations communiquées par un CPS transparent ont pour objectif d'améliorer la conscience de la situation, de calibrer la confiance et par extension, d'améliorer la performance. Selon la méta-analyse de van de Merwe et al. (2022) la transparence a été essentiellement utilisée dans le contexte militaire (ex : Selkowitz, Lakhmani, Larios & Chen, 2016 ; Stowers et al., 2017 ; Chen et al., 2018). Mais le concept de transparence a également été utilisé pour la gestion d'une voiture autonome (ex : Helldin, Falkman, Riveiro & Davidsson, 2013 ; Pokam, Debernard, Chauvin & Langlois, 2019), dans le domaine civil (ex : Bhaskara, Skinner et Loft, 2021), dans l'aviation (ex : Göritzlehner et al., 2014 ; Sadler et al., 2016 ; Lyons et al., 2016) et de façon moindre dans la robotique (ex : Guznov et al., 2020) et le nucléaire (ex : Skraaning & Jamieson, 2021). Pour formaliser ce concept et aider à son utilisation plusieurs modèles ont été proposés.

3.1.1. Modèles de transparence

La revue de Rajabiyazdi et Jamieson (2020) met en avant trois modèles prescriptifs. Le modèle du Situation Awareness-based agent Transparency (SAT) (Chen et al. 2014) et son

évolution le dynamic SAT model (Chen et al., 2018) sont les modèles les plus cités (473 citations cumulées, respectivement 228 et 245 citations²). Le troisième est le « Human-Robot Transparency model » proposé par Lyons (2013), cité 267³ fois. Pour Rajabiyazdi et Jamieson (2020) ces modèles sont prescriptifs, dans le sens où ils définissent un panel d'informations, sous forme de dimensions ou de niveaux, à communiquer à l'opérateur. Cependant, la littérature a montré qu'il manquait d'études empiriques sur les effets des dimensions de la transparence proposées par Lyons (Rajabiyazdi & Jamieson, 2020) et des inconsistances de résultats avec l'utilisation des modèles de Chen et al. (Bhaskara et al., 2020 ; Rajabiyazdi & Jamieson, 2020).

3.1.1.1. Human-robot transparency model

Le premier modèle proposé est celui de Lyons en 2013. Ce modèle a pour objectif de proposer un guide permettant d'améliorer la CHM. Pour Lyons, l'utilisation de ce modèle permet d'améliorer la compréhension du CPS par l'opérateur, mais également les informations que le CPS acquiert sur l'humain. Ce modèle scinde la transparence d'un CPS selon deux capacités : sa capacité à communiquer sur ses états et sa capacité à communiquer sur les informations acquises sur l'humain et sur l'équipe. La première capacité de transparence est « **robot-TO-human** » (rTOh) et correspond aux informations que le CPS va transmettre sur lui-même et sur son environnement. Cette transparence permet au CPS de communiquer sur ses buts, sur la façon dont il les atteint, ses raisonnements et sur l'environnement qui contraint ses actions. La seconde capacité de transparence est « **robot-OF-human** » (rOFh) et correspond aux informations que le CPS a acquis sur l'opérateur et sur l'équipe. Si l'opérateur doit comprendre comment fonctionne le CPS, l'inverse est vrai, et le CPS doit pouvoir acquérir des informations sur l'opérateur pour s'adapter en conséquence. Cette capacité de transparence comprend également la faculté de communiquer sur la répartition des tâches entre l'opérateur et le CPS. De ce fait, ces informations peuvent se rapprocher du méta-niveau du COFOR, selon le modèle de CHM de Hoc (2001). Un CPS avec cette capacité permet à l'opérateur de se figurer les informations dont dispose le CPS sur les

² Selon le moteur google scholar au 14/06/2023

³ Selon le moteur google scholar au 14/06/2023

opérateurs. Ces deux capacités de transparence permettent au CPS de communiquer sur plusieurs dimensions que Lyons nomme « modèles » (Figure 3.1).

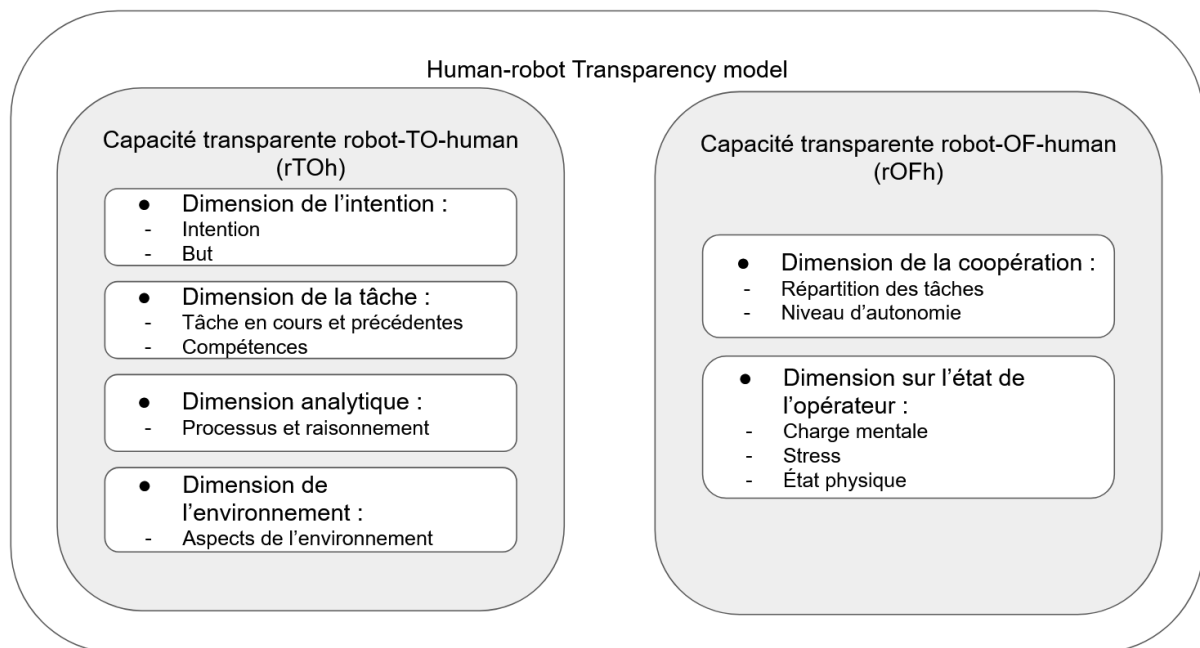


Figure 3.1 : Human-robot transparency model (Lyons, 2013)

Lyons définit quatre dimensions pour la capacité de transparence rTOh :

- **La dimension de l'intention.** Un CPS transparent sur les informations relatives à ses intentions permet de rendre compte de son but, et de ce pourquoi il a été conçu. La première façon de rendre tangible ce but est de passer par l'aspect physique du CPS. De la même façon que l'affordance permet de comprendre que l'on peut s'asseoir sur une chaise, un CPS transparent sur son intention permet à l'opérateur de savoir comment l'utiliser et dans quel but. En plus du « pourquoi » le CPS a été créé, les informations contenues dans la dimension de l'intention servent également à transmettre à l'opérateur le « comment » il va réaliser ses objectifs. Lyons précise que ces informations ne rentrent pas dans le détail, tâche par tâche, mais plutôt sur les lignes directrices des actions du CPS. Il dresse un parallèle avec les lois de la robotique d'Asimov (Asimov, 1950). Ces lois permettent de comprendre la philosophie globale du CPS, sans rentrer dans le détail des tâches. Ici un CPS préparateur de pizza aurait l'apparence d'un équipier de cuisine.

- **La dimension de la tâche.** Un CPS transparent sur cette dimension permet à l'opérateur de comprendre les actions que va effectuer le CPS. Contrairement à la dimension de l'intention, le CPS va exprimer tâche par tâche ce qu'il fait exactement, et dans quel but. Le CPS sera transparent sur l'objectif final, par exemple réaliser une pizza, et pourra expliciter tâche par tâche ce qu'il va faire (i.e. faire la pâte, ajouter la sauce tomate, etc.). Les informations contenues dans cette dimension comprennent également la progression dans la tâche, et ce à un niveau de granularité plus ou moins fin. Ces informations permettent également au CPS d'être transparent sur la qualité de ses compétences dans un environnement donné. Enfin, le CPS peut communiquer sur ses potentielles erreurs, permettant à l'opérateur de calibrer sa confiance.
- **La dimension analytique.** Les informations contenues dans cette dimension sont assimilables au concept de XAI. Le CPS a la capacité d'utiliser des données plus conséquentes qu'un opérateur. C'est pourquoi il doit communiquer à l'opérateur les données qu'il a utilisées et le raisonnement qu'il a mis en œuvre. Être transparent sur ces informations permet à l'opérateur de comprendre ce que le CPS utilise et améliore le modèle mental qu'il se fait du CPS. Le CPS préparateur de pizza communiquera ici qu'il abaisse la cuisson du four car la pâte commence à brûler mais que le fromage n'est pas fondu.
- **La dimension de l'environnement.** Un CPS agit dans un environnement qu'il doit prendre en compte lorsqu'il réalise ses actions. Les cobots ont par exemple des systèmes de sécurité les bloquant ou les ralentissant lorsqu'un opérateur se trouve à proximité. Le CPS doit dès lors être transparent sur ces éléments d'environnement et communiquer à l'opérateur qu'il les a pris en compte. Le CPS peut également communiquer en quoi cet environnement impacte ses compétences. L'opérateur peut alors adapter son modèle mental pour calibrer sa confiance dans le CPS. Ce modèle comprend également les contraintes temporelles. Le CPS pourra communiquer qu'il ne lui reste que quelques minutes pour finir sa commande.

Lyons inclut deux dimensions sur lesquelles s'appuie la capacité de transparence rOFh :

- **La dimension de la coopération.** Le CPS va être transparent sur la répartition des tâches entre l'opérateur et lui et également sur son niveau d'autonomie

(Parasuraman, Sheridan & Wickens, 2000). L'idée est qu'en étant transparent sur les informations de la coopération, l'opérateur va améliorer son COFOR sur le niveau de la planification et l'aider dans la gestion des futures interférences. Le CPS pourra communiquer qu'il s'occupe du pétrissage de la pâte, mais que c'est à l'opérateur de faire la garniture.

- **La dimension sur l'état de l'opérateur.** De la même façon qu'un coéquipier humain peut le faire, le CPS doit être capable d'acquérir des informations sur les états émotionnels, cognitifs et physiques de l'opérateur. L'utilisation de capteurs et IOT permettent au CPS d'acquérir cette sentience d'autrui. En fonction de l'acquisition de ces éléments, le CPS pourra (demander à) modifier son niveau d'automatisation pour venir en aide à l'opérateur, et le laisser se concentrer sur d'autres tâches. Lorsque le CPS perçoit certains signes, il peut également alerter l'opérateur d'un danger et le mettre au courant de la situation. Le CPS pourra communiquer qu'il perçoit que l'opérateur a trop de pizza à réaliser en même temps, et lui proposer une solution comme prendre en charge une partie de ses commandes.

Le modèle de Lyons permet à l'opérateur de comprendre le CPS lorsqu'il réalise une tâche. Lorsque le CPS fait une proposition, ce modèle permet de comprendre le pourquoi de la proposition et les conséquences liées à cette proposition, aussi bien sur les autres opérateurs que sur l'environnement. La bidirectionnalité de ce modèle permet de rapprocher la coopération avec le CPS d'une coopération avec un coéquipier humain.

3.1.1.2. Situation awareness-based agent transparency

Les deux autres modèles prescriptifs sont le SAT proposé par Chen et al. (2014) et le dynamic SAT, qui vise à compléter le SAT (Chen et al., 2018). Ces modèles se basent sur le concept de conscience de la situation (SA) proposé par Endsley (1995) (voir 2.2.2). Pour les auteurs, une des façons d'améliorer la coopération est de faciliter la conscience de la situation de l'opérateur. Pour améliorer cette conscience de la situation, les auteurs ont intégré le

Belief, Desires, Intention⁴ (BDI) agent framework (Rao & Georgeff, 1995) et les 3P (Purpose, Process, Performance⁵) de Lee et See (2004) aux trois niveaux de SA (Figure 3.2).

Dans le premier niveau de ce modèle, le CPS est transparent sur les informations permettant à l'opérateur d'améliorer sa perception de la situation (SA niveau 1). Au niveau 1 le CPS communique sur trois types d'informations : ses objectifs, ses processus et sa performance. Pour ses objectifs, le CPS communique les informations sur les buts qu'il veut atteindre (ex : aider à la cuisine et faire des pizzas). Pour ses processus il communique sur les tâches qu'il réalise pour atteindre ses buts et ce qui lui reste à faire (ex : « Je pétris la pâte et après je la laisse reposer. »). Enfin pour sa performance il communique sur ses échecs ou ses réussites concernant la tâche actuelle (ex : « Statut de pétrissage de la pâte « OK ». »).

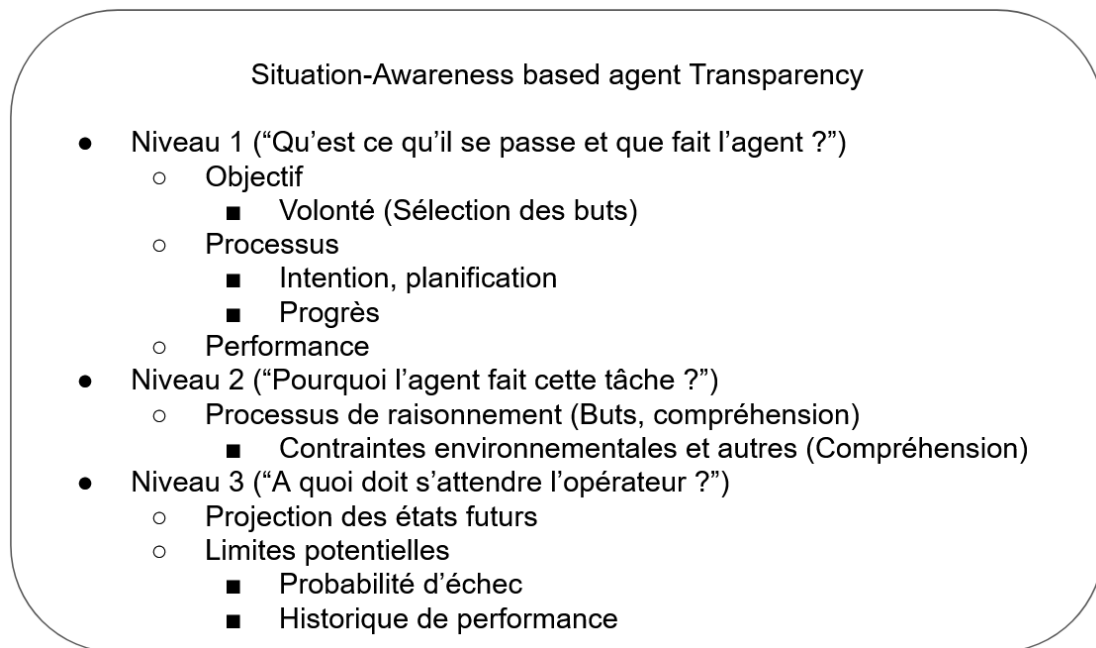


Figure 3.2 : Situation Awareness based-agent Transparency model (Chen et al., 2014)

Dans le second niveau le CPS est transparent sur les informations permettant à l'opérateur d'améliorer sa compréhension de la situation (SA niveau 2). Ici le CPS communique sur les informations relatives à ses processus de raisonnement. Nous retrouvons ici des informations comparables au XAI car le CPS sera transparent sur les données et les contraintes qu'il a pris en compte pour choisir et réaliser sa tâche. Ces contraintes peuvent-

⁴ Nous traduisons ici par Compréhension, Volonté et Intention

⁵ Nous traduisons ici par But, Processus, Performance

être environnementales (ex : « Je mets la pizza au four car il a atteint la bonne température. »), organisationnelle (ex : « Je change de commande car celle-ci n'est pas prioritaire. ») ou autre (ex : « Je prépare une pizza « reine » car la probabilité d'une commande de ce type est élevée. »).

Dans le troisième niveau le CPS est transparent sur les informations permettant à l'opérateur d'améliorer sa projection de la situation future (SA niveau 3). Dans ce troisième niveau le CPS communique sur deux types d'informations : la projection des états futurs et les limites potentielles. Pour la projection des états futurs cela passe par le statut de la situation suite à la fin de la tâche actuelle (ex : « La pizza sera cuite. »). Le CPS peut également communiquer ses limites et faire en sorte que l'opérateur soit conscient de la probabilité d'échec (ex : « Le taux de pizzas non conformes réalisées. »). Chen et al. (ibid) précisent que ce modèle n'est pas forcément cumulatif, et le CPS peut très bien n'être transparent que sur certaines informations selon le besoin de l'opérateur.

En 2018, Chen et al. ont proposé d'intégrer la perception de l'environnement du CPS et de ce que le CPS acquiert comme informations sur l'opérateur. Cet ajout se trouve au premier niveau du SAT (Figure 3.3 en rouge). Communiquer ces informations à l'opérateur lui permet d'améliorer sa conscience de la situation, et de savoir ce que le CPS perçoit.

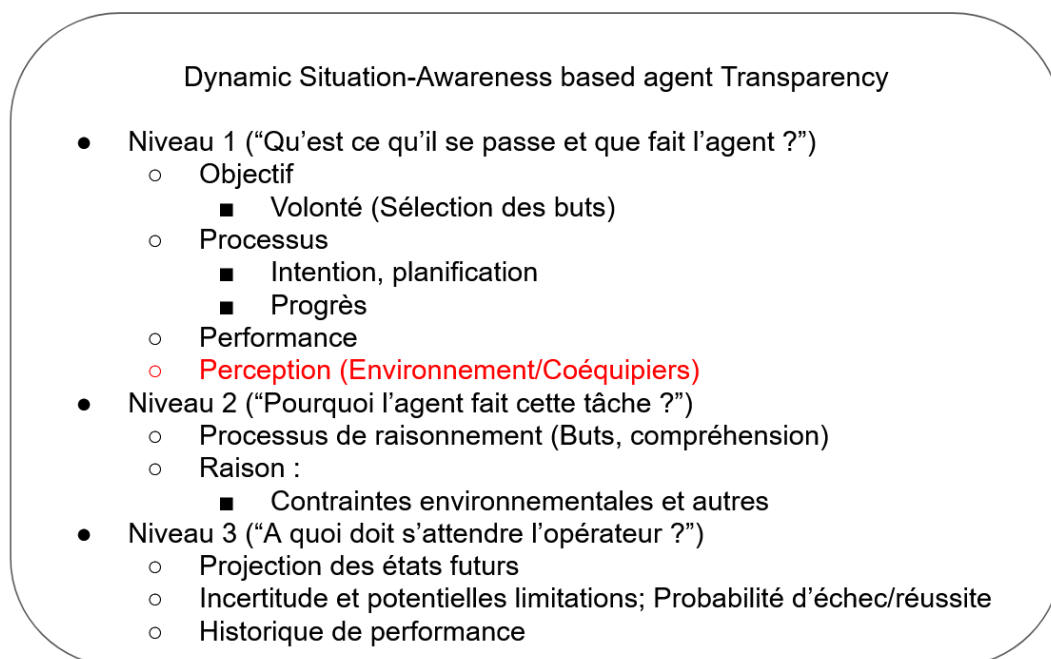


Figure 3.3 : Dynamic Situation Awareness based-agent Transparency model (Chen et al., 2018)

L'utilisation du SAT et du Dynamic SAT permet à l'opérateur d'avoir une meilleure compréhension du fonctionnement du CPS. Cette meilleure compréhension vient calibrer la confiance de l'opérateur, lui permettant de savoir s'il doit intervenir pendant la tâche du CPS ou si celui-ci fonctionne correctement. Lorsque le CPS fait des propositions, ce modèle permet à l'opérateur de comprendre la proposition et de pouvoir prendre sa décision avec une bonne conscience de la situation.

3.1.1.3. Articulation des modèles de transparence

Les différents modèles présentés permettent de définir la transparence selon des dimensions ou des niveaux d'informations communiquées par le CPS. Afin de mieux spécifier les informations communiquées par le CPS, nous proposons d'utiliser conjointement les deux modèles (Simon, Guérin, Rauffet, Chauvin & Martin, 2023a) (Figure 3.4).

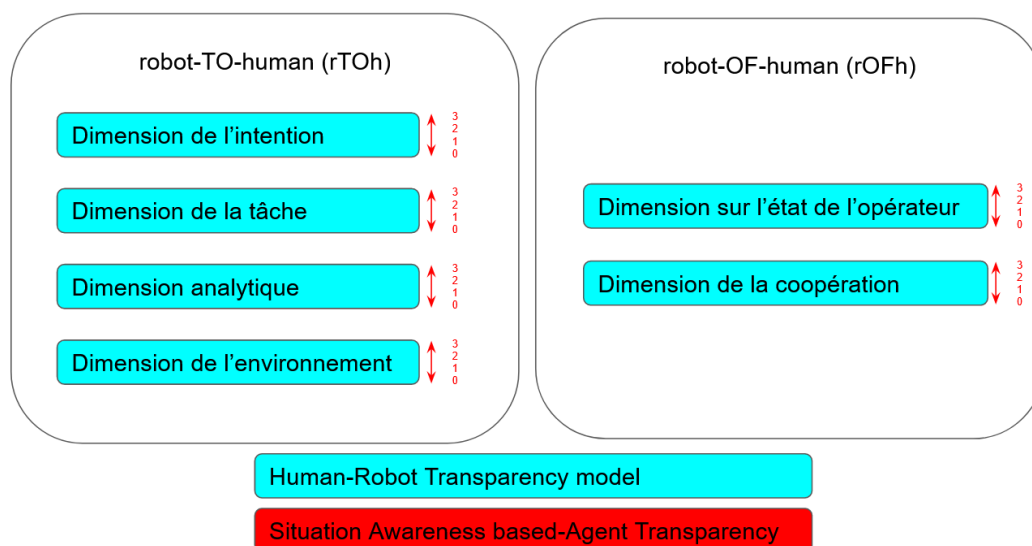


Figure 3.4 : Articulation des modèles de Lyons (2013) et Chen et al. (2018) (adapté)

L'utilisation conjointe des deux modèles passe par l'application des niveaux proposés par Chen et al. (2014 ; 2018) sur les dimensions de Lyons (2013). Chaque dimension peut, selon son niveau, apporter des informations à l'opérateur pour améliorer sa conscience de la situation. Notons qu'un nouveau niveau, le niveau 0, est ajouté par rapport à la proposition de Chen et al. (ibid). Ce niveau correspond à une absence d'information sur cette dimension. L'articulation de ces deux modèles, permet d'avoir une compréhension plus fine des effets provoqués par les informations transmises par un CPS transparent. Par exemple, un CPS avec une transparence SAT de niveau 0 ne transmettra aucune information. Un CPS qui est

transparent sur ses intentions à un niveau 1 du SAT n'indiquera que son intention (exemple : « Je sers à fabriquer des pizzas. »), à un niveau 2 le CPS alimentera son intention de ses processus (exemple : « Je pétris la pâte et mets la pizza au four. ») et au niveau 3 le CPS communiquera sur l'effet de son intention (exemple : « Avec mon aide le nombre de pizza fabriquée augmentera. »). En reprenant le cas d'un CPS pour la fabrication de pizza, il est possible de décliner cet exemple sur les différents niveaux et pour chaque dimension (Tableau 3.1).

Tableau 3.1 : Dimensions de transparence selon le niveau de transparence appliqué à un CPS d'aide à la fabrication de pizza

Dimensions de la transparence (Lyons, 2013)	Niveau de transparence (Chen et al., 2018)	Exemple
Intention	1	« J'aide à fabriquer des pizzas. »
	2	« Je viens en aide car la demande en pizzas a augmenté. »
	3	« Avec mon aide le nombre de pizza fabriquée augmentera. »
Tâche	1	« Je pétris la pâte. »
	2	« Suite à l'intégration des ingrédients je pétris la pâte. »
	3	« Après avoir pétri la pâte je la laisserais reposer. »
Analytique	1	« Je prépare plus de pizza pour anticiper une hausse des commandes. »
	2	« Les données des jours précédents et d'aujourd'hui me font percevoir une probabilité de hausse de commande. »
	3	« La probabilité que les commandes augmentent à 19h est de 95%. »
Environnement	1	« Je ne peux pas mettre d'autres pizzas à chauffer. »
	2	« Les données du four indiquent que le four est en surchauffe et m'empêche de mettre une nouvelle pizza à chauffer. »
	3	« Le four risque de prendre feu. »
Coopération	1	« Je m'occupe de pétrir la pâte et l'opérateur de mettre les ingrédients. »
	2	« L'opérateur est plus précis que moi pour mettre les ingrédients. »
	3	« Cette répartition des tâches permettra de répondre aux commandes. »
Opérateur	1	« L'opérateur est surchargé. »
	2	« La hausse des commandes fait que l'opérateur est surchargé. »
	3	« L'opérateur retrouvera une charge de travail normale dans dix minutes. »

Cette déclinaison des dimensions de la capacité de transparence sur plusieurs niveaux permet d'être plus précis sur la compréhension des informations communiquées par le CPS. L'utilisation de cette typologie permet de modéliser avec finesse la transparence du CPS. Nous proposons d'appeler le croisement entre une dimension et un niveau « **modalité de transparence** ». Cette proposition est une réponse au constat de van de Merwe et al. (2022) sur le manque d'une compréhension fine des effets des informations transmises par le CPS.

Les modèles présentés ont également été mis en application pour observer les effets de la transparence sur la CHM, et notamment sur les déterminants de l'opérateur présentés dans le chapitre 2 (confiance, conscience de la situation, perception du risque et charge mentale).

3.1.2. Transparence et composantes de la confiance

Il a été observé une forte disparité dans l'utilisation et la formalisation du concept de transparence à travers les études (van de Merwe et al., 2022). Selon les études, les auteurs utilisent de façon stricte le SAT (Mercado et al., 2016 ; Stowers et al., 2020 ; Bhaskara et al., 2021) ou le modèle de Lyons (Pokam, Debernard, Chavin & Langlois, 2019) et parfois utilisent le concept de façon plus vague telle que « low/high transparency » (Kizilcec, 2016 ; Roth et al., 2020). Nous voyons ici la complexité qu'il peut y avoir à tirer des principes généraux sur les effets de la transparence. Notons cependant que plusieurs travaux montrent que l'utilisation d'une interface transparente augmente la perception d'utilisabilité (Patel, Ramaswamy, Li & Pinciroli, 2021 ; Pokam et al., 2019 ; Mercado et al., 2016) et d'utilité (Lyons et al., 2016). La transparence est également corrélée avec la performance (Bhaskara et al., 2020 ; van de Merwe et al., 2022 ; Chen, 2021). Ce lien entre la transparence et la performance est explicable par l'effet de la transparence sur la confiance apprise et la confiance situationnelle (cf. section 2.2 « Les déterminants de l'opérateur dans la prise de décision assistée par un CPS »).

Transparence, confiance apprise et acceptation. Selon Wortham (2020), la transparence d'un CPS est un élément essentiel à la construction de la confiance. La transparence aide à l'élaboration d'un modèle mental du CPS en accord avec la réalité, et donc de calibrer la confiance. Ce constat est étayé par les résultats empiriques à disposition (Bhaskara et al.,

2020 ; Chen, 2021). Lorsque l'on compare des CPS transparents à des CPS non-transparentes, ce sont les CPS transparents qui obtiennent les plus hauts scores de confiance (Helldin, 2014 ; Mercado et al., 2016 ; Sadler et al., 2016 ; Lyons et al., 2016 ; Selkowitz et al., 2016 ; Patel et al., 2021 ; Liu, 2021). Il est intéressant de noter qu'en croisant la transparence d'un CPS avec sa fiabilité, Kluy et Roesler (2021) observent un effet d'interaction. La transparence sur un agent moins fiable amène à moins de confiance. Ce résultat indique que la confiance permet une calibration (i.e. que la confiance envers le CPS correspond à ses capacités réelles). Nous retrouvons cet effet de calibration chez Kizilcec (2016) où la transparence permet de réajuster la confiance des participants ayant perçu une violation de leurs attentes (i.e. le jugement de l'IA ne correspond pas à ce qu'ils attendaient). La transparence sur les explications modère cette baisse de la confiance. Par exemple Yang, Unhelkar, Li et Shah (2017) montrent que, suite à la violation de l'attente du participant, l'explicitation du raisonnement permet de calibrer la confiance en évitant un effet de « l'enfant qui criait au loup » (i.e. beaucoup d'alarmes sonnantes sans raison). La transparence vient aider l'opérateur à reconnaître les alarmes provenant d'une vraie menace.

Dans sa méta-analyse, van de Merwe et al. (2022) identifie six études dont le protocole implique d'accepter ou refuser une proposition. Dans ces expérimentations nous voyons que plus un agent communique d'informations (i.e. plus il est transparent), plus la confiance est élevée (Mercado et al., 2016 ; Sadler et al., 2016 ; Stowers et al., 2020). Roth et al. (2020) observent cette même tendance mais elle n'est pas significative. Bhaskara et al. (2021) observent eux un effet de la transparence sur la calibration de la confiance. Dans ces expérimentations nous observons également un effet de la transparence sur l'acceptation. Göritzlehner et al. (2014) observent par exemple un effet d'interaction entre la transparence et la complexité de la situation sur le fait d'accepter la proposition. Une transparence plus élevée dans une situation complexe amenait à plus d'acceptation. Plus la transparence est élevée, plus l'acceptation sera adéquate (i.e. savoir accepter et refuser les propositions au bon moment) (Mercado et al., 2016 ; Sadler et al., 2016 ; Stowers et al., 2020). Nous constatons ici que plus la transparence est élevée, plus la confiance est calibrée.

Transparence et confiance situationnelle (conscience de la situation). Les modèles proposés par Chen et al. (2014 ; 2018) accordent une place centrale à la conscience de la situation. De

ce fait, de nombreuses études ont mesuré cette conscience de la situation pour déterminer si la transparence permettait de l'améliorer. Plusieurs études observent qu'un agent transparent améliorera la conscience de la situation (Chen et al., 2014 ; Chen, Campbell, Gonzales & Copin, 2015 ; Selkowitz et al., 2016 ; Selkowitz, Larios, Lakhmani & Chen, 2017). Notons que l'amélioration de la conscience de la situation sera plus forte lorsque l'on passe d'un CPS non-transparent à un CPS transparent. L'amélioration est moins forte lorsque l'on compare un agent étant transparent à un niveau 2 du SAT et un agent transparent à un niveau 3 du SAT (Selkowitz et al., 2016 ; Wright et al., 2017). Cependant les résultats d'autres études ne montrent pas d'effet de la transparence sur la conscience de la situation (Pokam et al., 2019 ; Wright, Chen, Barnes & Hancock, 2020 ; Skraaning & Jamieson, 2021). Ces différences seraient dû à la disparité d'emploi de la transparence et à la façon dont la conscience de la situation est mesurée (van de Merwe et al., 2022). Dans les six articles identifiés par van de Merwe et al. (2022), seul celui de Roth et al. (2020) mesure la conscience de la situation. Leurs résultats montrent un lien positif entre une transparence plus élevée et une meilleure conscience de la situation.

Il est intéressant de noter qu'aucune étude, à notre connaissance, n'a exploré le lien entre la transparence et la perception du risque. Par exemple, Loft et al., (2021) ont manipulé le risque dans leur expérimentation (fort ou faible) mais n'ont pas mesuré la perception du risque des participants.

Transparence et confiance situationnelle (charge mentale). Les travaux sur la psychologie cognitive ont montré un lien entre le nombre d'informations présenté et la charge mentale. La hausse du nombre d'informations entraînant une hausse de la charge mentale (Baddeley & Logie, 1999). De ce constat, il était attendu à ce qu'un CPS plus transparent soit corrélé avec une augmentation de la charge mentale. Cependant les résultats des différentes études divergent. Plusieurs expérimentations n'ont pas observé de lien entre la transparence et la charge mentale (Selkowitz et al. 2017 ; Wright et al., 2020). Chen et al. (2015) observent eux un effet de la transparence sur la charge mentale. Ici, la transparence vient réduire la charge mentale. Kunze et al. (2019), à l'inverse, observent qu'un CPS transparent sur son incertitude augmente la charge mentale. Les résultats semblent encore diverger sur le lien entre transparence et charge mentale (Chen, 2021), même si la tendance est à considérer qu'une

augmentation de la transparence n'entraîne pas de charge mentale (Bhaskara et al., 2020 ; van de Merwe et al., 2022). Cinq des six expérimentations utilisant un paradigme où les participants devaient accepter ou refuser une proposition du CPS ont mesuré la charge mentale. Sur ces cinq études, aucune n'a identifié d'effet de la transparence sur la charge mentale (Görizlehner et al., 2014 ; Mercado et al. 2016 ; Roth et al., 2020 ; Stowers et al., 2020 ; Bhaskara et al., 2021).

Comme vu dans le modèle de Lyons, le CPS doit être capable de s'adapter à l'opérateur en fonction de ses états. Pour le SAT et le dynamic SAT, le CPS peut communiquer que sur certains niveaux. Selon ce modèle, le CPS est donc en capacité de s'adapter en fonction de l'individu ou de la situation. Proposer des CPS avec une capacité de transparence adaptative pourrait être un outil pour aider à calibrer la confiance d'un opérateur (Matthews et al., 2020).

3.2. Vers une transparence adaptative

L'idée d'adapter les interfaces aux opérateurs ou l'environnement n'est pas nouvelle. Les premières propositions remontent aux années 80 (Innocent, 1982). Greenberg et Witter (1985) montrent très tôt l'intérêt d'utiliser une interface qui s'adapte à l'opérateur par rapport à l'utilisation d'interfaces « fixes ». Plus globalement, l'utilisation d'interface adaptative permet de répondre à différentes problématiques telles que la différence de contexte d'utilisation (via des contraintes externes comme la luminosité ou le nombre d'informations) ou la différence entre les opérateurs (experts ou novice par exemple) (Brown, Norman & Riches, 1990 ; Thévenin, 2001). Pour Calvary et Coutaz (2002) une interface adaptative (qu'ils nomment « plastique ») est : « *Par analogie avec la plasticité des matériaux qui, sans rompre, s'étirent et se contractent au gré de la chaleur, la plasticité d'une interface dénote de sa capacité à s'adapter au contexte d'usage dans le respect de son utilisabilité* » (page 4). Pour Sarter (2007), l'interface adaptative s'ajuste par soi-même selon le contexte d'utilisation et de la demande, à l'inverse d'une interface adaptable pour laquelle c'est l'opérateur qui aura la main sur les changements d'informations. Les interfaces adaptatives s'adaptent selon des « Événements » qui vont entraîner des « Réactions » (Figure 3.5).

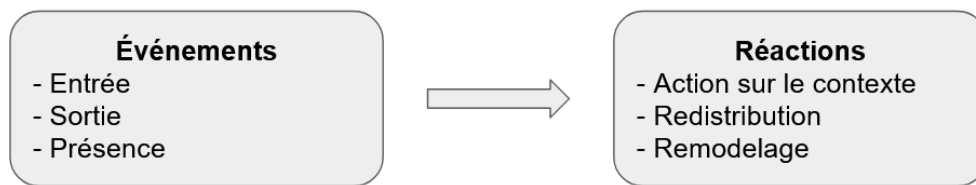


Figure 3.5 : Événements et réactions dans une interface adaptative (adapté de Calvary & Coutaz, 2002)

Les événements. Dans le cadre d'une interface adaptative, c'est l'utilisation qui fait le lien entre l'événement et la modification de l'interface. Pour une interface adaptative, c'est un algorithme sous-jacent qui détectera et interprètera les événements pour modifier l'interface. Calvary et Coutaz (2002) identifient trois grands événements pour les interfaces adaptatives : l'entrée dans un contexte (exemple : arriver sur l'autoroute), la sortie d'un contexte (exemple : sortir de l'autoroute) et la présence d'un contexte (exemple : Ntemps passé à rouler sur l'autoroute). Les trois types d'événements se basent sur la détection de ce que l'on nomme « déclencheurs ». La littérature propose tout type de déclencheurs pour les interfaces adaptatives (Scerbo, 1996 ; Rothrock, Koubek, Fuchs, Haas & Salvendy, 2002 ; Sarter, 2007 ; Ulahannan, Jennings, Oliveira & Birrel, 2020). Il est possible d'utiliser des mesures physiologiques à partir d'un électro-encéphalogramme (Saiwaki et al., 1996 ; Aricò et al., 2016) ou le parcours oculaire avec un eye-tracker (Komogortsev, Holland & Camou, 2011). Il est également possible de prendre en compte la performance (Morrisson, Cohen & Gluckman, 1993) ou les caractéristiques de l'opérateur (Sarter, 2007). Plus précisément dans les caractéristiques de l'opérateur, certains auteurs ont utilisé les préférences culturelles (Reinecke & Bernstein, 2011) ou la fonction d'opérateur (i.e. son métier, Akiki, 2013). Enfin, les états cognitifs de l'opérateur peuvent également servir de déclencheur. Il est possible d'utiliser la charge mentale (Kaber & Riley, 1999) ou la conscience de la situation (Harper, Guarino & Zacharias, 2000).

Les réactions. La présence d'un événement va entraîner des réactions dans l'interface. Ces réactions peuvent-être de l'ordre de l'action sur le contexte, de la redistribution ou du remodelage. L'action sur le contexte se définit lorsque l'interface s'adaptera au contexte d'utilisation : l'entrée dans le contexte « nuit » entraînera une réaction sur le contexte comme par exemple « baisse de la luminosité de l'interface ». La redistribution est le fait de faire migrer l'interface sur un autre support. Par exemple, l'action « sortie de la voiture » entraîne la redistribution de l'interface de gestion musicale qui était sur l'écran de la voiture vers celui

du téléphone. Le remodelage est le fait de transformer l'interface en termes d'information, de présentation d'information. Par exemple la présence du contexte « non-réponse à un alarme » qui donnera lieu au remodelage « mise en avant de cette alarme » par une surbrillance. Les trois types de réactions ne feront pas appel aux mêmes leviers techniques et aux mêmes compétences de conception. L'action sur le contexte et la redistribution nécessiteront avant tout des changements pour favoriser la lisibilité et sur quel support transmettre les informations. Le remodelage de l'interface en revanche va jouer sur la façon dont se déroule la coopération en modifiant les informations ou en automatisant certaines tâches. Rothrock et al. (2002) définissent quatre grands types de remodelage. Pour remodeler une interface il est possible de modifier :

- Le contenu de l'information : l'interface doit pouvoir modifier les informations qu'elle transmet (comme rendre transparente ou occulter certaines informations), leurs quantités (créer ou enlever des redondances), ajouter des informations (augmenter les feedback), la mise en page (regrouper certaines informations) et leurs modalités de présentations (ajouter des informations sonores).
- La structure du dialogue : l'interface doit pouvoir modifier sa structure de navigation (liens, hypertextes, etc.) pour s'adapter à l'opérateur.
- L'allocation des tâches : Rothrock et al. (ibid) ne différencie pas l'interface du CPS. Ils proposent donc d'intégrer le changement d'allocation des tâches du CPS comme réaction d'une interface adaptative. « L'interface » doit pouvoir augmenter ou réduire son automatisation.
- Le moment et la vitesse d'adaptation : en fonction du contexte ou de l'opérateur.

Canito, Mota, Marreiros, Corchado et Martins (2021), définissent que pour l'industrie 4.0, les déclencheurs importants à utiliser sont ceux liés aux contextes d'utilisation et ceux liés aux opérateurs. Si la conscience de la situation et la charge mentale ont déjà été utilisées, il semble possible d'utiliser d'autres des métriques identifiées dans la CHM. Nous pensons tout particulièrement à la confiance qui pourrait servir de déclencheur lorsque l'opérateur se trouve dans une zone de surconfiance ou de sous-confiance. Nous voyons également ici un lien naturel entre la transparence et le contenu de l'information. Plus précisément le fait de pouvoir modifier les informations que transmet le CPS par le biais de son interface.

Cette réflexion sur l'utilisation de la confiance comme déclencheur pour un CPS et de la transparence comme une capacité adaptative a été mise en application lors de trois études. Akash, McMahon, Reid et Jain (2020) et Akash, Jain et Misu (2020) utilisent un algorithme afin de déterminer la probabilité de faire confiance au CPS via des modèles de Markov. Ces modèles de Markov prennent en compte à la fois la confiance et la charge mentale de l'opérateur. La prise en compte de cette probabilité de survenue d'un état par le CPS déclenchant sa réaction de modification de transparence. Sur les deux études, les auteurs observent que l'utilisation d'une transparence adaptative a des effets sur la confiance. Sur le fait de se fier à une voiture autonome (Akash et al., 2020a) et sur la confiance dans un contexte militaire (Akash et al., 2020b). La troisième expérimentation a été réalisée par Okamura et Yamada (2020). Le déclencheur utilisé est la présence d'un comportement de surconfiance envers le CPS. Lorsque ce comportement est détecté, le CPS communique des informations sur des indices de confiance (i.e. « es-tu sûr de faire le bon choix ? »). Ils ont constaté un effet de l'utilisation adaptative de la transparence sur l'utilisation du CPS. La présence de l'interface adaptative réduit l'utilisation du CPS dans les situations où il est plus avantageux pour le participant de faire l'opération lui-même. Cependant, l'utilisation d'une interface adaptative n'est peut-être pas applicable de la même façon à tous les opérateurs. Son utilisation et ses potentiels avantages semblent différer en fonction des caractéristiques et des motivations des opérateurs (Kortschot, Jamieson & Prasad, 2022). Kortschot et al. (ibid) ont classifié les opérateurs d'interfaces adaptatives dans différents groupes selon leurs performances. Selon le type d'adaptation de l'interface, il existe des effets bénéfiques dans un groupe mais pas dans un autre. Tous les participants n'ont pas la même réceptivité aux interfaces adaptatives.

3.3. Synthèse du chapitre 3

L'effet que peut avoir la transparence sur la confiance apprise, la confiance situationnelle et l'acceptation fait de ce concept un élément clé dans la prise de décision assistée par un CPS. En croisant le modèle proposé par Lyons (2013) et le modèle présenté dans la partie 2.4, nous pouvons mettre en évidence le lien entre les dimensions et les composantes de la confiance (Figure 3.6). La dimension de l'intention, de la tâche, analytique et de l'environnement viendraient influencer la confiance apprise. Il est possible de mettre

en lien les dimensions de la transparence avec les facteurs influençant la confiance apprise proposés par Hoff et Bashir (2015). La dimension de l'intention fait écho à la prévisibilité et à l'utilité, qui viennent influencer la confiance apprise dynamique (CAD). Les dimensions de la tâche et analytique sont à rapprocher de la performance du système pour la CAD. Les trois dimensions aident à la compréhension du système pour la confiance apprise initiale (CAI). La dimension « environnement » donne accès à une meilleure compréhension du fonctionnement et actions du système dans un environnement donné, il est en lien avec la CAI. Cette dimension viendra également influencer la confiance situationnelle en étant transparente sur des éléments de l'environnement susceptibles de modifier la conscience de la situation et la perception du risque de l'individu. Les dimensions de la coopération et de l'état de l'opérateur viendront également influencer la confiance situationnelle. La dimension de la coopération correspond à la notion d'organisation citée par Hoff et Bashir et qui est en lien avec la confiance situationnelle. Un CPS transparent sur l'état de l'opérateur donnera accès aux informations qu'il a acquis sur l'opérateur comme la charge mentale ou son stress. Ces deux éléments venant influencer la confiance situationnelle.

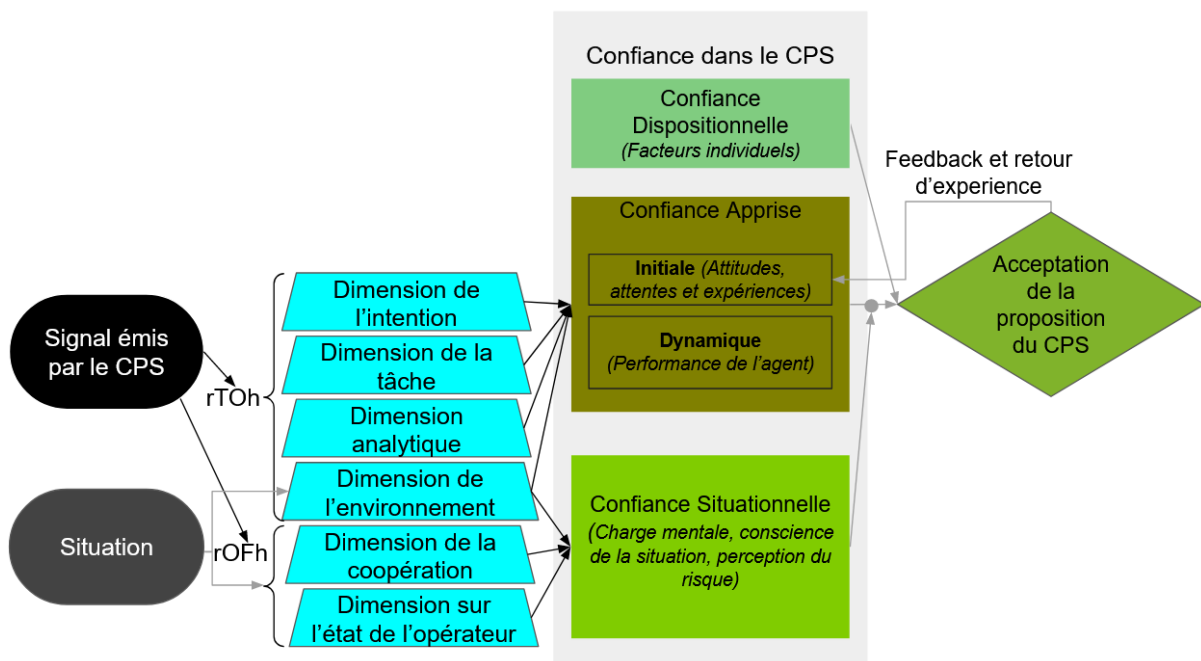


Figure 3.6 : Dimensions de la transparence sur les composants de la confiance dans une prise de décision assistée par un CPS

Notons que des informations sur la situation peuvent également être transmises par le CPS à l'opérateur. Ces informations sont transmises par la dimension de l'environnement

(exemple : les risques de la situation), la dimension de la coopération (exemple : l'allocation des fonctions) et la dimension sur l'état de l'opérateur (exemple : la situation est complexe et l'opérateur est en situation de stress).

Le concept de transparence est d'autant plus important que dans le cadre de l'industrie 4.0, des informations sont partagées entre l'humain et le CPS (données, processus ou propositions par exemple). Nous avons pu voir que les différents modèles couramment utilisés dans les études sur la transparence montrent certaines limites pour la compréhension des effets de la transparence (van de Merwe et al., 2022). Selon leur revue, qui corrobore les conclusions de Kluy et Roesler (2021), les études portant sur la transparence n'utilisent pas ce concept de manière suffisamment détaillée et fine. Une amélioration de la compréhension de la transparence par le biais des informations transmises permettra de mieux comprendre ses effets sur l'opérateur. L'objectif étant de comprendre les effets de chaque capacité de transparence (rTOh et rOFh) et des niveaux de transparence (1, 2 ou 3) sur les composantes de la confiance. De plus, l'effet de la transparence sur la perception du risque reste à ce jour peu, voire pas, étudié. Or le risque associé à une situation peut modérer le lien entre la confiance et le comportement de l'opérateur (Chancey et al., 2017 ; Hoesterey & Onnasch, 2022). Il semble donc important d'explorer cette thématique.

La coopération entre l'humain et la machine ne se joue pas sur des situations décorréelées les unes des autres mais est un processus continu où les situations passées viennent influencer les situations futures et la confiance dans le CPS (Hoff & Bashir, 2015). L'étude de la prise de décision assistée par un CPS sur du long terme plutôt que sur des situations décorréelées est également un challenge identifié par Steyvers et Kumar (2022). C'est pourquoi l'idée de faire évoluer les interfaces de communication entre l'humain et la machine dans le but de calibrer la confiance est une piste de recherche. Quelques études ont tenté d'utiliser la transparence de façon adaptative (Okamura & Yamada, 2020 ; Akash et al., 2020a ; Akash et al., 2020b). Ces études sont encore peu nombreuses et ce champ de recherche reste donc à investiguer.

Dans le cadre de cette thèse, la transparence sera étudiée sous :

- Le prisme de son effet sur les composantes de la confiance et sur l'acceptation de l'opérateur lors d'une prise de décision assistée par un CPS.
- Son intérêt en tant que capacité adaptative du CPS pour calibrer la confiance de l'opérateur

Pour nous aider à répondre à ces deux points en suspens dans la littérature de la transparence, nous nous intéressons aux méthodologies employées pour mesurer la confiance apprise ainsi que les facteurs liés à la confiance situationnelle.

Chapitre 4 : Méthodes et outils pour l'évaluation de la confiance

Ce chapitre a pour objectif de présenter les différentes approches permettant de quantifier la confiance envers un CPS lors d'une coopération. Nous ne traitons pas des mesures de la confiance dispositionnelle car celle-ci se base sur des facteurs individuels (Hoff & Bashir, 2015). Ces facteurs individuels sont récupérés par des questionnaires socio-démographique et de personnalités. Nous dressons une liste des outils à notre disposition permettant de mesurer les composantes de la confiance apprise et situationnelle. Nous nous appuyons sur les méthodes de quantification proposées dans la littérature. Nous catégorisons les différentes propositions de méthodologie selon quatre propriétés (Tableau 4.1). Cette catégorisation des mesures nous permet de définir quelle méthodologie de quantification semble la plus appropriée à utiliser dans notre contexte d'étude.

Tableau 4.1 : Propriétés des méthodologies de mesure

Nom de la mesure	Non-Invasive	Non-arrêt de la tâche	Sensible	Objectif/Subjectif
<i>Nom de la mesure</i>	<i>La mesure ne nécessite pas d'équiper le participant</i>	<i>La mesure ne nécessite pas d'interrompre la tâche</i>	<i>La mesure permet de différencier les variations</i>	<i>La mesure est objective ou subjective</i>

4.1. Mesurer la confiance apprise et l'acceptation

Ermisch, Gambetta, Laurie, Siedler et Noah (2021), tout comme Hoff et Bashir (2015), distinguent la confiance et le comportement qui en découle, et proposent des approches différentes pour les mesurer. La première approche se base sur des questionnaires utilisant des échelles, dites de Likert (Likert, 1932). Par cette approche, c'est l'appréciation du participant sur la confiance qu'il ressent envers le système cyber-physique (CPS) qui est investiguée. La seconde approche est une approche comportementale. Cette approche considère les réactions et comportements du participant, qui sont l'expression comportementale de la confiance envers le CPS en action. Les différentes méthodologies sont résumées dans le Tableau 4.2.

Afin de savoir si une personne fait confiance envers autrui, la méthode la plus souvent adoptée est l'utilisation d'échelles (Ermisch et al., 2021). Un score plus élevé sur l'échelle sera

révélateur d'une confiance plus élevée. Cette mesure par échelle de Likert permet de quantifier la confiance. Étant donné l'évolution cyclique de la confiance (le retour d'expérience venant modifier la confiance apprise initiale (Hoff & Bashir, 2015)), Yang, Schemanske et Searle (2021) conseillent de prendre une mesure de confiance après chaque essai ou interaction avec un CPS. De nombreux auteurs font des propositions de questionnaires (Chita-Tegmark, Law, Rabb & Scheutz, 2021 ; Law et Scheutz, 2021) ce qui fait que les outils à disposition sont très divers. Certains questionnaires sont créés pour une expérimentation spécifique et ont seulement quelques items (Rossi, Holthaus, Dautenhahn, Koay & Walters, 2018 ; Shu et al., 2018 ; Morgan, Williams, Flower, Alford & Parkin, 2019). D'autres questionnaires sont utilisés de façon plus générale et comportent une dizaine d'items (Jian, Bisantz & Drury, 2000 ; Schaefer, 2013).

Chita-Tegmark et al. (2021) ont comparé quatre questionnaires. Cette comparaison avait pour objectif de définir le questionnaire le plus facilement compréhensible et celui qui pouvait être appliqué à tous types de CPS. Il ressort de leur étude que le questionnaire le plus généralisable et compréhensible est l'IMOT-RIS proposé par Lyons & Guznov (2019) (Annexe E.1). Ce questionnaire, en dix items, se base sur le questionnaire à quatre items proposés par Schoorman, Mayer et Davis (1996) (cité par Schoorman, Mayer & Davis, 2016). Ce questionnaire a pour but premier les coopérations humains-humains, et les intitulés ont donc été modifiés pour correspondre à une coopération humain-machine (CHM). Lyons et Guznov ont intégré six items supplémentaires à ce premier questionnaire. Leur questionnaire s'intéresse à la question de la confiance sur la base de l'intention d'adopter un comportement de confiance. Comme pour la plupart des questionnaires une moyenne des items est faite, et plus le score est élevé plus la confiance dans le CPS est forte.

Le second type de mesure est celui des mesures comportementales (Luster & Pitts, 2021). Le but est d'observer si le participant a un comportement de confiance, ou non, envers le second agent. Law et Scheutz (2021) distinguent quatre types de mesures comportementales :

- **L'intervention dans la tâche.** Cette mesure consiste à regarder le nombre de fois où le participant reprend la main pendant que le CPS réalise une tâche. Ce type de mesure est souvent utilisé lorsque l'objectif est de tester des CPS amenés à réaliser des tâches

de tous les jours (exemple des voitures autonomes qui sont vouées à être utilisées au quotidien) (Pedersen, Andersen, Kösllich, Weigelin & Kuusinen, 2018 ; Chen et al., 2018 ; Heldin et al. 2013). Dans ce cadre, on considère que plus le participant reprend la main, moins il fait confiance au CPS.

- **La délégation de la tâche.** Cette mesure consiste à regarder le choix qu'un participant fait entre confier une tâche à un CPS ou confier une tâche à un autre opérateur. Cette comparaison peut également se faire entre deux CPS. On considère ici que le choix du participant correspond à l'agent (humain ou autonome), auquel il fait le plus confiance. Cette mesure est utilisée pour différencier dans quel type d'agent (humain ou autonome) la confiance est plus élevée pour réaliser une tâche prédéfinie (Sanders et al., 2019 ; Xie, Bodala, Ong, Hsu & Soh, 2019).
- **Le changement de comportement.** Cette mesure consiste à observer le comportement qu'a le participant face au CPS. Cette mesure utilise différents indices qui permettent de définir si le comportement du participant est associé à la présence ou à l'absence de confiance. Ces indices sont multiples et peuvent être par exemple : le temps de réponse face à une proposition (Kintz et al., 2022) ; la distance entre le participant et un véhicule autonome (Jayaraman et al., 2019) ; des signes de stress (Weigelin, Mathiesen, Nielsen, Fischer & Nielsen, 2018). Ces indices sont différents en fonction de la situation observée.
- **Le suivi de proposition.** Cette mesure consiste à regarder si un participant accepte et donc suit la proposition ou le conseil d'un CPS. Cette mesure est souvent utilisée lorsque le participant et le CPS sont coéquipiers ou lorsqu'ils ont un but commun. Cette mesure est parfois appelée « agreement rate » (Chen, Mishler & Hu, 2021) ou « compliance » (Wang et al., 2016 ; Akash et al., 2020b ; Pynadath, Gurney & Wang, 2022). Si un participant accepte une proposition d'un CPS, on va considérer qu'il fait confiance.

Bolton (2022) souligne qu'il est plus que primordial de considérer le comportement (d'accepter ou de se fier) dans la CHM. Comme nous l'avons montré dans le second chapitre, la confiance dans sa définition est une attitude qui s'exprime par un comportement (Mayer et al., 1995). La confiance est la probabilité d'adopter un comportement. Il est donc opportun

de multiplier les mesures à différents moments : la confiance grâce à des questionnaires ; l'acceptation grâce à l'analyse de la décision du participant.

Tableau 4.2 : Propriétés des mesures de confiance

Nom de la mesure	Non-Invasive	Non-arrêt de la tâche	Sensible	Objectif/Subjectif
Questionnaire	X	-	X	Subjectif
Intervention dans la tâche	X	X	X	Objectif
Délégation de la tâche	X	X	X	Objectif
Changement de comportement	X	X	X	Objectif
Suivi de proposition	X	X	X	Objectif

4.2. Mesurer la confiance situationnelle : conscience de la situation et la perception du risque

Depuis le début des travaux sur la conscience de la situation plusieurs méthodologies ont été proposées. Parmi ces méthodologies, nous retrouvons : des mesures sur les processus, des mesures sur la performance, des mesures utilisant des questionnaires d'échelle et des mesures utilisant des questionnaires de connaissances (Endsley, 2021a). Endsley (2021a) différencie ces quatre méthodologies selon qu'elles utilisent des mesures directes ou des mesures indirectes. Les mesures de processus et de performance sont considérées comme indirectes et les mesures par questionnaire sont des mesures directes. Nous revenons sur chacune de ces méthodologies pour expliciter les outils employés (Tableau 4.3) :

- **Mesures sur les processus.** Cette catégorie inclut plusieurs méthodologies comme l'eye-tracking, l'étude des communications ou encore l'étude des mesures physiologiques. L'avantage de ces méthodologies est leur utilisation en continu.
 - Pour l'eye-tracking il existe un double inconvénient. Premièrement l'absence d'informations sur le recueil de données autre par le participant. Se baser

uniquement sur la méthodologie de l'eye-tracking occulte tout les autres formes de récupération d'information (auditive, sensitive) utilisées par le participant. Secondement, il est impossible de savoir si les informations sont bien intégrées dans la conscience de la situation (Endsley, 1995a ; Nguyen, Lim, Nguyen, Gordon-Brown & Nahavandi, 2019). Le participant peut regarder sans comprendre ou interpréter l'objet.

- L'analyse des communications se fait sur la base de communications spontanées, c'est-à-dire le dialogue entre deux agents d'un système (Chauvin, Coppin & Chené, 2010 ; Lassalle et al., 2017). L'inconvénient de cette méthodologie est que tout le monde ne verbalise pas de la même façon ce qui empêche un accès complet à la conscience de la situation (Endsley, 2021a).
- Les mesures physiologiques pour mesurer la conscience de la situation sont multiples (du type électroencéphalogramme, image par résonance magnétique ou rythme cardiaque, Wilson, 2000). Il existe cependant peu d'études sur le lien existant entre ces mesures physiologiques et la conscience de la situation. Les mesures physiologiques sont en effet difficilement différenciables d'autres éléments pouvant provoquer des variations (Sirkin, Martelaro, Johns & Ju, 2017 ; Nguyen et al., 2019).
- **Mesures sur la performance.** Cette catégorie inclut le temps de réponse et les erreurs. L'avantage de ces mesures est qu'elles peuvent être récoltées sans intervention spécifique du participant. En revanche, il peut être dangereux de considérer que la performance équivaut à une bonne conscience de la situation car cela exclut d'autres facteurs (facilité de la tâche ou utilisation de compétences de bas niveau (SRK, Rasmussen, 1983) par exemple). Le temps de réponse et le taux d'erreur sont des indices, mais ne permettent pas de statuer sur le niveau de conscience de la situation.
- **Mesures par questionnaire d'échelle.** Les mesures subjectives se présentent sous la forme de questionnaires comportant des échelles de ressenti de la situation. Un exemple de ce type de questionnaire est le Situation Awareness Rating Technique (Taylor, 1990 dans Taylor, 2011). Cette échelle à l'avantage d'être applicable à plusieurs domaines sans nécessité d'être adaptée, ce qui la rend facile d'utilisation. Cependant, cette échelle intègre d'autres concepts tels que la charge mentale ou l'attention, pouvant limiter la validité de la mesure. Autre inconvénient, elle nécessite

d'arrêter la situation et de mesurer le ressenti a posteriori. La principale limite de cette échelle est qu'elle ne permet pas de s'assurer que le participant a bien connaissance de son environnement, son ressenti peut être différent de la situation. Auquel cas sa conscience de la situation n'est pas adéquate, mais le participant n'en a pas conscience.

- **Mesures par questionnaire de connaissances.** Les questionnaires utilisés ici sont des questionnaires de connaissances sur la situation. Un exemple de questionnaire est le Situation Awareness Global Assessment Technique, SAGAT (Endsley, 1995a). L'avantage de ce questionnaire est d'obtenir un score objectif, ce qui permet de déterminer si le participant a une conscience de la situation appropriée. De plus, chaque questionnaire est fait pour une situation déterminée ce qui fait qu'il n'y a pas de risque d'apprentissage. Cet avantage représente également un inconvénient majeur, car chaque nouvelle situation nécessite la création d'un nouveau questionnaire. Ce nouveau questionnaire nécessite, pour sa création, d'avoir les connaissances relatives à la situation. De plus, le questionnaire, tout comme les échelles, impose d'arrêter la passation pour que le participant puisse répondre aux questions. Cela le force également à se remémorer des éléments, pouvant créer un biais. En revanche, cette méthode permet de corrélérer directement une conscience de la situation plus élevée avec de meilleurs scores aux questionnaires. Toujours avec l'exemple du SAGAT, plusieurs items seront créés pour chaque niveau de la conscience de la situation. Au premier niveau, sont utilisés des items portant sur la perception du danger et plus globalement sur la perception de la situation. Pour le second niveau de conscience de la situation, les items portent sur ce qui a amené à cette situation. Au troisième niveau, les items portent sur les conséquences de la situation et les projections à venir.

Pour Endsley (2021b), les mesures directes sont plus sensibles (i.e. elles différencient mieux deux situations) et permettent de meilleurs diagnostics.

Pour rappel, nous considérons ici la perception du risque comme corollaire à la conscience de la situation. La méthodologie la plus souvent utilisée pour quantifier la perception du risque est celle de questionnaires composés d'échelles (Sjöberg, Moen &

Rundmo, 2004). Ces questionnaires sont utilisés aussi bien pour quantifier la perception du risque de façon générale (Wilson, Zwickle & Walpole, 2018) que sur des thématiques spécifiques comme par exemple les études tabacologiques (Kaufman, Persokie, Twersten & Bromberg, 2018). Pour ces questionnaires, plus un score est élevé plus la perception du risque est évaluée comme élevée. De récentes études ont tenté d’explorer le lien entre la perception du risque et l’utilisation de métriques physiologiques (Perello-March, Burns, Birrel, Woodman & Elliott, 2022 ; Petit, 2022). Les auteurs ont notamment exploré le lien entre le rythme cardiaque, la conductivité de la peau et un questionnaire. Selon Petit (2022), l’utilisation de mesures physiologiques pourrait être complémentaire aux mesures par questionnaire.

Tableau 4.3 : Propriétés des mesures de conscience de la situation et de perception du risque

Mesures	Non-Invasive	Non-arrêt de la tâche	Sensible	Objectif/Subjectif
<i>Mesure sur les processus (eye-tracking et physiologique)</i>	-	X	-	<i>Objectif</i>
<i>Mesure sur les processus (analyse des communications)</i>	X	X	-	<i>Objectif</i>
<i>Mesure sur la performance</i>	X	X	-	<i>Objectif</i>
<i>Questionnaire (échelles de Likert)</i>	X	-	X	<i>Subjective</i>
<i>Questionnaire (de connaissance)</i>	X	-	X	<i>Objectif</i>

La notion de perception du danger et de perception du risque forment un continuum. En amont, et directement dans la conscience de la situation, il y a la perception du danger (i.e. ce qui peut causer des dommages), et en aval la perception du risque (i.e. la probabilité que ce dommage survienne). Il semble donc intéressant d’utiliser un questionnaire, permettant de mesurer la perception du danger, et d’utiliser un second questionnaire, pour la perception du risque.

4.3. Mesurer la confiance situationnelle : charge mentale

La charge mentale a fait l’objet de nombreuses recherches pour déterminer les méthodes les plus aptes à la quantifier. Selon la proposition de Kostenko (2017) il est possible

de mesurer la charge mentale selon trois dimensions. Nous nous intéressons ici seulement à la dimension « effet » de la charge mentale (i.e. effet de la situation sur le participant). Nous ne nous intéressons pas à la dimension cause (i.e. la situation en elle-même) ni à la dimension médiatrice (i.e. les stratégies déployées par le participant). Au sein de cette dimension « effet », Kostenko (2017) inclut trois catégories : les conséquences physiologiques, la performance et les questionnaires. Brunken, Plass et Leutner (2010) placent ces différentes métriques selon leur relation causale (indirecte/directe) et leur objectivité. Les différentes méthodologies sont résumées dans le Tableau 4.4.

Il existe deux types de mesures physiologiques. Le premier type de mesure est indirect car ce sont des mesures physiologiques qui sont le reflet d'une activité du cerveau. Cette activité du cerveau crée des activités électriques à travers le corps (Kostenko, 2017), et ce sont ces activités électriques qui sont observées. Le second type de mesure est direct car l'observation se fait directement au niveau de l'activité électrique du cerveau. L'intérêt d'utiliser ces deux types de mesure est que le participant n'est pas interrompu dans sa tâche. Les mesures physiologiques sont également standardisables, facilement comparables entre plusieurs études et elles ne nécessitent pas de grands échantillons de participants (Tao et al., 2019). Il existe de nombreuses mesures physiologiques permettant d'inférer sur la charge mentale (Kramer, 1991 dans Kramer, 2020 ; Lean & Shan, 2012 ; Kostenko, 2017 ; Tao et al., 2019) :

- **Activité électrodermale** (indirect/objectif). Le principe est d'observer les variations de conductivité électrique de la peau, via l'utilisation de capteurs, pour inférer sur l'état du participant. L'existence de variations indique une activation du système nerveux, provoqué par un changement dans la charge mentale (Mehler, Reimer, Coughlin & Dusek, 2009 ; Collet, Salvia & Petit-Boulanger, 2014). Le point négatif est que ces variations peuvent être induites par d'autres éléments tels que le stress ou l'émotion (Kostenko, 2017). Cette limite rend donc difficile la discrimination de ce qui a provoqué le changement de conductivité électrique.
- **Activité cardiaque** (indirect/objectif). Il existe deux méthodologies permettant de vérifier l'activité cardiaque. La première permet de mesurer le volume sanguin périphérique en regardant les variations de quantité sanguine. Cette variation de la

quantité sanguine permet d'inférer sur le rythme cardiaque. La seconde méthode est l'utilisation d'électrocardiogramme pour étudier l'activité électrique du cœur. Si les variations cardiaques sont corrélées avec la charge mentale (Meshkati, 1988 ; Mehler et al., 2009), au même titre que l'activité électrodermale, elles peuvent être provoquées par de multiples facteurs (stress, effort physique, effort mentale, etc.) ce qui rend la discrimination difficile.

- **Activité neuronale** (direct/objectif). L'utilisation de l'électroencéphalogramme permet de mesurer les variations électriques du cerveau du participant. Les ondes produites par le cerveau sont captées grâce à des électrodes placées sur le crâne du participant. C'est leur variation qui permet de définir la charge mentale du participant. Certains types d'ondes, comme les ondes alpha, auront tendance à baisser lors de la hausse d'une charge mentale, lorsque d'autres ondes (θ , β , δ et γ) auront tendance à augmenter (Tao et al., 2019).
- Les **activités oculaires** sont issues d'une grande famille de mesures physiologiques et jusqu'à treize mesures différentes peuvent être recensées. Parmi ces treize, trois sont particulièrement utilisées (Tao et al., 2019) :
 - Le **diamètre pupillaire** (indirect/objectif). Ici c'est la variation du diamètre pupillaire qui permet d'indiquer une variation dans la charge mentale. Beatty (1982) a montré, dans une méta-analyse, que ce diamètre pupillaire est corrélé avec différents processus cognitifs tels que l'utilisation de la mémoire de travail (se remémorer des séries de chiffres), les processus langagiers (reconnaissance de lettre ou du sens d'une phrase), les raisonnements (résolution mathématique), la détection de signal (auditif ou visuel) et l'attention sélective (reconnaître un son parmi d'autres). Lorsque le diamètre pupillaire s'élargit, la charge mentale augmente (Tsai, Viirre, Strychacz, Chase & Jung, 2007). Encore une fois les variations pupillaires peuvent être influencées par d'autres facteurs, rendant cette mesure peu discriminante (Cegarra & Chevalier, 2008).
 - **Fixation** (indirect/objectif). Le temps de fixation est analysé afin de mesurer sa corrélation avec la charge mentale du participant. Le sens de cette relation est en revanche variable selon les études. Dans certains cas, le temps de fixation augmente lorsque la charge mentale augmente (Greef, Lafeber, Oostendorp &

Lindenberg, 2009 ; Lean & Shan, 2012) alors que dans d'autres études le temps de fixation baisse (Tao et al., 2019).

- **Clignement** (indirect/objectif). Le temps entre chaque clignement des yeux du participant est comptabilisé afin d'en déduire la charge mentale. Lorsque le temps entre chaque clignement augmente, cela indique que la charge mentale augmente (Tao et al., 2019 ; Lean & Shan, 2012).

Lors d'une expérimentation, toutes les outils permettant de mesurer ces données physiologiques nécessitent d'équiper le participant. C'est ce qui rend ces mesures complexes à mettre en place et qu'elles peuvent représenter un frein chez les participants. Comme le soulignent Cegarra et Chevalier (2008), si les mesures physiologiques sont sensibles (elles varient suffisamment entre deux situations pour être détectées), elles restent peu discriminantes sur ce qui les provoque. Il est difficile de discerner si les variations (électrodermales, cardiaques ou oculaires) sont bien provoquées par une hausse de la charge mentale ou par d'autres causes (stress, peur, etc.).

La seconde catégorie proposée par Kostenko (2017) pour mesurer la charge mentale est la performance. La performance est une mesure indirecte et objective de la charge mentale (Brunken et al., 2010). La performance d'un participant peut être comparée à une performance de référence pour permettre d'inférer sur la charge mentale. Cegarra et Chevalier (2008) distinguent deux mesures de performance :

- **Performance à une tâche.** L'utilisation du temps de réponse, du nombre d'erreurs ou de la vitesse permet d'établir la performance à une tâche. Cette performance est corrélée avec la charge mentale du participant, une bonne performance est révélatrice d'une charge mentale appropriée (i.e. ni trop faible, ni trop élevée) (Tsai et al., 2007 ; Mehler et al., 2009).
- **Performance à une tâche secondaire.** L'utilisation d'une tâche secondaire (voire de tâches multiples comme dans le simulateur Multi-Attribute Task Battery (MATB-II)), permet de savoir si la charge mentale à disposition est suffisante pour gérer ces différentes tâches simultanément. Tout comme pour la performance à la tâche principale, une performance plus faible aux tâches secondaires est un indicateur d'une charge mentale élevée.

La troisième catégorie est celle des questionnaires. Les questionnaires utilisent notamment des échelles de Likert. Les échelles vont servir à interroger le participant sur son ressenti de sa charge mentale. Un des questionnaires couramment utilisé est le NASA-TLX proposé par Hart et Staveland (1988) (au moins 550 études l'ont utilisé entre 1988 et 2006, Hart, 2006). Le NASA-TLX est découpé en 6 dimensions, avec un item explicite par dimension. Ces dimensions sont l'exigence temporelle, l'exigence physique, l'exigence mentale, l'effort, la performance et la frustration ressentie durant la tâche. Chaque dimension est ensuite pondérée et comparée aux autres dimensions, puis les scores finaux sont moyennés permettant d'obtenir un indice global de charge mentale. Un second questionnaire, plus simple dans sa passation, est l'Instantaneous Self-Assessment (Tattersall & Foord, 1996). Ce questionnaire se base sur une question unique posée au participant. Le participant est invité à répondre, sur une échelle de Likert en 5 points, sur le ressenti de sa charge mentale. Durant l'activité, on mesure si sa charge mentale était très utilisée (avec la modalité « *excessive* ») ou sous-utilisée (avec la modalité « *sous-utilisée* »). Ce questionnaire, même si basique, est corrélé avec les mesures physiologiques et à d'autres questionnaires de mesure de la charge mentale. Que ce soit le NASA-TLX ou l'ISA, l'usage de questionnaire a un inconvénient majeur qui est l'interruption de la tâche en cours. Cet écueil est cependant à modérer si le participant a des pauses préexistantes entre les différentes tâches. Les mesures subjectives sont intéressantes pour plusieurs raisons. Elles sont non-intrusives (contrairement aux mesures physiologiques), elles sont faciles à implémenter et elles sont étayées théoriquement. De plus, les réponses d'un même participant à travers plusieurs questionnaires sont consistantes (Cegarra & Chevalier, 2008).

Le constat partagé est que pour observer avec précision la charge mentale et ses variations, l'utilisation de plusieurs méthodologies (physiologiques, performances et questionnaires) est la solution la plus intéressante (Cegarra & Chevalier, 2008 ; Kostenko, 2017).

Tableau 4.4 : Propriétés des mesures de charge mentale

Mesures	Non-Invasive	Non-arrêt de la tâche	Sensible	Objectif/Subjectif
<i>Activité électrodermale</i>	-	X	X	<i>Objectif</i>
<i>Activité cardiaque</i>	<i>Selon la technologie employée</i>	X	X	<i>Objectif</i>
<i>Activité neuronale</i>	-	X	X	<i>Objectif</i>
<i>Activité oculaire</i>	<i>Selon la technologie employée</i>	X	X	<i>Objectif</i>
<i>Performance (sur la tâche et secondaire)</i>	X	X	X	<i>Objectif</i>
<i>Questionnaire</i>	X	-	X	<i>Subjectif</i>

4.4. Synthèse du chapitre 4

Nous avons vu dans ce chapitre qu'il existe de nombreuses méthodologies permettant de quantifier les composantes de la confiance de l'opérateur et sa prise de décision dans la CHM. Dans le cadre de nos recherches nous devons choisir quels indicateurs privilégier pour quantifier ces composantes. L'utilisation de toutes ces méthodes de façon concourante a comme risque de surcharger les futurs participants des expérimentations, en créant des protocoles expérimentaux trop complexes. C'est pour cela que nous effectuons des choix parmi les différentes méthodologies et pour les différentes composantes :

- La confiance étant notre premier sujet d'investigation, nous souhaitons la mesurer en détail, en utilisant des « capteurs » sur l'attitude et le comportement. Premièrement, nous observons la confiance grâce à l'utilisation de questionnaires. Secondement, notre cadre d'étude portant sur la prise de décision assistée par un CPS, la mesure de l'acceptation de la proposition apparaît comme une évidence (Tableau 4.5).
- Pour la conscience de la situation, nous investiguons plus particulièrement la perception du danger et la perception du risque qui en découle. Les mesures

physiologiques existantes dans la littérature ne permettent pas, à priori, de pouvoir les quantifier avec précision. C'est pourquoi nous choisissons d'utiliser des questionnaires de connaissances portant spécifiquement sur la perception du danger (items de niveau 1 du SAGAT) et des questionnaires utilisant des échelles de Likert pour la perception du risque. (Tableau 4.5).

- Nous contrôlons la charge mentale par l'utilisation de questionnaire, par la mesure de la performance et par l'utilisation de capteurs physiologiques (Tableau 4.5). Ainsi, conformément à la proposition de Cegarra et Chevalier (2008), nous multiplions les méthodes de mesure de la charge mentale.

Tableau 4.5 : Propriétés des mesures utilisées selon les composantes de confiance

Mesures	Non-Invasive	Non-arrêt de la tâche	Sensible	Objectif/Subjectif	Composantes de confiance et comportement
<i>Confiance (questionnaire)</i>	X	-	X	<i>Subjectif</i>	<i>Composantes CAD/CAI/CS</i>
<i>Acceptation (mesure comportementale)</i>	X	X	X	<i>Objectif</i>	<i>Comportement</i>
<i>Compréhension de la situation (questionnaire de connaissance)</i>	X	-	X	<i>Objectif</i>	<i>Composante CS</i>
<i>Compréhension de la situation (questionnaire)</i>	X	-	X	<i>Subjectif</i>	<i>Composante CS</i>
<i>Charge mentale (questionnaire)</i>	X	-	X	<i>Subjectif</i>	<i>Composante CS</i>
<i>Charge mentale (performance)</i>	X	X	X	<i>Objectif</i>	<i>Composante CS</i>
<i>Charge mentale (mesure oculaire)</i>	-	X	X	<i>Objectif</i>	<i>Composante CS</i>

Pour chacune des dimensions que nous explorons, nous utilisons au moins un questionnaire. Cette pratique est courante dans les expérimentations s'intéressant aux trois dimensions : confiance, conscience de la situation et charge mentale. (Kintz et al., 2022). Notons que si ces

questionnaires nécessitent d'interrompre la tâche, ce n'est pas un inconvénient pour les expérimentations, telles que nous les avons prévues et qui sont décrites dans les trois chapitres expérimentaux. Celles-ci sont composées de situations courtes avec un temps d'arrêt entre situations. Les questionnaires ne viennent donc pas interrompre la tâche, mais sont remplis à la fin de chaque situation. Il ne semble pas exister à ce jour de consensus scientifique sur le lien entre les métriques physiologiques et le concept de conscience de la situation. C'est pourquoi, nous privilégions des mesures non-invasives. Cela permet de simplifier le protocole expérimental, son acceptabilité auprès des participants et simplifiera l'analyse des résultats. Cependant, au vu de l'importance de croisement de données pour la charge mentale, l'intégration des données oculaires semble être le choix le plus pertinent par sa facilité de captation (lunettes ou barre d'eye-tracking).

Chapitre 5 : Problématique

5.1. Objectifs des travaux

L'intégration des systèmes cyber-physique (CPS) dans l'industrie 4.0 amène dans son sillage de nouvelles interactions humains-machines. Ces changements de coopération sont notamment liés à un nouveau paradigme où l'opérateur coopère avec un CPS (Human Autonomy Teaming) en lieu et place d'interagir avec un automate (Human Autonomy Interaction). La coopération sur laquelle nous portons notre intérêt est plus précisément la prise de décision assistée par un CPS. Cette classe de problème est peu explorée du point de vue des facteurs humains (Steyvers & Kumar, 2022).

Les objectifs de cette thèse sont de mieux comprendre le rôle des composantes de la confiance dans la prise de décision et de mieux comprendre les effets de la transparence sur ces composantes. Ces objectifs répondent à des enjeux d'amélioration de la coopération. En effet, une meilleure coopération implique de meilleures performances et une plus grande sécurité pour l'opérateur. Une meilleure coopération a aussi pour objectif de maintenir l'opérateur dans la boucle (et éviter une perte de compétence par exemple) ou qu'ils ne fassent un mauvais usage du CPS (voir les célèbres travaux de Parasuraman & Riley (1997) : Use, Misuse, Disuse, Abuse). Afin de répondre aux objectifs de cette thèse nous avons identifié trois verrous scientifiques à investiguer (Figure 5.1).

Premièrement, améliorer les connaissances sur les composantes de la confiance sous-jacentes à la prise de décision assistée par un CPS.

Secondement, améliorer la compréhension du rôle des modalités de transparence sur les composantes de la confiance dans la prise de décision par une utilisation croisée des modèles de Lyons (2013) et de Chen et al. (2018).

Troisièmement, investiguer l'utilisation de mécanismes adaptatifs de la transparence en vue de calibrer la confiance de l'opérateur envers le CPS.

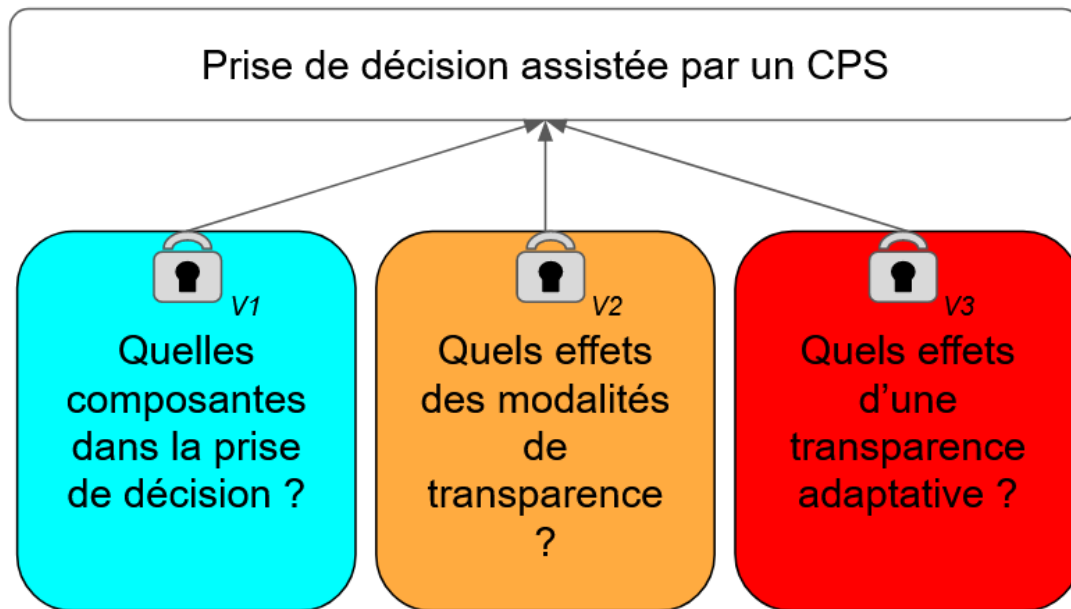


Figure 5.1 : Synoptique des verrous scientifiques

5.2. Verrou 1 : Compréhension du rôle des composantes de la confiance dans la prise de décision de l'opérateur

Le travail de recherche proposé ici vise à comprendre le rôle des composantes de la confiance (Hoff & Bashir, 2015) et parmi celles-ci, lesquelles sont des déterminants de la prise de décision de l'opérateur lorsqu'il est assisté par un CPS (Figure 5.2). Comme signalé par Steyvers et Kumar (2022), les facteurs liés à l'opérateur sont souvent délaissés dans les études portant sur ces situations de coopération. Dans cette thèse, nous inspectons les différentes composantes de la confiance de l'opérateur pour comprendre lesquels déterminent la prise de décision de l'opérateur. Cette prise de décision (i.e. accepter ou refuser une proposition) est considérée comme un indicateur objectif de la confiance du point de vue l'opérateur (Chen et al., 2021). D'où notre première question de recherche :

Q1a : Quel est le rôle des composantes de la confiance sur la prise de décision assistée par un CPS ?

Nous investiguons également l'interaction entre la confiance apprise et la confiance situationnelle dans le cadre d'une prise de décision assistée par un CPS (Figure 5.2). En effet, il est couramment accepté que le risque joue un rôle dans la prise de décision d'un opérateur (Williams & Noyes, 2007), même si son rôle reste le sujet de questionnements scientifiques

(Hoesterey & Onasch, 2022 ; Atchley et al., 2023). Les travaux sur l'automatisation sont souvent réutilisés dans le cadre des études impliquant des CPS (Mc Neese, Demir, Cooke & Myers, 2018). C'est pourquoi nous réutilisons les travaux de Chancey et al. (2017), basés sur l'automatisation, portant sur la prise de décision d'un opérateur relative à une alarme. Leur proposition de modèle inclut un rôle modérateur du risque dans cette prise de décision. Nous rappelons également que la perception du risque est, selon le modèle de Hoff et Bashir (ibid), un processus pleinement intégré à la confiance situationnelle. D'où notre deuxième question de recherche :

Q1b : *Quel est le lien entre la confiance situationnelle et la confiance apprise dans la prise de décision assistée par un CPS ?*

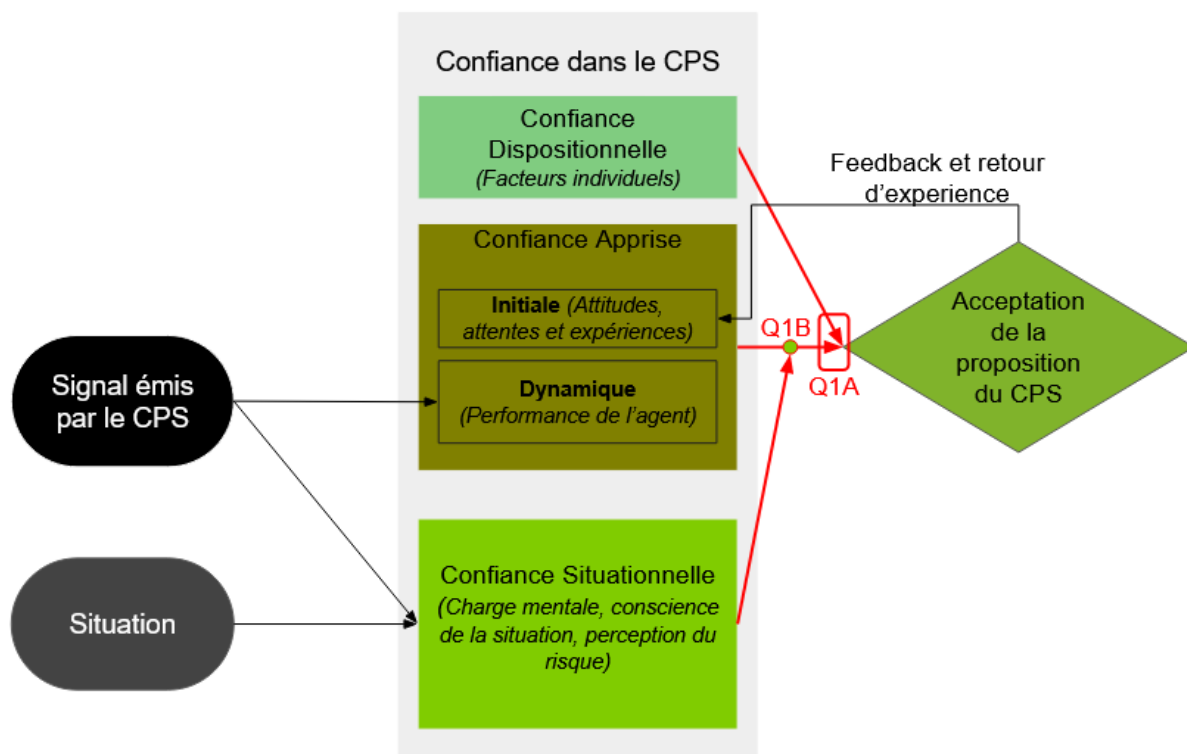


Figure 5.2 : Composantes de la confiance de l'opérateur dans le cadre de la prise de décision assistée par un CPS

Ces questions de recherche sont explorées dans le chapitre 6 : « *Confiance et transparence dans une prise de décision opérationnelle avec un CPS de fiabilité inconnue* », et dans le chapitre 7 : « *Confiance et transparence dans une prise de décision tactique avec un CPS fiable* ».

5.3. Verrou 2 : Compréhension des effets des modalités de la transparence

Les différents travaux menés sur le concept de transparence ont permis de montrer son utilité dans la coopération humain-machine (CHM) (Chen, 2021) mais plusieurs auteurs ont pointé un manque de connaissance sur les apports des différentes informations de la transparence dans la CHM (Kluy & Roesler, 2021 ; van De Merwe et al., 2022). En effet, les études expérimentales existantes utilisent les modèles de manière générale dans leurs protocoles. Les expérimentations que nous avons identifiées dans la littérature utilisent par exemple les niveaux de transparence proposés par Chen et al. (2014 ; 2018) en utilisant les trois niveaux comme variables (pour quelques exemples de protocoles voir : Mercado et al., 2016 ; Selkowitz et al., 2016 ; Stowers et al., 2020 ; Bhaskara et al., 2021). D'autres expérimentations restent floues sur leur utilisation du concept de transparence, avec par exemple la simple distinction « peu transparent/très transparent » (quelques exemples avec : Sadler et al., 2016 ; Pokam et al., 2019 ; Skraaning & Jamieson, 2021 ; Tatasciore, Bowden & Loft, 2023). Ces mêmes études ne distinguent pas les effets de la transparence selon différentes modalités (par exemple la dimension de l'état de l'opérateur (Lyons, 2013) selon leurs niveaux 1, 2 ou 3 (Chen et al., 2018). Notons également que les études portant sur la transparence, et qui ont introduit la notion de risque dans leur protocole, ne mesurent pas la perception du risque (Loft et al., 2021), laissant de côté la possibilité de comprendre le rôle de la transparence sur cette perception. Dans cette thèse nous chercherons donc à croiser les modèles du SAT (Chen et al., 2018) et du Human-Robot-Transparency model (HRT) (Lyons, 2013) pour comprendre de manière plus fine l'effet de différentes modalités sur la prise de décision et sur les composantes de la confiance. Nous appelons modalités le fait de croiser une dimension issue du HRT selon un niveau du SAT. Il n'existe pas à notre connaissance d'étude ayant exploré de façon croisée ces deux modèles. D'où nos deux questions de recherche (Figure 5.3) :

Q2a : Quels sont les effets des modalités de la transparence sur l'acceptation d'une proposition lors d'une prise de décision assistée par un CPS ?

Q2b : Quels sont les effets des modalités de la transparence sur les composantes de la confiance de l'opérateur dans la prise de décision assistée par un CPS ?

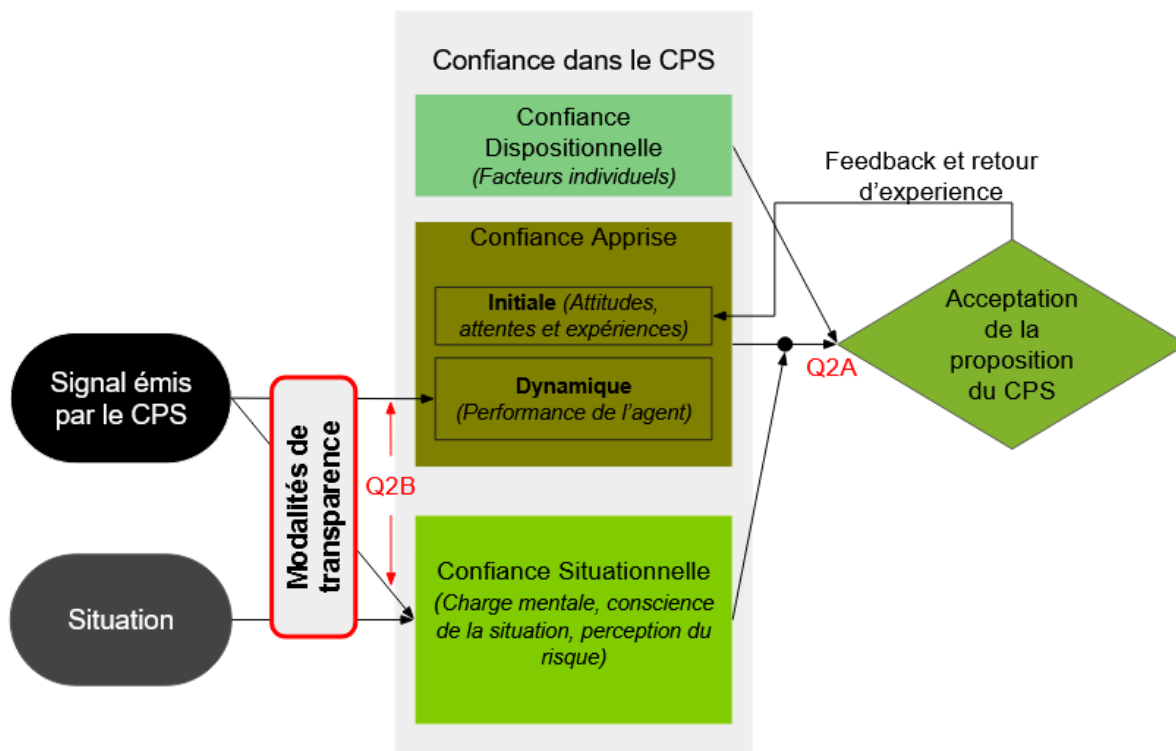


Figure 5.3 : Compréhension des modalités de transparence dans le cadre de la prise décision assistée par un CPS

Ces questions de recherche sont explorées dans le chapitre 6 : « *Confiance et transparence dans une prise de décision opérationnelle avec un CPS de fiabilité inconnue* », dans le chapitre 7 : « *Confiance et transparence dans une prise de décision tactique avec un CPS fiable* ».

5.4. Verrou 3 : Compréhension des effets d'une transparence adaptative

La question de l'adaptabilité des interfaces des CPS est cruciale pour répondre aux besoins des opérateurs qui peuvent évoluer au cours du temps (fatigue, manque de connaissance ou confiance inappropriée par exemple) (Steyvers & Kumar, 2022). Akash et al. (2020a ; 2020b) ont utilisé la transparence de façon dynamique avec un système basé sur la confiance et la charge mentale et ont obtenu des résultats encourageants. Une récente étude a également utilisé la transparence de façon dynamique en fonction d'une surutilisation ou d'une sous-utilisation d'un système de détection de nid de poule (Okamura & Yamada, 2020). Dans cette expérimentation, les chercheurs ont fait varier la transparence en utilisant un signal indiquant un potentiel mauvais choix de l'opérateur (Figure 5.4). D'où notre troisième et dernière question de recherche :

Q3 : La transparence peut-elle être utilisée de façon dynamique pour calibrer la confiance d'un opérateur ?

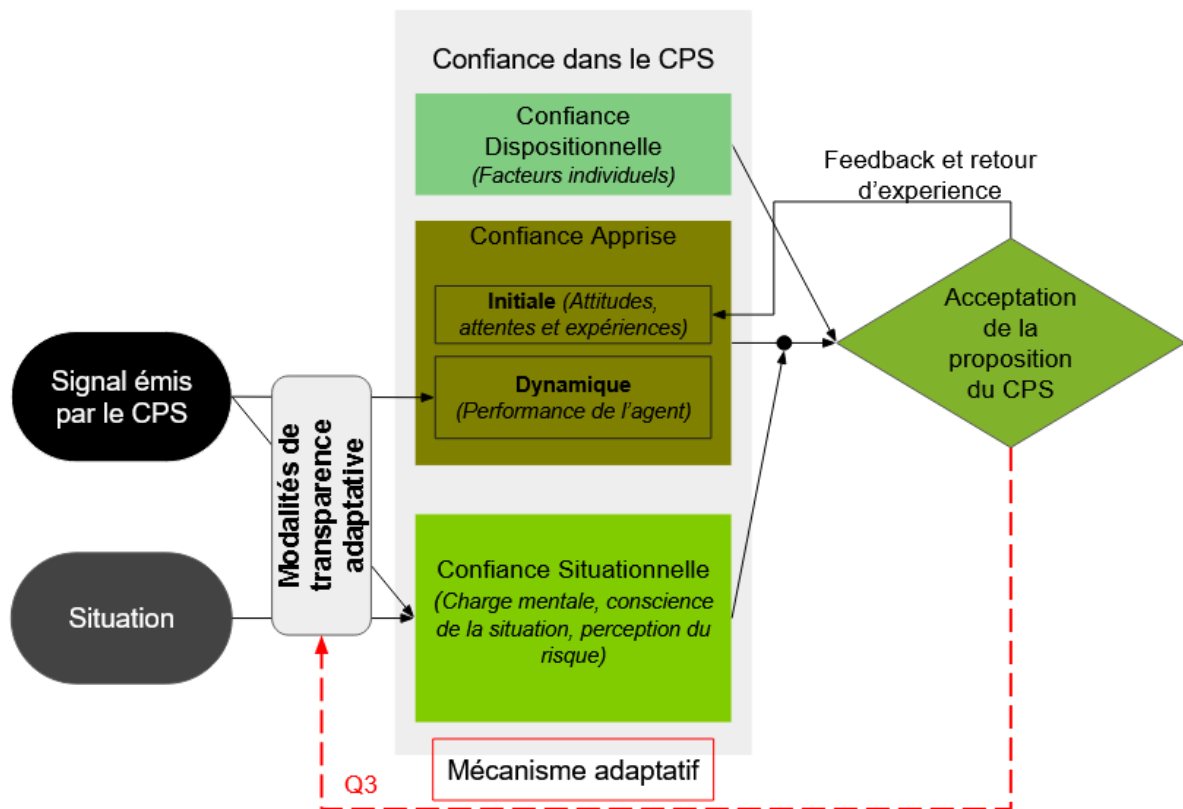
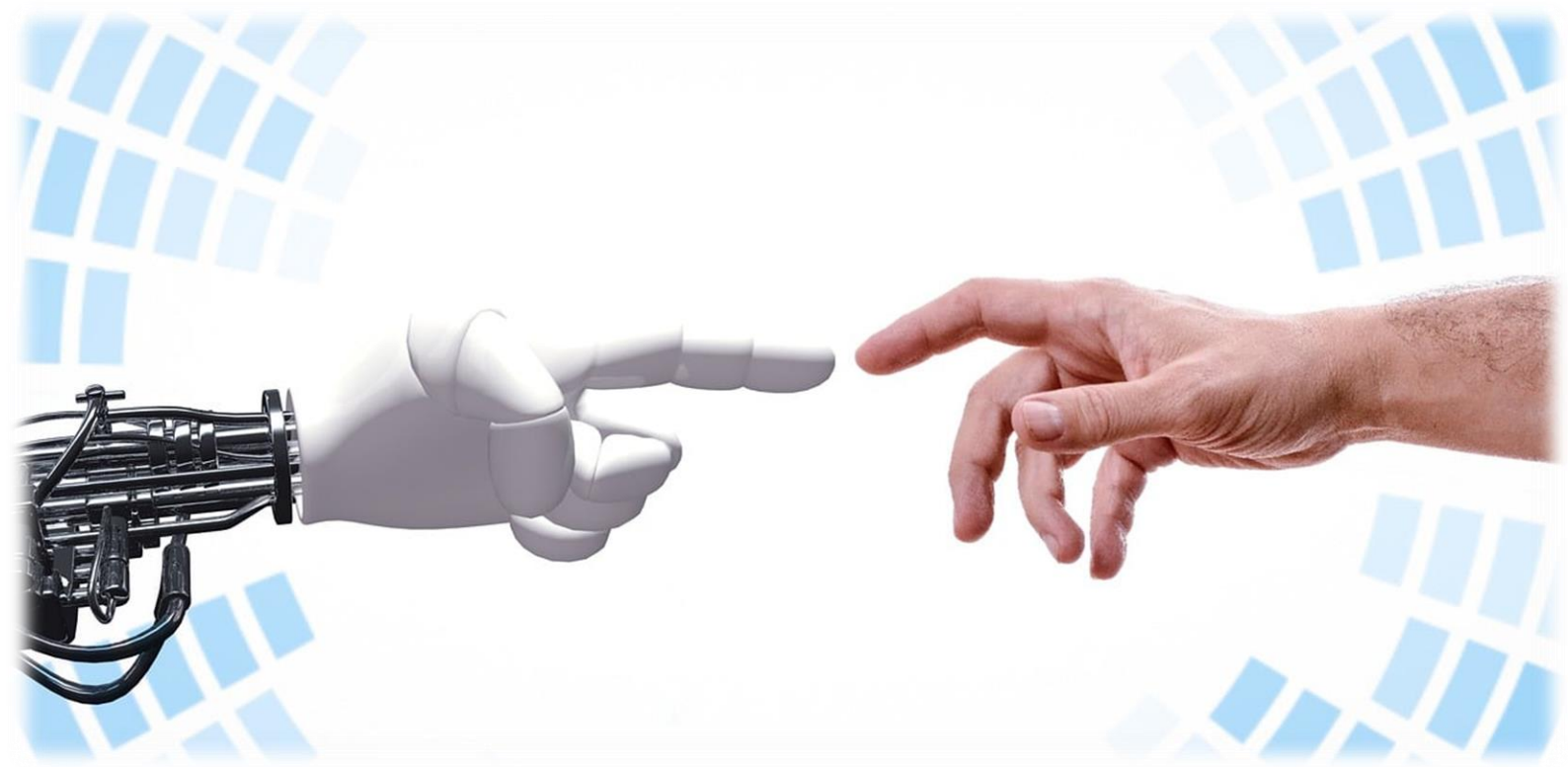


Figure 5.4 : Intégration d'un mécanisme adaptatif sur la transparence dans le cadre de la prise de décision assistée par un CPS

Cette question de recherche est explorée dans le chapitre 8 : « *Adaptation dynamique de la transparence pour calibrer la confiance* ».



Partie empirique

Composantes de la confiance et
modalités de transparence dans la
prise de décision assistée par un CPS

Les verrous scientifiques, et questions de recherches associées, que nous souhaitons investiguer dans cette thèse sont explorés au travers des deux projets présentés dans le chapitre 1 (partie 1.4). Dans le cadre de ces projets, trois expérimentations sont réalisées. Nous faisons ici le résumé des points communs et des différences qui existent entre ces trois expérimentations (Tableau II.1). Le point commun principal est que nous sommes toujours dans une situation de prise de décision assistée par un système cyber-physique (CPS). Ce CPS fait des propositions à l'opérateur et lui présentent plusieurs informations liées à ses propositions. L'opérateur doit alors d'accepter ou refuser cette proposition. Il y a deux principales différences. La première porte sur les modalités de transparence des CPS. La seconde porte sur la fiabilité des CPS. Pour rappel, nous appelons « modalité » de transparence le croisement entre les dimensions de la transparence (issu de Lyons, 2013) et leur niveau (issu de Chen et al., 2018).

Tableau II.1: Points communs et différences entre les projets Humanism et Seanatic

	Humanism	Seanatic (non-dynamique)	Seanatic (dynamique)
Questions de recherche abordées	QR1, QR2	QR1, QR2	QR3
Contexte et situation			
Contexte	Production en milieu industriel	Maintenance en milieu maritime	
Rôle de l'opérateur	Superviseur et opérateur	Superviseur	
Indépendance des situations	Situations indépendantes	Situations indépendantes	Situations dépendantes
Caractéristiques des agents			
Type de CPS	Cobot	Intelligence Artificielle	
Capacités du CPS	Myope	Modèle prédictif fiable	
Fiabilité du CPS	Inconnue	Haute	

	Humanism	Seanatic (non-dynamique)	Seanatic (dynamique)
Modalités d'interaction			
Niveau de l'interaction (Rauffet et al., 2018)	Tactique pendant la production	Tactique avant la maintenance	
Type d'interférence (Hoc, 2001)	Condition préalable, Bilatérale, Contrôle mutuel	Condition préalable, Bilatérale, Contrôle mutuel	
Niveau de l'interférence	Coopération dans l'action et opérationnelle	Coopération dans la planification et tactique	
Gestion de l'interférence (Hoc, 2001)	Acceptation ou imposition		
Type de coopération (Schmidt, 1991 ; Millot & Mandiaux, 1995)	Intégrative, débative et verticale		
Paradigme de présentation de la proposition (Steyvers & Kumar, 2022)	Concourante		
Modalités de transparence			
Modalité de transparence : Dimension (Lyons, 2013) * Niveau (Chen et al., 2018)	Dimension analytique (2 ou 3) Dimension sur l'état de l'opérateur (0 ou 1)	Dimension analytique (2 ou 3) Dimension sur l'environnement (0 ou 3)	
Adaptation dynamique de la transparence	Non-dynamique	Non-dynamique	Dynamique

Les travaux présentés dans ce manuscrit ont fait l'objet de plusieurs communications. Concernant le projet Humanism, nous avons publié un article à comité de lecture (ACL) publié dans une revue indexée (*PsycINFO, Medline, ISI, Scimago-Scopus Q1*), et portant sur nos résultats expérimentaux :

Simon, L., Guérin, C., Rauffet, P., Chauvin, C., & Martin, É. (2023a). How Humans Comply With a (Potentially) Faulty Robot: Effects of Multidimensional Transparency. *IEEE Transactions on Human-Machine Systems*, 1-10. <https://doi.org/10.1109/THMS.2023.3273773>

Dans le cadre du projet Seanatic, plusieurs communications orales dans des colloques nationaux et internationaux à comité de lecture ont été réalisées. Ces communications, qui ont fait l'objet de publications dans des actes, ont porté sur des résultats d'expérimentations et sur des aspects théoriques liés à la conception des interfaces écologiques.

- Communications sur les résultats expérimentaux :

Simon, L., Guerin, C., & Rauffet, P. (2023b, juillet 4). *Utilisation dynamique de la transparence pour la coopération humain-machine*. 12e Colloque EPIQUE Paris. <https://hal.science/hal-04111732>

Simon, L., Rauffet, P., Guérin, C., & Seguin, C. (2022b, juillet). Trust in an autonomous agent for predictive maintenance : How agent transparency could impact compliance. *13th AHFE Conference (Applied Human Factors and Ergonomics) 2022*. <https://doi.org/10.54941/ahfe1001602>

- Communications sur les ancrages théoriques et sur les méthodologies de conception :

Simon, L., Guérin, C., Rauffet, P., & Diguët, J.-P. (2022a). Integrating Transparency to Ecological Interface Design. *Proceedings of the 32nd European Safety and Reliability Conference (ESREL 2022)*, 3301-3308. <https://hal.archives-ouvertes.fr/hal-03601183>

Simon, L., Guérin, C., Rauffet, P., & Lassalle, J. (2021a). Using cognitive work analysis to develop predictive maintenance tool for vessels. *31st European Safety and Reliability Conference*.

Simon, L., Rauffet, P., Guérin, C., & Lassalle, J. (2021c). Exploiter la méthode CWA pour le design écologique d'une interface de maintenance 4.0. *Onzième colloque de Psychologie Ergonomique EPIQUE 2021*, 210.

Enfin plusieurs communications ont porté sur les deux projets de façon conjointe :

Rauffet, P., Simon, L., & Guerin, C. (2023, juin 14). *Transparence et Confiance au sein des équipes Humains Systèmes de l'Industrie 4.0*. CIGI QUALITA MOSIM 2023 - Propulser la performance. Interconnectivité et collaboration dans un contexte d'intelligence artificielle. <https://hal.science/hal-04111778>

Simon, L., Rauffet, P., & Guérin, C. (2021b). Développement de règles de conception et des mécanismes adaptatifs des interfaces dans le dialogue humain-machine- Application à l'industrie 4.0. *11e Colloque EPIQUE: Doctoriales 2021*.

Chapitre 6 : Confiance et transparence dans une prise de décision opérationnelle avec un CPS de fiabilité inconnue

Ce premier chapitre expérimental vise à explorer le rôle de la transparence dans la prise de décision de l'opérateur assistée par un système cyber-physique (CPS). Ceci correspond au **premier** et au **second verrous** identifiés. Le premier verrou porte sur la **compréhension du rôle des composantes de la confiance dans la prise de décision** (section 5.2) et le second porte sur **la compréhension des effets des modalités de transparence sur l'acceptation et la confiance** de l'opérateur (section 5.3).

Nous répondons à nos questions de recherche sous le prisme d'une prise de décision assistée par un CPS portant sur *un niveau opérationnel* (gestion d'une rupture de stock) et émanant d'un CPS *de fiabilité inconnue* et avec *un défaut d'acquisition des données* (myope). La myopie représente le manque d'information d'un système cyber-physique concernant la portée (entre un niveau local et un niveau global) et l'horizon temporel (entre le futur proche et le futur lointain) du processus supervisé et des autres entités du système (Adam, Gregor, Hevner & Morana, 2021).

Nous regardons en premier lieu le rôle de la confiance dispositionnelle (CD), de la confiance apprise initiale (CAI), de la confiance apprise dynamique (CAD) et de la confiance situationnelle (CS) dans la prise de décision de l'opérateur. Nous regardons également l'effet modérateur de la CS sur la CA dans l'acceptation de la proposition du CPS. Pour manipuler la transparence, nous utilisons un croisement des deux modèles couramment utilisés dans la littérature que sont le modèle du SAT (Chen et al., 2018) et le modèle de Lyons (2013) (section 3.1.1.3 « Articulation des modèles de transparence »). L'utilisation combinée de ces modèles est une proposition nouvelle et répond au besoin de mieux comprendre les effets de la transparence sur différents types d'informations (van de Merwe et al., 2022). Nous regardons en quoi ces modalités de transparence ont un effet sur l'acceptation des opérateurs, sur les métriques de confiance apprise initiale (CAI), de confiance apprise dynamique (CAD), sur la perception du danger et sur la charge mentale (liés à la confiance situationnelle).

Pour ce faire, les participants ont été placés dans une situation de supervision d'une ligne de production où un cobot leur faisait une proposition suite à un problème de stock de

ressources. Dans le cadre de cette expérimentation les participants avaient à accepter ou refuser la proposition d'un cobot. Nous mesurons la confiance du participant, sa perception du danger, sa performance à une tâche secondaire, son ressenti de sa charge mentale et son activité oculométrique.

6.1. Méthode

6.1.1. Participants

Cette étude a eu lieu entre décembre 2019 et janvier 2020 à l'ENSIBS de Lorient. La campagne de recrutement s'est déroulée durant trois semaines par l'envoi de mail, par la pose de poster et avec des interventions en début de cours. Les participants étaient invités à participer à une expérimentation sur la coopération humain-machine grâce à une simulation de ligne de production. Pour leur participation à l'étude, les participants étaient dédommagés avec un bon d'achat de 10 euros. La participation se faisait en présentiel. 53 participants issus de cursus ingénieur (Mécatronique et Génie Industriel) y ont participé. L'effectif est composé de 17 femmes et 36 hommes et la moyenne d'âge est de 21,34 ans ($E-T = 1,67$ ans). Aucun participant n'a de problème de vision impactant la lisibilité des couleurs (dyschromatopsie).

6.1.2. Protocole expérimental

Dans cette expérimentation, les participants ont pour objectif de superviser une ligne de production. Cette ligne est composée d'un cobot et d'un agent humain qui, tout comme le participant, remplissent des plateaux avec des cubes de couleur (Figure 6.1). Les participants sont amenés à prendre des décisions sur la gestion des ressources entre le cobot et le coéquipier humain. A chaque situation, le cobot émet une proposition pour transférer des ressources depuis le stock du coéquipier humain. Cette proposition émerge à la suite d'une prévision, par le cobot, d'une rupture de stock sur la ressource en question. Cette analyse est basée sur son propre stock et sur le stock de l'entrepôt. Cependant, il arrive que l'information sur le stock de l'entrepôt ne soit pas utilisée par le cobot pour calculer sa rupture de stock. Cela était dû au fait que le cobot peut être myope. Par exemple, le cobot peut formuler sa proposition au participant alors que les ressources en stock dans l'entrepôt sont suffisantes pour résoudre sa rupture. Les participants sont invités à décider s'ils acceptent ou

non la proposition du cobot (c'est-à-dire la nécessité de prélever des ressources auprès du coéquipier humain).

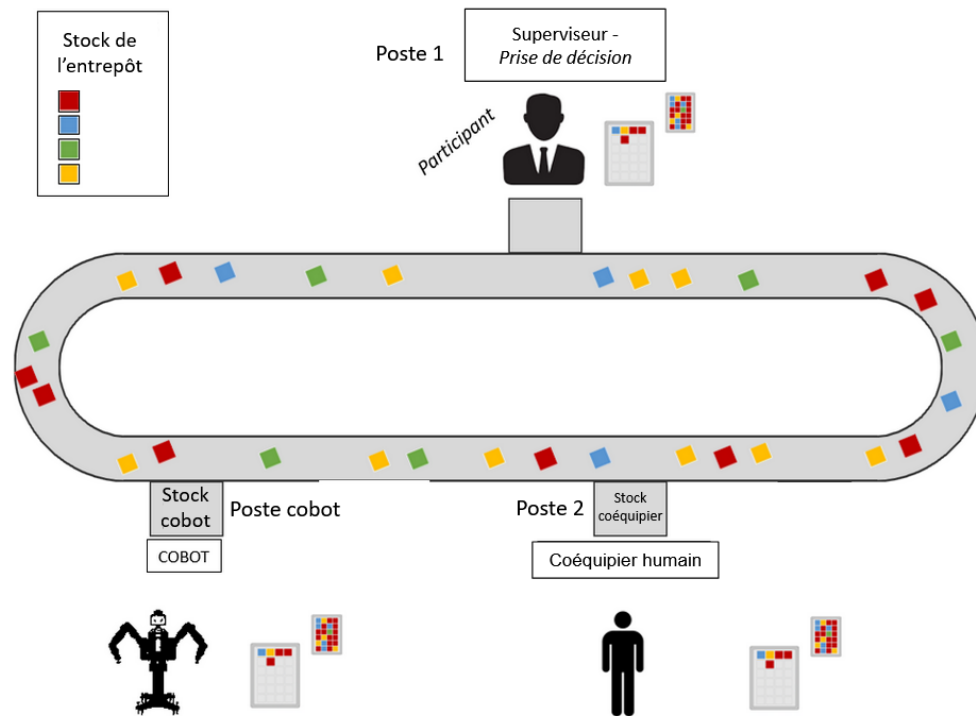


Figure 6.1 : Contexte expérimental et rôle des agents humains et du cobot

Accepter la demande implique un effet sur le stock du coéquipier dû au transfert. Pour répondre à cette tâche de décision, les participants sont placés dans deux modalités de transparence sur la dimension analytique (A) :

- Dans la première modalité (appelée par la suite A L2), le cobot fournit une demande et des informations de raisonnement (i.e. il envoie une alerte de rupture de stock), mais il n'est pas transparent sur la totalité de son raisonnement (i.e. il n'indique pas s'il a pris en compte le stock de l'entrepôt lors de sa demande). Ici le participant peut penser que le cobot a pris en compte le stock de l'entrepôt.
- Dans la seconde modalité (appelée par la suite A L3), les mêmes informations sont fournies mais le cobot précise qu'il n'a pas pris en compte le stock de l'entrepôt pour formuler sa proposition. Le cobot est transparent sur sa myopie. Ici, les participants sont certains que le cobot n'a pas pris en compte le stock de ressources présent dans l'entrepôt.

Pour étudier l'effet combiné de la transparence sur la dimension « analytique » (A) avec la transparence sur la dimension « état de l'opérateur » (EO), nous reproduisons les deux conditions précédentes en affichant des informations relatives aux capacités du coéquipier humain. Cette capacité est représentée par le taux de charge du coéquipier humain. Celui-ci peut être plus élevé ou plus faible que celui du cobot, ce qui peut impacter la capacité du coéquipier à préparer les commandes à son poste. Le cobot ne prend pas en compte cette information pour faire sa demande mais peut la détecter et la transmettre au participant. Ces informations supplémentaires ne sont pas liées à la tâche de décision mais peuvent être considérées par les participants pour estimer l'impact du cobot sur l'activité du coéquipier humain. La demande du cobot peut être correcte, mais, dû au transfert de ressources, les résultats de la proposition du cobot peuvent avoir un impact négatif sur le coéquipier humain. Cette information est, de ce fait, liée à la perception du danger pour le participant. Dans cette expérience, la modalité EO montre le taux de charge relatif. Ce facteur indique si le coéquipier humain est plus occupé (valeur supérieure à 100%) ou moins occupé (valeur inférieure à 100%) que le cobot. Les deux modalités de transparence A sont combinées avec trois modalités de transparence EO :

- La première modalité est une absence de transparence sur la dimension sur l'état de l'opérateur (désigné par la suite comme EO L0).
- La seconde modalité est la présentation d'information sur un taux de charge du coéquipier inférieur à 100%. Le cobot perçoit que le coéquipier humain est **moins occupé** par ses commandes que lui (appelé par la suite EO L1+).
- La troisième modalité est la présentation d'information sur un taux de charge du coéquipier supérieur à 100%. Le cobot perçoit que le coéquipier humain est **plus occupé** par ses commandes que lui (désigné plus tard par EO L1-).

La combinaison de ces modalités génère six conditions expérimentales (Tableau 6.1). Chaque participant est confronté deux fois aux six conditions expérimentales. De ce fait, chaque participant a à prendre douze décisions. Les situations sont randomisées pour éviter qu'un participant n'ait deux fois la même modalité de transparence « A » à la suite. Cela sert également à contrebalancer un éventuel effet d'ordre. Chaque décision est indépendante les unes des autres et n'a aucun impact sur le déroulement du scénario. Les participants sont

placés dans des situations à court terme, évitant d'établir des projections concernant la situation existante. Les participants n'ont pas de retour sur les conséquences de leurs décisions. En plus de cette activité de supervision, les participants ont une activité opérationnelle qui consiste à remplir des commandes selon un modèle qui leur est présenté.

Tableau 6.1 : Modalités de transparence dans les six conditions expérimentales

Conditions expérimentales	Modalités de transparence	
	Modalités de transparence Analytique	Modalités de transparence État Opérateur
	Niveaux (Chen et al., 2018) pour la dimension Analytique (Lyons, 2013)	Niveaux (adaptés de Chen et al., 2018) pour la dimension État de l'opérateur (Lyons, 2013)
1	L2	L0
2		L1+
3		L1-
4	L3	L0
5		L1+
6		L1-

Déroulé de la passation. Les participants étaient accueillis, et après avoir eu une présentation des objectifs de l'expérimentation, signaient un accord de consentement libre et éclairé (Annexe A). Ensuite, l'expérimentateur présentait le contexte industriel et le scénario expérimental à l'aide de supports visuels. Les participants étaient informés qu'ils étaient chargés de superviser la ligne de production. A la suite de cela, les modalités de transparence A étaient présentées aux participants, comme décrites précédemment. En particulier, pour A L3, les participants étaient informés que le cobot communiquait sur sa myopie due à une collecte d'informations incomplète. Il était donc précisé que le cobot était transparent sur le fait qu'il n'avait pas pris en compte le stock de l'entrepôt. Cette potentielle limitation pouvait entraîner une demande d'aide inadéquate de la part du cobot. Le participant ne pouvait pas en être sûr ni même le vérifier car il n'avait pas accès au niveau de stock de l'entrepôt. Pour finir, les modalités de transparence EO du cobot étaient expliquées.

Les participants étaient informés que cette information n'était pas utilisée par le cobot pour faire son diagnostic mais qu'il s'agissait d'une information supplémentaire communiquée par le cobot. Ces informations pouvaient aider à évaluer l'impact du transfert des ressources depuis le stock du coéquipier humain vers le stock du cobot :

- Pour la modalité EO L1+, il était signalé aux participants que les conditions étaient favorables à un transfert de ressources.
- Pour la modalité EO L1-, les participants étaient informés que les conditions pouvaient impacter négativement le stock du coéquipier humain.
- Pour la modalité EO L0, les participants n'avaient pas d'information sur cet éventuel impact sur le stock du coéquipier.

Les participants devaient reformuler leur compréhension des différentes modalités de transparence auprès de l'expérimentateur. Au début de l'expérimentation, les participants remplissaient un questionnaire démographique sur la familiarité des participants avec des jeux tels que le Poker ou le Blackjack (considérant que la décision était prise avec des informations manquantes) (annexe B.1). Suite à ces questionnaires les participants étaient confrontés à deux situations (et donc deux prises de décision) pour prendre en main le logiciel. Les douze situations faisaient suite à cette phase et chaque situation était indépendante l'une de l'autre. Une passation durait entre 45 minutes et 1 heure.

6.1.3. Matériel

Pour réaliser cette expérimentation les participants ont accès à un ordinateur. Les différentes échelles et questionnaires que les participants complètent se trouvent dans le logiciel « Humanism ». Le logiciel « Humanism » est composé de deux interfaces. La première interface est l'onglet de gestion de production et d'interaction avec le cobot (Figure 6.2). Depuis cet onglet les participants ont 90 secondes pour remplir leur commande, selon un modèle à disposition, ce qui fait office de tâche secondaire (Figure 6.2 à gauche). A droite de cet onglet se trouve l'encart de dialogue avec le cobot où celui-ci émet sa proposition (Figure 6.2 à droite). A partir de cet encart, les participants doivent décider s'ils sont en accord avec la demande du cobot (en cliquant sur « accepter » signifiant que le cobot prend des ressources dans le stock du coéquipier), ou s'ils sont en désaccord avec la demande du cobot

(en cliquant sur « refuser » ; ou en cliquant sur « alternative » signifiant que le cobot devrait prendre en compte la quantité de stock de l'entrepôt). Accepter la demande peut éventuellement résoudre le problème du cobot au prix d'un possible impact négatif sur le stock du coéquipier. Au contraire, refuser la demande n'aide pas à résoudre le problème du cobot mais permet au participant d'éviter d'impacter le coéquipier. Après leur décision, les participants n'ont pas de retour sur les conséquences de leur choix.

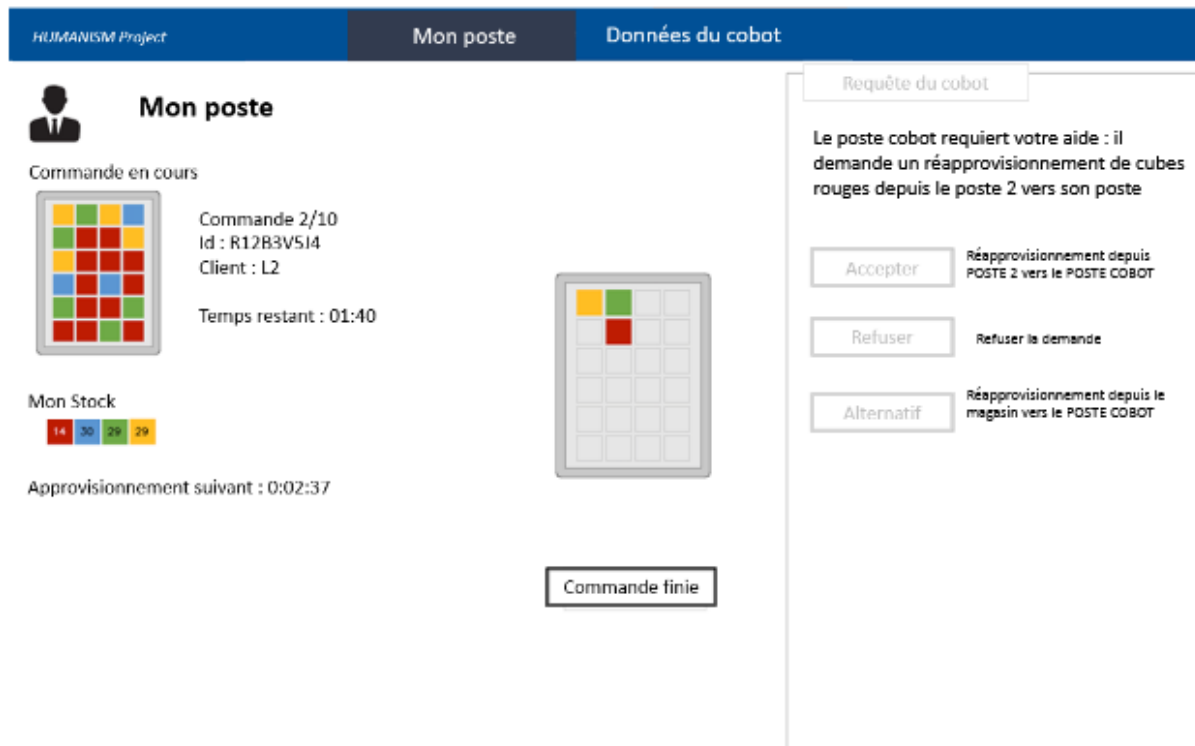


Figure 6.2 : Interface de complétion de tâche et de proposition

La deuxième interface du logiciel permet d'accéder aux informations transmises par le cobot (Figure 6.3). La partie haute de l'onglet représente les modalités de la transparence Analytique (A), où le cobot communique les informations sur son stock local (cf. risque de rupture) et sur le stock général de l'entrepôt (cf. stock et magasin). La partie basse de l'onglet représente les modalités de la transparence état de l'opérateur (EO) où le cobot communique des informations sur le taux de charge du coéquipier humain. Les informations fournies par le cobot au superviseur peuvent varier en fonction des niveaux de transparence de l'agent. La Figure 6.3, montre les différentes options d'affichage (cf. en pointillés bleus) concernant l'indicateur de stock de l'entrepôt (relatif aux modalités A) et l'indicateur taux de charge relatif (relatif aux modalités EO). Le carré noir représente le manque de transparence de l'agent sur :

- La collecte d'informations et sur la myopie du cobot pour la dimension A (si $A = L2$)
- Les ressources du coéquipier humain pour la dimension EO (si $EO = L0$).

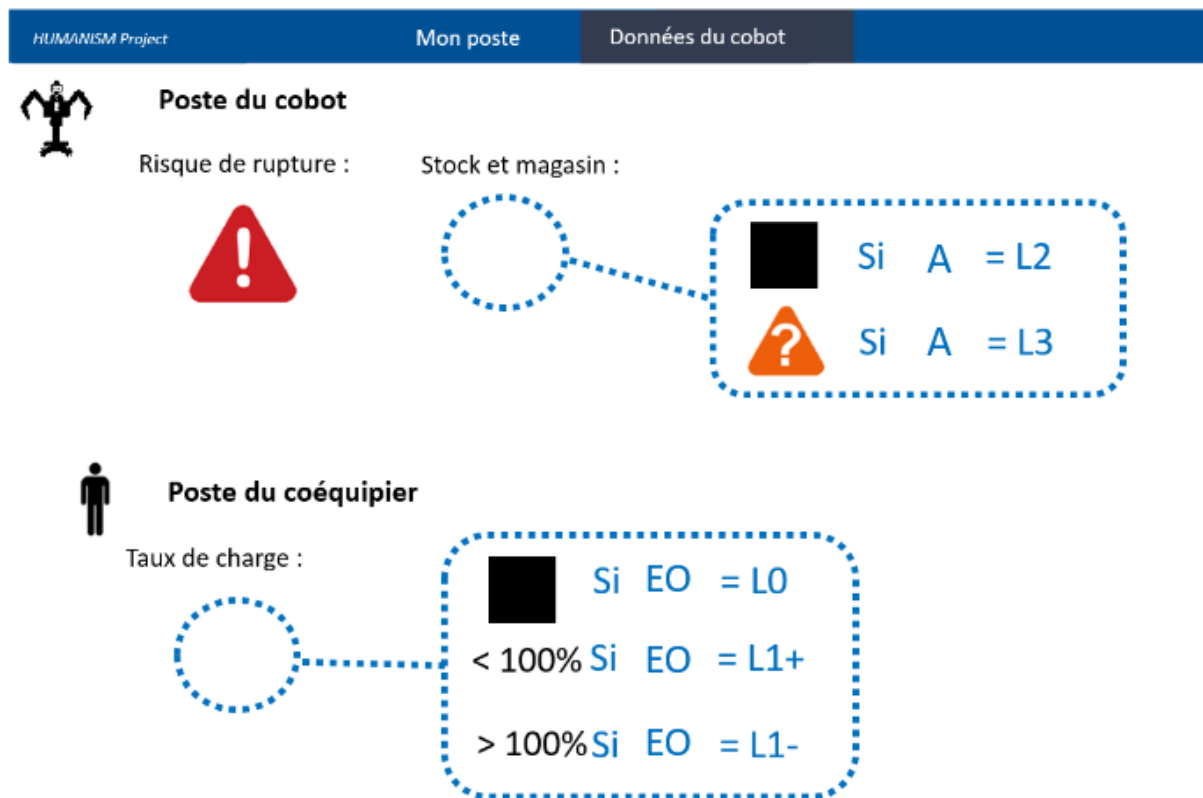


Figure 6.3 : Interface des données du cobot selon les niveaux de transparence. Le point d'exclamation indique un risque de rupture de stock communiqué par le cobot

6.1.4. Observables, méthode d'analyse des résultats, et hypothèses

Les différentes données collectées sont sauvegardées directement par le logiciel « Humanism » ce qui permet de les extraire pour les analyser par la suite. Nous avons mesuré deux variables de façon objective :

- **L'acceptation**, relative au choix du participant à la suite de la proposition du cobot : « Accepter » est codé comme « accepter la proposition », « Refuser » et « Alternatif » sont codés comme « refuser la proposition ». Le choix « alternatif » est codé comme un refus car cela montre que les participants ne sont pas en accord avec la proposition du cobot.
- **La confiance situationnelle (CS) est mesurée à partir de plusieurs indicateurs :**
 - La charge mentale grâce à l'eye-tracking et des mesures de performances (temps de complétion et taux d'erreur)

Nous avons mesuré six variables de façon subjective :

- La **confiance dispositionnelle (CD)** :
 - La propension au risque grâce à deux items créés pour les besoins de l'expérimentation (annexe B.2)
 - La tendance des participants à faire confiance grâce à un item créé pour les besoins de l'expérimentation (annexe B.3)
- La **confiance apprise** :
 - **Initiale (CAI)** grâce à deux items créés pour les besoins de l'expérimentation (Annexe B.4).
 - **Dynamique (CAD)** : grâce à un item créé pour les besoins de l'expérimentation (Annexe B.5).
- La **confiance situationnelle (CS)** :
 - La perception du danger grâce un item créé pour les besoins de l'expérimentation (Annexe B.6).
 - La charge mentale grâce à l'échelle d'Instantaneous Self-Assessment (Tattersall & Foord, 1996) (Annexe B.7)

Pour analyser les résultats nous utilisons R (R Core Team, 2021), et notamment le package lme4 régression logistique ordinaire (fonction polr pour étudier les effets sur des variables ordinales comme la confiance ou l'ISA), et des analyses linéaires des effets mixtes (fonction lmer pour étudier les effets sur des variables continues, comme le temps de décision, la réalisation de la tâche) (Bates, Maechler & Bokler, 2012). Pour l'analyse des effets linéaires, l'inspection visuelle des graphiques de résidus ne révèle pas de déviations évidentes de l'homoscédasticité ou de la normalité. Pour tous les modèles statistiques, nous introduisons les modalités A (transparence analytique) et les modalités EO (transparence état de l'opérateur) comme effet fixe (avec des termes d'interaction) dans le modèle complet. Comme effet aléatoire, nous avons un intercept pour les participants. En ce qui concerne les effets fixes, une sélection de modèles par étapes et par AIC (stepAIC) est effectuée. A chaque étape, un nouveau modèle est ajusté, dans lequel un des termes du modèle est éliminé et testé par rapport à l'ancien modèle.

Nous répondons aux questions de nos **premier** et **deuxième verrous** au travers des hypothèses exploratoires suivantes (Figure 6.4) :

H1. Les composantes de la confiance seraient corrélées avec l'acceptation (Hoff & Bashir, 2015). Par ailleurs, nous considérons qu'il existerait une interaction entre la confiance apprise et la confiance situationnelle (Chancey et al., 2017) (**Verrou 1**).

H2. Les modalités de transparence de la **dimension analytique** et de la **dimension de l'état opérateur** d'un **cobot à fiabilité inconnue** seraient corrélées avec l'acceptation d'une proposition (Sadler et al., 2016 ; Selkowitz et al., 2016 ; Stower et al., 2016 ; Bhaskara et al., 2021 ; Pynadath, Gurney & Wang, 2022) (**Verrou 2**). Comme suggéré par Monsaigneon, et al., (2020) et Zhang et al. (2022), on s'attendrait à ce que cette fiabilité soit modérée par la dimension de l'état opérateur.

H3. Les modalités de transparence de la **dimension analytique** et de la dimension de l'état **opérateur** d'un **cobot à fiabilité inconnue** seraient corrélées avec les composantes de la confiance de l'opérateur (Selkowitz et al., 2016 ; Mercado et al., 2016) (**Verrou 2**).

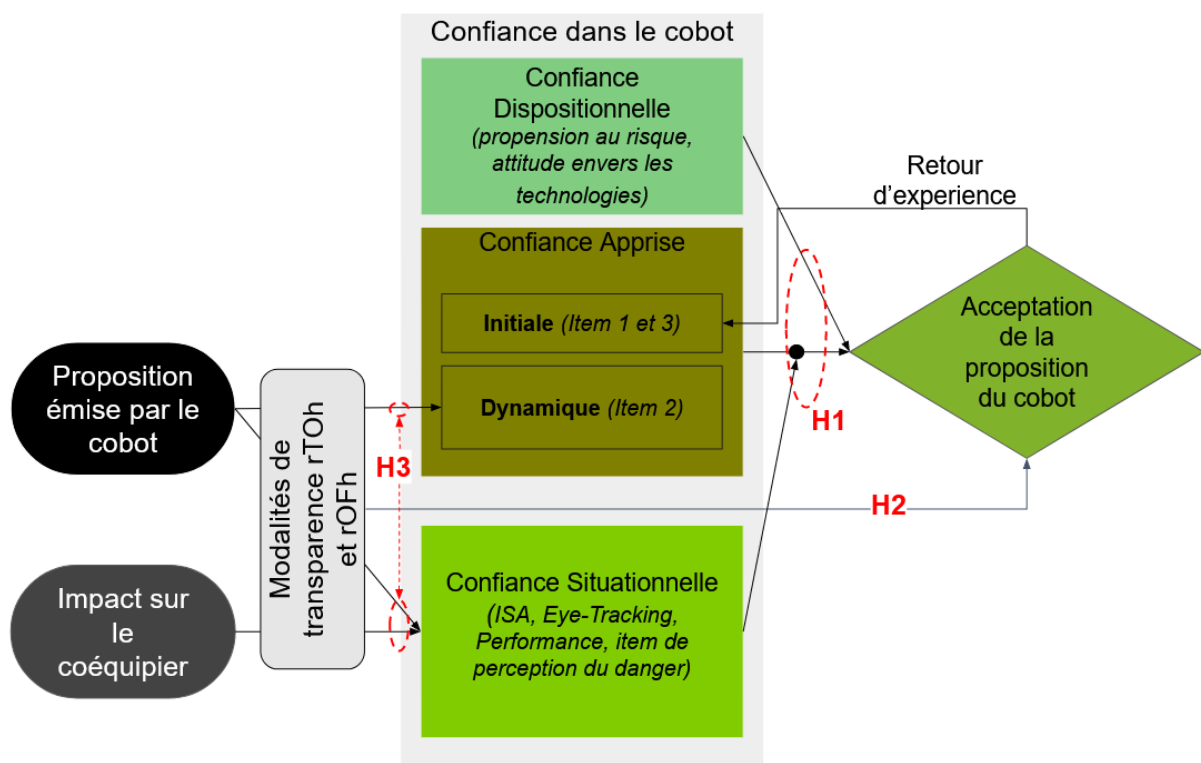


Figure 6.4 : Positionnement des questions de recherche dans le cadre de la prise de décision assistée par un cobot

6.2. Résultats

6.2.1. Relations entre les composantes de la confiance et l'acceptation

Relations entre la confiance dispositionnelle et l'acceptation. Les analyses de Mann-Whitney montrent que **la confiance dispositionnelle n'est pas corrélée avec l'acceptation**. Les scores des participants n'ont montré aucun lien significatif entre les questions démographiques et l'acceptation des participants (respectivement affinité envers les technologies $Mdn = 2.97$, $Mdn = 2.46$, $W=37152$, $p = 0.75$, propension à la confiance $Mdn = 2.28$, $Mdn = 1.8$, $W=37482$, $p = 0.623$ et propension au risque $Mdn = 2.19$, $Mdn = 1.6$, $W=34559$, $p = 0.287$).

Relations entre la confiance apprise et l'acceptation. Les analyses de Mann-Whitney ont révélé, et ce pour les trois items, que **la confiance apprise (dynamique et initiale) est corrélée positivement avec l'acceptation**. Les participants qui acceptent la proposition du cobot rapportent des scores significativement plus élevés que ceux qui refusent sur la compréhensibilité du cobot (CAI) (respectivement $Mdn = 5.05$ et $Mdn = 4.95$, $W=50846$, $p = .008$), s'il est digne de confiance (CAI) (respectivement $Mdn = 4.97$, $Mdn = 4.25$, $W=55450$, $p < .001$) et sa fiabilité (CAD) (respectivement $Mdn = 4.97$, $Mdn = 4.49$, $W=51383$, $p = .004$).

Relations entre la confiance situationnelle (perception du danger) et l'acceptation. De même, il existe une association significative entre la perception du danger et l'acceptation. **La confiance situationnelle est corrélée négativement avec l'acceptation**. Les participants qui acceptent la proposition du cobot rapporte une perception du danger significativement plus faible ($Mdn = 4$) que ceux qui refusent ($Mdn = 2$) ($W=23378$, $p = .038$).

Relations entre la confiance apprise, la confiance situationnelle (perception du danger) et l'acceptation. Nous ne constatons pas d'interaction entre **la confiance situationnelle et la confiance apprise sur l'acceptation**.

Ces résultats soutiennent en partie H1.

6.2.2. Relations entre les modalités de transparence du cobot et l'acceptation

Nous étudions d'abord l'effet des modalités de transparence analytique (A) et état de l'opérateur (EO) sur l'acceptation de la proposition de l'IA par le participant. Nous effectuons une régression logistique qui révèle des effets principaux de A et de EO (Tableau 6.2).

Tableau 6.2 : Régression logistique, effet des modalités A et EO sur l'acceptation

Modalités	OR ¹	95%IC ²	<i>p-value</i>
A (ref L2)			
L3	0.14	-2.75, -1.24	<.001***
EO (ref L1-)			
L0	8.47	1.46, 2.82	<.001***
L1+	26.34	2.30, 4.25	<.001***
A * EO (ref L2 * L1-)			
L3*L0	2.5	-0.11, 1.94	.08
L3*L1+	7.29	0.51, 3.46	.008**

¹OR = Odd Ratio, ²IC = intervalle de confiance

Relations entre les modalités de transparence de la dimension analytique et l'acceptation. La transparence analytique (A) a un effet négatif sur l'acceptation. L'acceptation est plus faible lorsque le cobot est plus transparent et montre que sa proposition est basée sur une observation partielle de la situation (A L3, i.e. le cobot est myope). Comme nous pouvons le voir dans la Figure 6.5, la probabilité d'acceptation est significativement plus faible lorsque la modalité de transparence du cobot est de A L3 par rapport à A L2 ($OR = 0.14$, $p < .001$). Ce résultat soutient l'hypothèse H2.

Relations entre les modalités de transparence de la dimension opérateur et l'acceptation. La transparence sur l'état de l'opérateur (EO), avec un impact minimal sur le stock du coéquipier, a un effet positif sur l'acceptation. Nous pouvons observer que la probabilité d'acceptation est significativement plus élevée lorsque le cobot communique un impact

minimal sur le stock du coéquipier humain (EO L1+) par rapport à un impact négatif potentiel sur le coéquipier humain (EO L1-) ($OR = 26.34, p < .001$) (Figure 6.5). La probabilité d'acceptation est significativement plus élevée lorsqu'il ne communique pas sur cet impact (EO L0) par rapport à EO L1- ($OR = 8.47, p < .001$). Ce résultat soutient l'hypothèse H2.

Relations entre les modalités de transparence de la dimension état de l'opérateur, de la dimension analytique et l'acceptation. La transparence sur l'état de l'opérateur (EO) est un modérateur du lien entre la transparence analytique (A) et l'acceptation. Nous constatons qu'il existe une interaction significative augmentant la probabilité d'acceptation avec EO L1+ et A L3 par rapport à la référence (EO L1- et A L2) ($OR = 7.29, p < .01$). Par ailleurs, si le résultat n'est pas significatif, nous observons que la probabilité d'acceptation a tendance à augmenter pour l'interaction EO L0 et A L3 par rapport à la référence ($OR = 2.5, p = 0.08$). Ces effets sont soulignés dans la Figure 6.5.

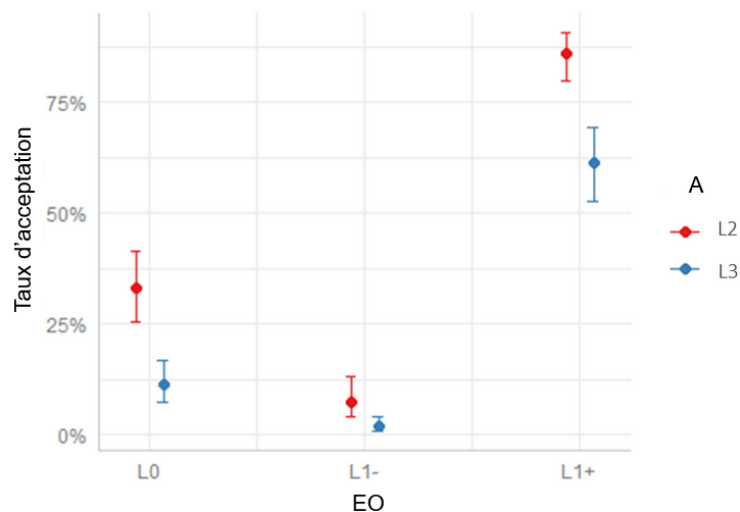


Figure 6.5 : Probabilité d'acceptation de la proposition du cobot en fonction des modalités de transparence

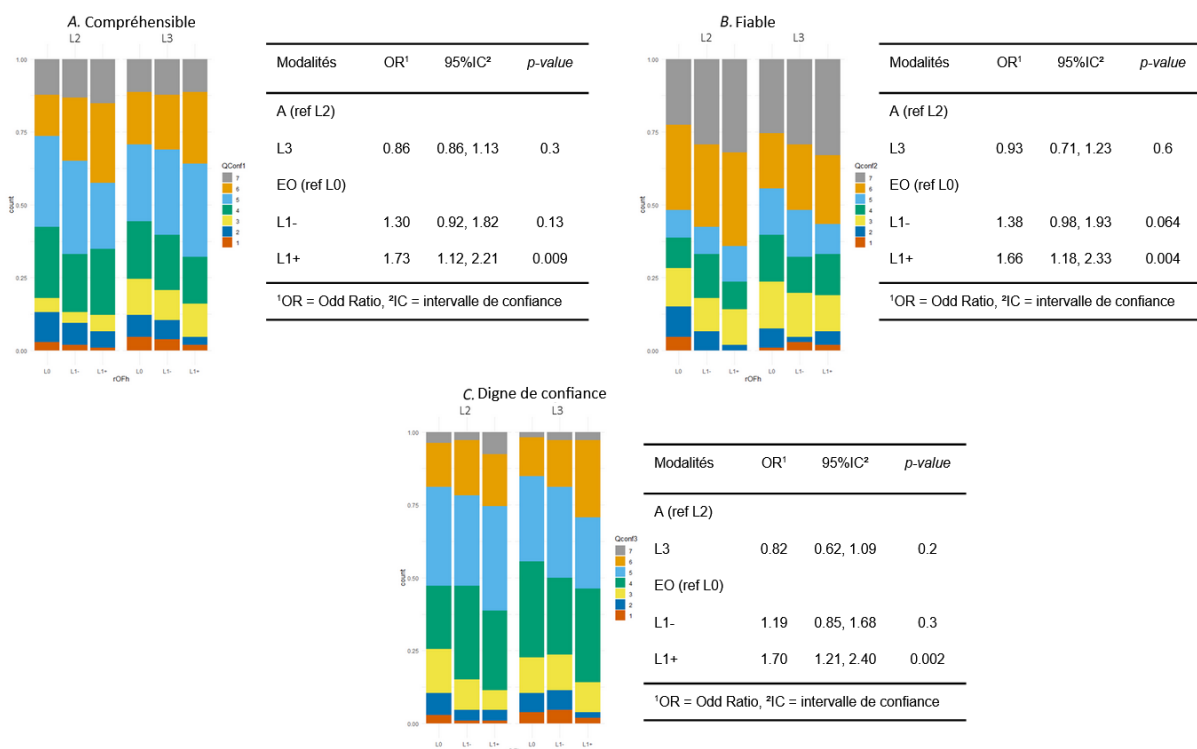
La probabilité d'acceptation de la proposition est toujours plus faible en A L3 qu'en A L2, quelle que soit la modalité de transparence EO associée. De plus, nous pouvons observer sur cette figure que la transparence EO a un effet important sur les variations de la probabilité d'acceptation en modérant l'effet de la transparence A. Lorsqu'aucune information n'est présentée sur le taux de charge du coéquipier humain (EO L0), le modèle prédit que les participants seront moins enclins à accepter avec A L2 (34%), et encore moins avec A L3, qui indiquent une observabilité partielle de l'agent pour sa demande (12%). Mais, en ajoutant des informations relatives au coéquipier, ces tendances à l'acceptation ou au refus changent

légèrement. Avec une situation positive indiquant un impact minimale sur l'autre coéquipier (EO L1+), le modèle prédit que les participants accepteront plus (86% avec A L2, 62% avec A L3). Au contraire, le modèle prédit très peu d'acceptation de la proposition du cobot (6% avec A L2, 1% avec A L3) lorsque l'impact sur le coéquipier humain est négatif (EO L1-). Ces résultats soutiennent en partie l'hypothèse H1.

6.2.3. Relations entre les modalités de transparence du cobot et les composantes de la confiance

Relations entre les modalités de transparence et la confiance apprise. En effectuant une régression logistique ordinaire sur les différents items du questionnaire de confiance, nous observons également un effet significatif des modalités EO (Tableau 6.3). **La transparence sur l'état de l'opérateur (EO), avec un impact minimal sur le stock du coéquipier, a un effet positif sur la confiance apprise.**

Tableau 6.3 : Régression logistique ordinaire, effet des modalités de transparence sur la confiance apprise (A. Compréhensible, B. Fiable, C. Digne de confiance)



Les participants trouvent l'agent plus compréhensible (CAI) ($OR = 1.57, p = .009$), plus fiable (CAD) ($OR = 1.66, p = .004$) et plus digne de confiance (CAI) ($OR = 1.70, p = .002$) lorsque la dimension EO est au niveau L1+ (c'est-à-dire avec un impact minimal sur le coéquipier) par

rapport à L0. A l'inverse, nous n'observons pas d'effet significatif des modalités de A ou de l'effet de leur interaction. Ce résultat ne permet pas de soutenir H3.

Relations entre les modalités de transparence et la confiance situationnelle (perception du danger). A l'aide d'une régression logistique ordinale, nous analysons l'effet de la transparence sur l'item destiné à estimer le niveau de perception du danger (Figure 6.6) et nous observons notamment un effet significatif des modalités EO. **La transparence sur l'état de l'opérateur (EO), avec un impact minimal sur le stock du coéquipier, a un effet négatif sur la confiance situationnelle.** Nous observons notamment un effet significatif des modalités EO (Figure 6.6 et Tableau 6.4).

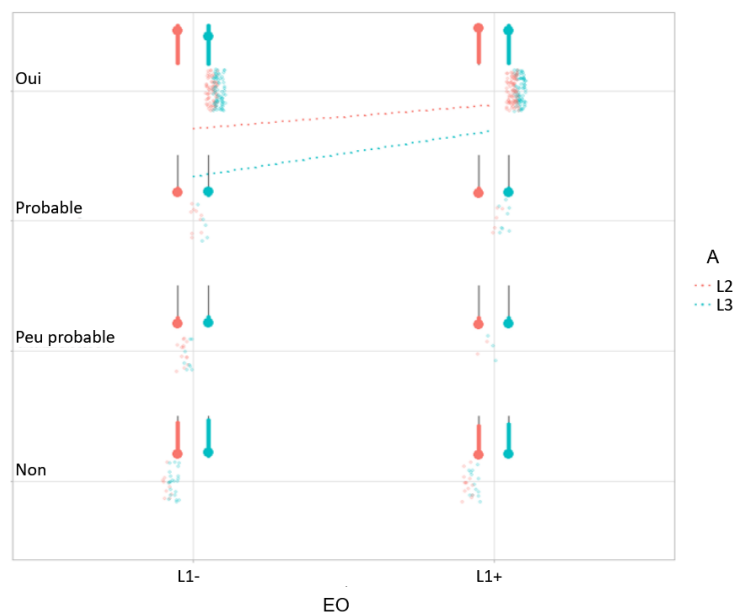


Figure 6.6 : Effet de la transparence sur les modalités de réponses aux questionnaires de confiance

Tableau 6.4 : Régression logistique ordinale, effet des modalités A et EO sur la perception du danger

Modalités	OR ¹	95%IC ²	p-value
A (ref L2)			
L3	0.79	0.51, 1.22	0.3
EO (ref L1-)			
L1+	1.73	1.12, 2.69	0.014*

¹OR = Odd Ratio, ²IC = intervalle de confiance

Le danger est perçu comme plus élevé et plus certain par les participants dans la modalité EO L1- que dans la modalité EO L1+ ($OR = 1.73, p = .014$), supportant en partie H3 et validant le lien que nous faisons entre EO et perception du danger.

Relations entre les modalités de transparence et la confiance situationnelle (charge mentale). Concernant les métriques de charge mentale, les modalités de la transparence A et de la transparence EO n'ont pas montré d'effet significatif sur la charge mentale et sur l'eye-tracking. **La transparence n'a pas d'effet sur la charge mentale.**

La transparence état de l'opérateur, avec un impact minimal sur le coéquipier, a un effet positif sur la performance. Le Tableau 6.5 montre une corrélation négative significative de EO L1+ sur le temps utilisé pour prendre une décision ; le temps nécessaire est réduit de - 3,92 secondes.

Tableau 6.5 : Modèle linéaire mixte de la performance des modalités de transparence sur la tâche secondaire et le temps avant la prise de décision

	Temps entre l'analyse des indicateurs et la décision ~ A + EO	Temps total ~ A	Taux de complétion ~ A
Intercept	15.22(1.22)***	52.33(1.11)***	99.90(0.89)***
A (ref L2)			
L3	4.27 (0.90)***	0.534 (0.82)***	99.90 (0.89)***
EO (ref L0)			
L1+	-3.92 (1.10)***		
L1-	-2.00 (1.10)		
AIC	5593.71	5716.97	4679.76
BIC	5621.06	5735.36	4697.57
Log vraisemblance	-2790.85	-2854.49	-2335.88

La transparence analytique (A) a un effet négatif sur la performance. Pour A L3, le temps total a significativement augmenté (+ 5,34 secondes) et plus particulièrement pour le

temps entre l'analyse des indicateurs et la prise de décision : + 4,27 secondes. Il y a une réduction significative de la performance sur le taux de remplissage de la tâche secondaire opérationnelle pour la A L3. Pour la modalité de transparence A L2, les participants ont un taux de remplissage de 99,9%, ce qui est supérieur à ce qui est observé pour A L3 avec un taux de remplissage de 98,2%. Nous n'observons pas d'effet significatif de leur interaction.

6.3. Discussion

Cette première expérimentation a permis de montrer que l'acceptation est corrélée avec la confiance situationnelle et la confiance apprise. Les résultats indiquent que les modalités transparence ont un effet sur l'acceptation. Les résultats indiquent également que la transparence a un effet sur la confiance apprise (dynamique et initiale) et sur la confiance situationnelle (*perception du danger*). Nous ne constatons pas d'effet de la transparence sur la charge mentale physiologie ou subjective, mais nous constatons un effet indirect sur celle-ci via les performances du participant. Également, nous montrons qu'il est possible d'utiliser conjointement deux modèles de transparence.

Relations entre les composantes de la confiance et l'acceptation. Contrairement à l'attendu (Hoff & Bashir, 2015) **nous n'avons pas constaté d'effet significatif de la confiance dispositionnelle sur l'acceptation du participant.** Nous évoquons deux possibilités pour expliquer ce résultat. Premièrement, les questionnaires que nous avons utilisés ne sont pas assez sensibles pour différencier les participants et ne mesurent pas les facteurs individuels de la confiance dispositionnelle avec précision. Secondement, la confiance dispositionnelle ne joue peut-être qu'un rôle mineur dans l'explicabilité de la prise de décision, celle-ci étant plus explicable par la confiance apprise et la confiance situationnelle.

Nous constatons que **la confiance apprise et la confiance situationnelle sont corrélées avec l'acceptation.** Nous retrouvons ici un résultat conforme au modèle proposé par Hoff & Bashir (2015). La confiance est bien structurée en plusieurs composantes. Les deux composantes identifiées ici sont en accord avec le modèle proposé par Chancey et al. (2017) dans le sens où c'est la perception du danger basse, de la confiance situationnelle, et une confiance apprise haute qui sont corrélées avec plus d'acceptation. Ces deux composantes semblent être des déterminants de la prise de décision de l'opérateur.

Si Chancey et ses collègues n'avaient pas réussi à soutenir leur hypothèse selon laquelle le risque venait modérer la confiance (« *It is unclear from our results whether risk moderated the mediating effect of trust on compliance* », p. 342). Nos résultats ont apporté des preuves plus solides, en montrant que l'ajout d'informations sur le risque d'impacter un autre coéquipier (avec la dimension état de l'opérateur) modère le lien entre la confiance et l'acceptation d'une proposition provenant d'un cobot. Par rapport aux situations où le cobot ne communique aucun élément sur le risque, lorsqu'il communique que sa proposition n'a pas d'impact sur le coéquipier, le taux d'acceptation des participants augmentent. A l'inverse lorsque le cobot communique que sa proposition a un impact négatif sur le coéquipier, alors le taux d'acceptation des participants est plus bas. **Nous constatons que c'est l'ajout de l'information sur le risque qui vient modérer l'effet de la transparence de la dimension analytique sur l'acceptation.** Si le cobot est myope, mais qu'il communique que sa proposition n'aura pas d'impact sur le coéquipier, alors le participant fait confiance. Si nous n'avons pas pu constater directement d'effet modérateur de la confiance situationnelle sur la confiance apprise, ce résultat alimente notre compréhension du lien entre ces deux composantes et l'acceptation.

Notre première hypothèse est en partie validée et nos résultats permettent d'éclairer notre premier verrou, en améliorant notre compréhension du rôle des composantes dans la prise de décision de l'opérateur assistée par un système cyber-physique (CPS). **Nous constatons que parmi les trois composantes de la confiance proposées par Hoff & Bashir, (2015) deux d'entre elles sont des déterminants de l'acceptation.**

Relations entre les modalités de la transparence et l'acceptation. Nos résultats ont montré que l'acceptation des participants, lorsque l'agent raisonne avec des informations partielles et limitées, change de manière significative selon la transparence du cobot. **Une plus grande transparence du cobot sur la dimension analytique (L3) est liée à une plus faible acceptation** de sa proposition en comparaison d'un cobot cachant sa myopie au participant (de transparence analytique L2). Nous retrouvons ici un effet connu dans le milieu des voitures autonomes (Monsaingeon et al., 2020 ; Zang et al., 2022) où une fiabilité du cobot plus faible est associée à moins d'acceptation des propositions. Une plus grande transparence

sur la myopie du cobot permet de calibrer la confiance du participant en montrant indirectement, que le cobot n'était pas totalement fiable.

De plus, **l'acceptation varie radicalement lorsque le cobot communique sur les informations sur l'état de l'opérateur.** La transparence sur un impact minimal sur coéquipier (comme dans la condition état de l'opérateur L1+) augmente le taux d'acceptation par rapport à une condition sans transparence (pas d'information sur l'état opérateur). A l'inverse, la transparence sur un impact négatif sur le stock du coéquipier (état de l'opérateur L1-) a diminué l'acceptation par rapport à une absence de communication sur l'état opérateur.

L'interaction entre les informations de la dimensions analytique et de la dimension état de l'opérateur est significative lorsqu'une situation risquée est communiquée par le cobot (état de l'opérateur L1-). Ce résultat est conforme au modèle de Chancey et al. (2017) selon lequel le risque modère la confiance. En revanche, nos résultats sont en contradiction avec ceux de Wang et al. (2016) qui n'ont pas trouvé de corrélation entre la transparence et l'acceptation. Cette différence de résultat peut être expliquée par les modalités de transparence utilisées entre les deux études. Dans l'étude de Wang et ses collègues (2016), la transparence du cobot ne concerne que le raisonnement et l'explication de la décision du cobot. Dans notre expérience la transparence concernait également les potentiels effets irrévocables et nuisibles d'une acceptation de la proposition sur le coéquipier (son stock et par conséquent son activité de production). Ainsi, comme suggéré par Salem et al. (2015), lorsque la proposition est suivie d'effets irrévocables et négatifs, les participants ont moins tendance à vouloir suivre les propositions du cobot.

Nous pouvons ici valider notre seconde hypothèse et ainsi apporter des éléments pour améliorer notre compréhension du lien entre transparence et acceptation et aider à déverrouiller notre second verrou. **La transparence du CPS a un effet sur l'acceptation des opérateurs.** Les modalités utilisées par le CPS pour communiquer n'auront pas toutes les mêmes effets sur l'acceptation. Certaines modalités amenant l'opérateur à privilégier une acceptation de la proposition, et d'autres modalités amenant l'opérateur à privilégier le refus de la proposition.

Relations entre les modalités de transparence et les composantes de la confiance. Le fait que la transparence sur l'état de l'opérateur joue un rôle sur l'acceptation de la proposition par le participant est à mettre en lien avec son effet sur la confiance apprise. La modalité de transparence de l'état de l'opérateur (L1+) a également entraîné une plus grande confiance apprise des participants dans le cobot en comparaison d'une absence de transparence sur cette dimension. La confiance apprise (initiale et dynamique) est plus élevée dans cette modalité. **Lorsque le cobot fait une proposition dans un contexte où il n'y a pas de danger, la confiance apprise du participant est impactée positivement.** Concernant la dimension analytique, nous n'avons pas observé d'effet sur la confiance apprise. Ceci pourrait être expliqué par le fait que, certes, le cobot est plus transparent mais c'est sur une collecte incomplète d'informations (en d'autres termes sa myopie). Le participant a dès lors une meilleure connaissance des limites de compétence du cobot, ce qui lui permet de calibrer la confiance. Cette calibration prend la forme ici d'une confiance apprise qui reste stable.

La transparence a également un effet sur la confiance situationnelle. En effet, nous avons pu observer que lorsque le cobot était transparent sur un impact sans conséquence négatives (état de l'opérateur L1+), le facteur de perception du danger de la confiance situationnelle était plus faible que dans une situation où l'impact était négatif (état de l'opérateur L1-). Ce résultat est conforme à la littérature où la transparence peut servir à améliorer la conscience de la situation des participants (Chen et al., 2014). Conformément à l'attendu, la transparence n'a pas impacté le facteur de la charge mentale, d'un point de vue physiologique ou subjectif, de la confiance situationnelle (van de Merwe et al., 2022). En revanche, la transparence sur l'état de l'opérateur, lorsqu'il n'y a pas d'impact sur le stock du coéquipier, améliore les performances des participants sur la tâche. La dimension sur l'état de l'opérateur, pour des informations positives, allège la charge mentale que l'opérateur alloue à sa prise de décision et peut l'utiliser pour l'accomplissement d'autres tâches. A l'inverse, lorsque le CPS est transparent sur la dimension analytique, il y a une baisse de la performance du participant. La transparence sur la dimension analytique augmente la charge mentale en complexifiant la prise de décision avec l'incertitude liée à la myopie. Ce résultat est à rapprocher du constat de Chen (2021) sur l'inconsistance des résultats entre la transparence et la charge mentale. **Selon les informations transmises par le CPS, la transparence n'aura pas le même effet sur la charge mentale.**

Nous ne validons qu'en partie notre troisième hypothèse, mais cela permet de mieux comprendre le lien entre la transparence et les composantes de la confiance, tel qu'exprimé dans notre second verrou. De la même façon que les modalités de transparence n'ont pas le même effet sur l'acceptation, ces modalités n'auront pas le même effet sur les composantes. Nous constatons notamment que c'est la transparence sur l'état de l'opérateur qui a le plus d'effet sur la confiance des opérateurs.

Limites. L'une des limites de cette expérimentation est liée au fait que les participants étaient des novices. Des participants avec une expertise dans la prise de décision assistée par un cobot de niveau opérationnel pourraient modifier la façon de coopérer avec lui. A travers le paradigme expert/novice, il serait intéressant de considérer les populations familières et expertes de situations industrielles (par exemple, les opérateurs de ligne de production dans l'industrie). Pour compléter nos résultats sur le lien entre la perception du danger et la confiance, il serait intéressant d'approfondir les mécanismes qui concourent à la perception du risque.

Plus globalement, cette expérimentation ouvre des perspectives intéressantes sur le thème de l'Human-Autonomy Teaming pour l'industrie 4.0. Les questions clés concernent la conception des nouvelles interactions entre les opérateurs et les systèmes cyber-physiques de l'Industrie 4.0 afin d'améliorer la fiabilité du processus de décision et d'éviter les erreurs. Parasuraman et Riley (1997) ont montré qu'une confiance excessive envers les CPS n'est pas nécessairement bénéfique pour la performance. Les opérateurs peuvent ne pas remettre en question les choix des agents et suivre aveuglément leurs propositions. Cette surconfiance pourrait conduire à une mauvaise prise de décision, surtout si l'agent technologique n'est pas fiable. Dans l'industrie 4.0, où les opérateurs humains supervisent des CPS, deux comportements différents peuvent être observés : soit les opérateurs effectuent une analyse approfondie de la situation, augmentant leur niveau d'activité, soit ils effectuent une analyse superficielle et se contentent d'accepter la proposition faite par la machine. Ce dernier comportement est le phénomène de complaisance (Parasuraman & Riley, 1997). Proposer un CPS qui adapterait sa transparence afin d'éviter ces écueils est une voie récente et prometteuse (Okamura & Yamada, 2020 ; Akash et al., 2020a ; Akash et al., 2020b). Le calibrage dynamique de la confiance pourrait être basé sur le taux d'acceptation de

l'opérateur plutôt que sur sa confiance. La transparence pourrait être utilisée pour jouer sur les dimensions liées au raisonnement et à l'analyse de la situation par le cobot. Cette transparence pourrait aussi être utilisée sur les dimensions liées à la perception du danger ou du risque, qui peuvent avoir un rôle médiateur entre la confiance et l'acceptation de la proposition. Un tel processus dynamique permettrait d'éviter à l'opérateur d'accepter de manière inappropriée les propositions d'un cobot, et d'améliorer la coopération entre l'homme et le CPS.

6.4. Conclusion

Cette étude avait comme objet d'étude l'effet des modalités analytique et état de l'opérateur dans la prise de décision de niveau opérationnel d'un opérateur assistée par un cobot. Les effets de la transparence de l'agent ont été examinés à travers les niveaux de transparence du modèle de Chen et al. (2018) et appliqués aux dimensions de la transparence en utilisant le cadre de Lyons (2013). Nous avons observé les effets de cette transparence sur l'acceptation des opérateurs, leur confiance dans le cobot et leur perception du danger. En outre, conformément aux résultats d'études antérieures, nous n'avons pas observé d'augmentation de la charge mentale des participants malgré l'augmentation du niveau de transparence. Nous avons également constaté que les informations sur le risque pouvaient modérer les informations sur le raisonnement du cobot le cadre de cette prise de décision.

Notre première hypothèse est partiellement validée par nos résultats : **les composantes de confiance apprise et de confiance situationnelle sont corrélées avec la prise de décision** (Figure 6.7). Ce résultat permet d'apporter des éléments pour notre premier verrou. La validation de cette hypothèse consolide la proposition de modèle de Hoff et Bashir (2015) d'une constitution de la confiance en différentes composantes. Notons cependant que la confiance ne semble pas être un déterminant de la prise de décision de l'opération, contrairement à la confiance apprise ou la confiance situationnelle.

Notre seconde hypothèse est validée par nos résultats : **les modalités de transparence sont corrélées avec l'acceptation**. Mais notre troisième hypothèse n'est que partiellement validée par nos résultats : **les modalités de transparence sont en partie corrélées avec les composantes de la confiance** (Figure 6.7). Ces deux hypothèses alimentent notre second

verrou de recherche. Les résultats montrent que les modalités de transparence ont un effet sur la prise de décision de l'opérateur. En revanche, les composantes de confiance ont différentes sensibilités face aux modalités de transparence, et l'effet de ces modalités sera plus ou moins grand. Pour conclure, nous pensons que pour favoriser l'émergence d'un opérateur 4.0 en symbiose avec les systèmes cyber-physiques, l'optimisation de la coopération humain-machine doit s'appuyer sur un dialogue de qualité. Il serait donc intéressant d'explorer plus en détail l'effet des niveaux de transparence sur d'autres dimensions. Une meilleure compréhension des modalités de transparence et de ses effets sur la confiance et la perception du danger, ou du risque, pourrait être une voie pour calibrer dynamiquement la confiance afin d'éviter les phénomènes de surconfiance ou de sous-confiance.

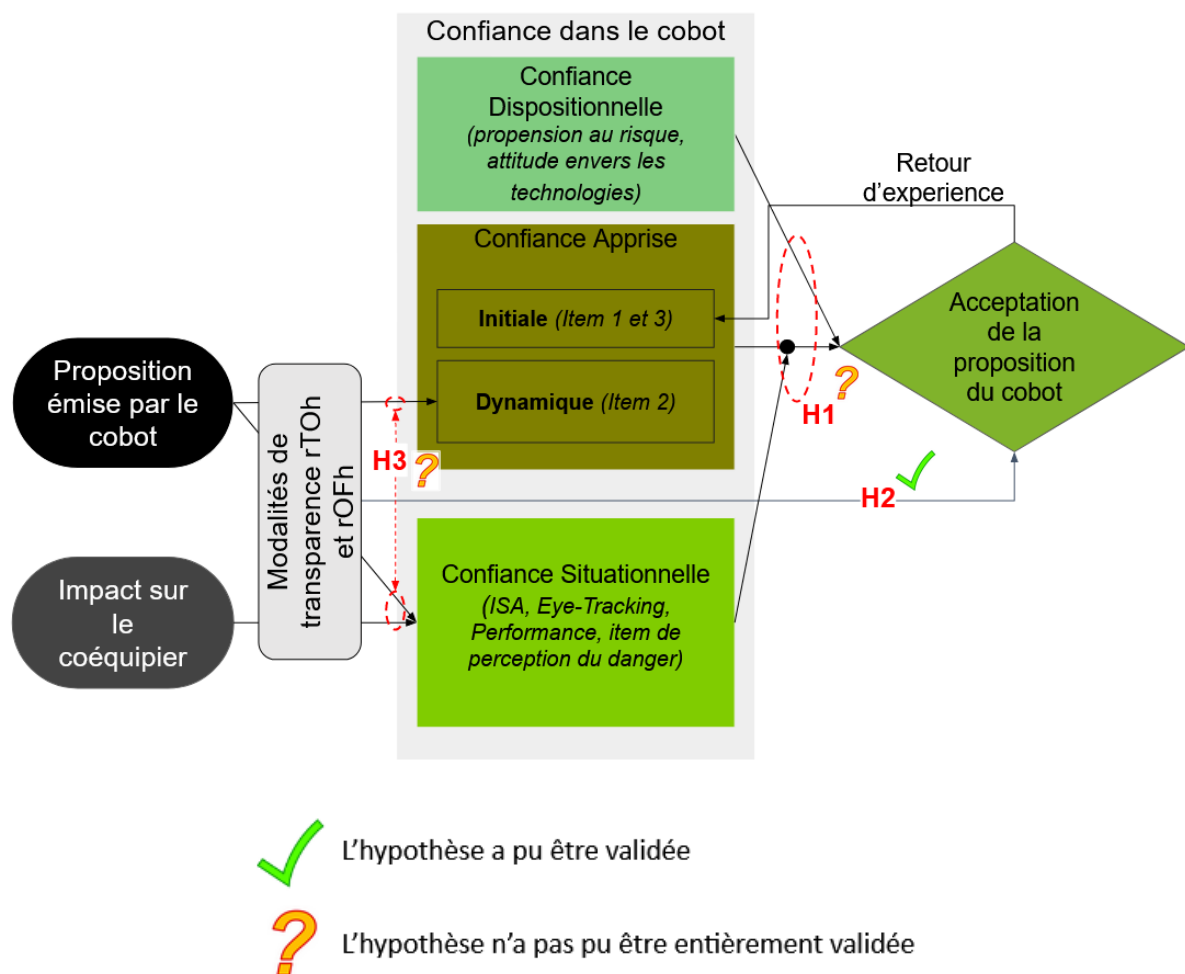


Figure 6.7 : Validation et synthèse des hypothèses initiales

Chapitre 7 : Confiance et transparence dans une prise de décision tactique avec un CPS fiable

Ce deuxième chapitre expérimental vise à explorer le rôle de la transparence dans la prise de décision de l'opérateur assistée par une IA. Ceci correspond à notre **premier** et à notre **second verrou** scientifique. Pour rappel, le premier verrou porte sur la **compréhension du rôle des composantes de la confiance dans la prise de décision** (section 5.2) et la seconde porte sur **la compréhension des effets des modalités de la transparence sur l'acceptation et la confiance de l'opérateur** (section 5.3).

A la différence du chapitre précédent, nous répondons ici à nos questions de recherche sous le prisme d'une prise de décisions assistée par un CPS portant sur *un niveau tactique* (gestion du calendrier de maintenance) et émanant d'une *IA de fiabilité élevée et connue*.

Nous cherchons à améliorer la compréhension du lien existant entre la confiance dispositionnelle (CD), la confiance apprise (CA) (initiale (CAI) et dynamique (CAD)), la confiance situationnelle (CS) et l'acceptation. Nous continuons d'investiguer le lien entre la CD, la CA, et l'acceptation. Par rapport à l'expérimentation précédente nous regardons plus en détail le rôle des facteurs individuels (qui forment la CD) dans la prise de décision d'un opérateur assistée par une IA. Comme cadre pour la transparence nous utilisons à nouveau le modèle de Situation Awareness Transparency de Chen et al. (2018) et celui de Lyons (2013). Nous utilisons les modalités de transparence telles que définies dans la section 3.1.1.3 « Articulation des modèles de transparence ». Nous nous intéressons aux effets de la transparence sur l'acceptation de la proposition, sur les métriques de CAI, de CAD et sur la CS (perception du risque et la charge mentale).

Dans cette expérimentation, les participants ont été mis dans le rôle d'un gestionnaire de calendrier de maintenance dans un cadre maritime (i.e. le rôle d'un chef mécanicien). Les participants avaient à accepter ou refuser la proposition d'une intelligence artificielle visant à modifier la date d'une opération de maintenance déjà planifiée par le logiciel de gestion de maintenance assistée par ordinateur (GMAO). Nous mesurons leur affinité envers les technologies, leur propension à faire confiance aux technologies et leur propension au risque.

Nous mesurons également la confiance apprise initiale, dynamique et situationnelle des participants.

7.1. Méthode

7.1.1. Participants

Cette étude a pris place en janvier 2022 à distance. La campagne de recrutement a duré 3 semaines et s'est faite par l'envoi de mail et par des interventions en début de cours. Les participants étaient invités à participer à une étude, à distance, portant sur la coopération humain-machine dans le cadre de la maintenance prévisionnelle et de l'industrie 4.0. Pour leur participation à l'étude les participants étaient rétribués avec un bon d'achat de 10 euros. 39 participants issus de cursus ingénieurs (Préparation aux études d'ingénieurs, Mécatronique et Génie Industriel) ont participé. L'effectif est composé de 25 hommes et 14 femmes. L'âge moyen est de 22,15 ans ($E-T = 2,77$). Dans l'effectif, 17 participants considèrent avoir des connaissances basiques sur le fonctionnement d'un moteur et 22 aucune connaissance. 12 participants ont déjà utilisé un logiciel de gestion de maintenance assistée par ordinateur et 27 n'en n'ont jamais utilisé.

7.1.2. Protocole expérimental

Les participants sont mis face à six situations. Chaque situation se compose d'une proposition émanant d'une intelligence artificielle d'aide à la maintenance prévisionnelle que les participants doivent accepter. Les situations présentées aux participants varient selon les deux variables indépendantes suivantes :

- La **criticité de la situation** (Tableau 7.1) : les modalités de criticité de la situation ont été élaborées avec des chefs mécaniciens et sur la base d'une analyse du domaine de travail. Nous utilisons deux types d'opérations de maintenance. Le premier type de maintenance concerne le remplacement de filtres à huile et à combustible. Le risque encouru d'une proposition sur cette opération est un risque mécanique (arrêt et casse moteur). La gravité de ce risque est **forte**. Le second type de maintenance concerne les nettoyages du turbocharger et du système de refroidissement. Le risque

encouru d'une proposition sur cette opération est un risque organisationnel (mise en tension du planning de maintenance). La gravité de ce risque est **moyenne**.

Tableau 7.1 : Modalités de criticité de la situation

<i>Proposition de l'IA</i>	<i>Risques encourus (dommages)</i>	<i>Probabilité du risque</i>	<i>Gravité du risque</i>	<i>Criticité de la situation</i>
Avancer un changement de filtre	Risque mécanique (Arrêt et casse moteur)	Faible probabilité	Gravité Forte	Modérée
Repousser un changement de filtre		Forte probabilité		Élevée
Avancer un nettoyage	Risque organisationnel (Mise en tension du planning de maintenance)	Forte probabilité	Gravité Moyenne	Élevée
Repousser un nettoyage		Faible probabilité		Modérée

Selon que l'IA propose d'avancer ou de reculer l'opération de maintenance, ces risques ont plus ou moins de probabilité de survenir. Repousser un changement de filtre est associé à une **forte probabilité** d'un arrêt et d'une casse moteur. Avancer une opération de nettoyage est, quant à elle, associée à une **forte probabilité** d'une mise en tension du planning des équipiers.

La maintenance en milieu maritime a peu de tolérance concernant les risques (Simon et al., 2021a). Les deux situations où ces propositions sont faites (i.e. **reculer la maintenance d'un filtre** et **avancer une opération de nettoyage**) sont considérées à criticité élevée car avec de forte probabilité de survenue de risque de gravité forte ou moyenne (Tableau 7.2).

Tableau 7.2 : Matrice d'analyse des risques

		Gravité		
		Faible	Moyenne	Forte
Probabilité	Faible		Repousser un nettoyage	Avancer un changement de filtre
	Moyenne			
	Forte		Avancer un nettoyage	Repousser un changement de filtre

Criticité faible	Criticité modérée	Criticité élevée
-------------------------	--------------------------	-------------------------

- La **transparence** (Tableau 7.3) : les modalités de transparence de l'IA sont manipulées sur deux dimensions et selon deux niveaux. Lorsque l'IA communique sur sa fiabilité, elle émet des informations issues de la **dimension analytique** de **niveau 3** (A3). L'IA communique de façon continue les données qu'elle a utilisées et lorsqu'elle n'émet pas d'information sur sa fiabilité elle est en modalité de transparence A2. Lorsque l'IA communique sur les risques à venir, elle émet des informations issues de la **dimension l'environnement** de **niveau 3** (E3). Lorsqu'elle n'est pas transparente sur l'environnement, l'opérateur n'a aucune information, l'IA est dans une modalité de transparence E0.

Tableau 7.3 : Modalités de transparence de l'IA utilisées dans l'expérimentation

<i>Types d'informations transparentes</i>	<i>Modalités de transparence</i>	<i>Informations transmises aux participants</i>
Fiabilité (F)	A3 et E0	Fiabilité de 90%
Fiabilité + Risque (F+R)	A3 et E3	Fiabilité à 90% et informations sur les risques encourus si acceptation ou si refus de la proposition
Risque (R)	A2 et E3	Informations sur les risques encourus si acceptation ou si refus de la proposition

A chaque nouvelle situation, les participants sont confrontés à des modalités de transparence différentes de la situation précédente. Le croisement des modalités de

transparence et des modalités de criticité de la situation permet de définir 6 conditions expérimentales (Tableau 7.4, Figure 7.1).

Tableau 7.4 : Conditions expérimentales selon la criticité de la situation et la transparence de l'IA

Condition expérimentale	Criticité de la situation	Transparence de l'IA
1	Modérée	F
2	Élevée	
3	Modérée	F+R
4	Élevée	
5	Modérée	R
6	Élevée	

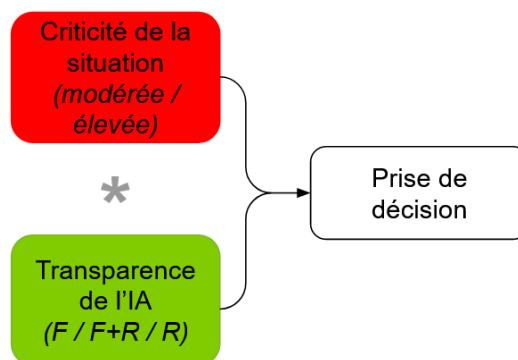


Figure 7.1 : Conditions expérimentales de l'expérimentation

Déroulé de la passation. Les expérimentations se déroulant à distance, un lien teams était communiqué aux étudiants. Une fois le participant connecté, l'expérimentateur lui présentait l'objectif de l'étude et le déroulement. Le participant devait ensuite remplir la feuille de consentement libre et éclairé (annexe C). L'expérimentateur présentait alors le contexte de la maintenance prévisionnelle en milieu maritime, les différentes maintenances qu'ils allaient rencontrer, puis les différentes interfaces que le participant allait utiliser. Suite à cela l'expérimentateur donnait le contrôle de son ordinateur à distance au participant pour qu'il puisse s'approprier l'interface du logiciel de maintenance prévisionnelle. Puis, les participants étaient amenés à remplir les questionnaires, sur le logiciel Limesurvey, portant sur leur affinité à la technologie (annexe D.1), leur propension à faire confiance à la technologie (annexe D.2), leur propension au risque (annexe D.3) et leur confiance apprise

initiale (annexe E.1). Les participants avaient à réaliser six situations où l'IA leur proposait de décaler une opération de maintenance. Chaque situation était indépendante et les participants n'avaient pas connaissance des répercussions de leur prise de décision. Entre chaque situation les participants avaient à répondre à un questionnaire permettant d'évaluer leur confiance apprise initiale (CAI) (annexe E.1), leur confiance apprise dynamique (CAD) (annexe E.2) et leur confiance situationnelle (CS) (annexes F.1 et F.2). Une fois les six situations réalisées, les participants étaient amenés à remplir un questionnaire sociodémographique (annexe G).

7.1.3. Matériel

7.1.3.1. Interfaces Seanatic

Nous présentons ici le logiciel servant de base aux interfaces de l'expérimentation et nous nous attardons plus en détail sur les interfaces utilisées dans cette expérimentation. Les différentes interfaces du logiciel ont été créées grâce à la méthodologie de conception d'interface écologique (Burns & Hajdukiewicz, 2004). Nous avons utilisé la méthode du Cognitive Work Analysis (CWA : Rasmussen & Vicente, 1989 ; Vicente, Christoffersen & Perekhita, 1995 ; Stanton, Salmon, Walker & Jenkins, 2017) couplé à l'utilisation de méthodes centrées opérateurs (Lallemand & Gronier, 2015). La mise en application de ces deux méthodes a permis d'inscrire la notion d'utilité au centre de la création du logiciel (Loup-Escande, Burkhardt & Richir, 2013). Le détail de l'analyse du travail de maintenance en milieu maritime de même que la création du logiciel Seanatic ont fait l'objet de publications lors de conférences (Simon et al., 2021a ; Simon, Rauffet & Guerin, 2021c) et de rapports de projets pour l'ADEME (Lassalle & Simon, 2020 ; Simon & Lassalle, 2022 ; Simon & Rauffet, 2022).

La création de l'interface Seanatic a nécessité plusieurs étapes. La première étape a été la compréhension du domaine de travail. Pour ce faire, des entretiens avec des chefs-mécaniciens (sujets experts) et l'utilisation de documents de référence (cours, normes, guides de maintenance) ont permis de créer une hiérarchie d'abstraction (annexe H). Cette hiérarchie d'abstraction a permis la représentation des objectifs, contraintes, tâches et outils d'un chef mécanicien au long cours (i.e. navigation longue durée). Une fois cette hiérarchie d'abstraction formalisée, elle a été soumise puis validée par les sujets experts. De ces

entretiens et de cette analyse du domaine de travail nous avons pu réaliser une analyse des tâches (annexes I.1 et I.2) que nous avons complétée avec une analyse de l'organisation sociale et des coopérations (annexe I.3). La première analyse permettait de mettre en avant les moments où étaient réalisées les tâches liées à la gestion de la maintenance (comme la re-planification). La seconde analyse permettait de mettre en avant les opérateurs habilités à réaliser ces tâches. Pour compléter ces analyses, des maquettes ont été réalisées conjointement avec des chefs-mécaniciens, des ergonomes, un expert IA et des concepteurs d'interface lors d'ateliers d'idéation et de conception (annexes J.1 à J.7). Les interfaces du logiciel ont été créées sur la base des éléments issus du CWA et des maquettes. Le logiciel a ensuite été amélioré par le biais de tests opérateurs avec des étudiants de l'École Nationale Supérieure et Maritime. Le logiciel final, dont les interfaces sont utilisées dans cette expérimentation, a été validé par des chefs mécaniciens lors de tests utilisateurs.

Le logiciel Seanatic est composé de quatre interfaces : un tableau de bord (permettant une vision rapide des propositions de l'intelligence artificielle), un planning (comportant des calendriers selon une vue hebdomadaire, mensuelle, ou annuelle, et un Gantt des opérations prévues), une interface de données (permettant la visualisation des données récoltées et utilisées par l'IA pour ses projections) et une interface de dialogue (comportant la liste des propositions de l'IA ainsi que d'éventuelles informations complémentaires relatives à ces propositions). Ces quatre interfaces permettent à l'IA d'être transparente sur les trois niveaux du SAT (Chen et al., 2018). Dans le cadre de cette expérimentation nous utilisons l'interface comportant les données (Figure 7.2) et l'interface de dialogue (Figure 7.3) pour faire varier la transparence.

L'IA est transparente sur la dimension analytique de niveau 2 dans toutes les situations (Figure 7.2). Cette interface permet à l'opérateur de prendre connaissance des données (réelles et projetées) utilisées par l'IA pour faire sa proposition. Sur ces données apparaissent la date proposée par l'IA et la date initiale (issue de la Gestion de Maintenance Assistée par Ordinateur).



Figure 7.2 : Représentation des données lorsque la proposition est d’avancer un changement de filtre

L’interface de dialogue (Figure 7.3) permet à l’opérateur d’interagir avec l’IA en acceptant ou refusant sa proposition (éléments entourés en blanc dans la Figure 7.3). Cette interface permet d’avoir accès à la proposition de l’IA au niveau 3 du SAT (Chen et al., 2018) avec la présence de la **fiabilité** (entourée en vert dans la Figure 7.3) et du **risque** (entourée en rouge dans la Figure 7.3).

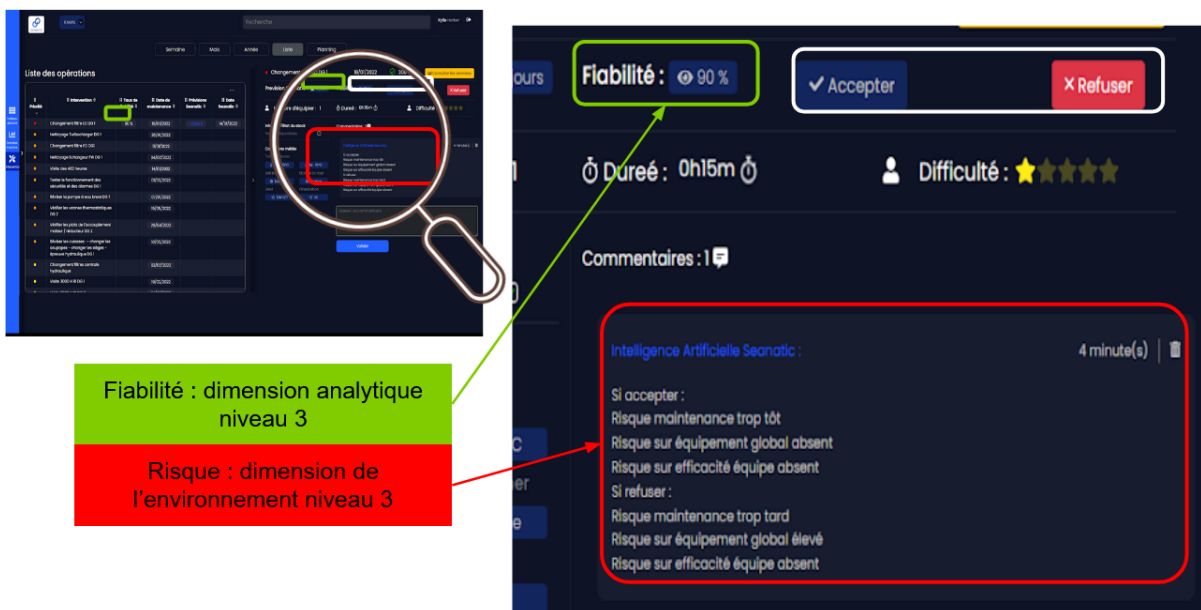


Figure 7.3 : Interface de dialogue lorsque la proposition est d’avancer un changement de filtre
Ce sont ces deux éléments que nous faisons varier, en les affichant ou en les occultant, selon les modalités de transparence. L’interface de dialogue de l’IA peut laisser apparaître la fiabilité

(dimension analytique de niveau 3), le risque (dimension de l'environnement de niveau 3) ou les deux à la fois.

7.1.4. Observables, méthode d'analyse des résultats, et hypothèses

Nous mesurons une variable de façon objective :

- **L'acceptation** - mesurée à chaque situation suite à la décision du participant (accepter ou refuser)

Nous mesurons sept variables de façon subjective :

- **La confiance dispositionnelle (CD) :**
 - L'affinité envers la technologie du participant grâce à l'échelle ATI (traduction française de Franke, Attig & Wessel, 2019) (Annexe D.1)
 - La propension du participant à faire confiance à la technologie grâce à l'échelle PTT (traduction française de Jessup, 2018) (Annexe D.2)
 - La propension générale au risque du participant grâce à l'échelle GriPS (traduction française de Zhang et al., 2019) (Annexe D.3)
- **La confiance apprise initiale (CAI)** du participant - mesurée au début, puis après chaque situation d'interaction avec l'échelle IMOTRIS (traduction française de Mayer et al., 1995, Lyons & Guznov, 2019) (Annexe E.1)
- **La confiance apprise dynamique (CAD)** du participant - mesurée grâce à un item de confiance dans le signal après chaque situation (Annexe E.2)
- **La confiance situationnelle (CS) :**
 - La perception du risque grâce à l'échelle de perception du risque (traduction française de Wilson, Zwickle & Walpole, 2018) (Annexe F.1). Nous utilisons une échelle de risque globale car les risques peuvent être multiples (organisationnels et matériels). Le contexte de l'expérimentation étant la maintenance en milieu maritime, tout risque matériel pouvait également entraîner un risque corporel (si le navire s'échoue).
 - La charge mentale grâce au Nasa-TLX (traduction française de Hart & Staveland, 1988) après chaque situation d'interaction. (Annexe F.2)

Nous utilisons le logiciel statistique R (2021) pour nos analyses. Les tests statistiques sont des régressions logistiques binomiale et ordinale, des régressions linéaires et des tests de corrélation. Nous testons le meilleur des deux modèles entre des effets indépendants de la transparence et de la situation ou un effet d'interaction en plus d'effets indépendants. Nous comparons leurs AIC, et pour les régressions binomiales leurs capacités prédictives, afin de sélectionner le modèle le plus pertinent.

Nous répondons aux questions de nos **premier** et **deuxième verrous** au travers des hypothèses exploratoires suivantes (Figure 7.4) :

H1. Les composantes de la confiance seraient corrélées avec l'acceptation (Hoff & Bashir, 2015). Par ailleurs, nous considérons qu'il existerait une interaction entre la confiance apprise et la confiance situationnelle (Chancey et al., 2017) (**Verrou 1**).

H2. Les modalités de transparence de la **dimension analytique** et de la **dimension environnement** d'une IA à **fiabilité haute** seraient corrélées avec l'acceptation de la proposition (Mercado et al., 2016 ; Sadler et al., 2016 ; Stowers et al., 2020 ; Zhang, Liao & Bellamy, 2020) (**Verrou 2**). Comme suggéré par Monsaingeon, et al., (2020) et Zhang et al. (2022), on s'attendrait à ce que cette fiabilité soit en modérée négativement avec la dimension de l'environnement.

Nous souhaitons également investiguer l'effet de la criticité de la situation, et son interaction avec les modalités de la transparence, sur l'acceptation (Hoesterey & Onnasch, 2022).

H3. Les modalités de transparence de la **dimension analytique** et de la **dimension environnement** d'une IA à **fiabilité haute** seraient corrélées avec les composantes de la confiance de l'opérateur (Selkowitz et al., 2016 ; Mercado et al., 2016 ; Wright et al., 2017 ; Javid & Estivill-Castro, 2021) (**Verrou 2**).

Nous souhaitons également investiguer l'effet de la criticité de la situation, et son interaction avec les modalités de la transparence, sur les composantes de la confiance (Hoesterey & Onnasch, 2022).

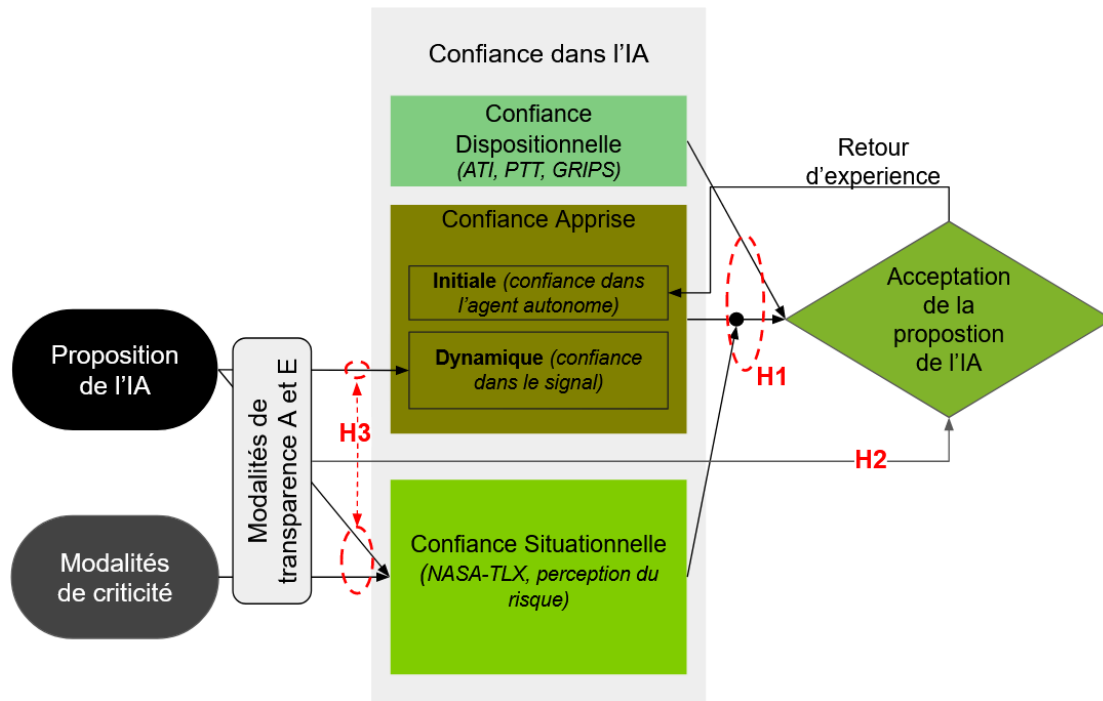


Figure 7.4 : Positionnement des hypothèses, des observables et des variables dans le cadre de la prise de décision assistée par une IA

7.2. Résultats

7.2.1. Relations entre les composantes de la confiance et l'acceptation

Relations entre la confiance dispositionnelle et l'acceptation. Les analyses de Mann-Whitney montrent que **la confiance dispositionnelle n'est pas corrélée avec l'acceptation**. Les scores à l'affinité envers la technologie (ATI) ne sont pas supérieurs chez les participant ayant accepté ($Mdn = 4$) et ceux ayant refusé ($Mdn = 4$) ($W=5686.5$, $p = 0.9907$). Pareillement pour les scores de propension à la confiance (PTT) ($Mdn = 3.8$, $Mdn = 3.8$, $W=5062.5$, $p = 0.17$) et de propension au risque (Grips) ($Mdn = 2.63$, $Mdn = 3.5$, $W=4939.5$, $p = 0.11$).

Relations entre la confiance apprise et l'acceptation. **La confiance apprise initiale n'est pas corrélée à l'acceptation**. La confiance dans l'IA ne suivant pas une loi normale, nous utilisons une analyse Mann-Whitney. Nous ne constatons pas de différence significative de confiance dans l'IA entre les participants ayant accepté ($Mdn = 4.4$), et les participants ayant refusé ($Mdn = 4.3$) ($W = 5295$, $p > .05$).

La confiance apprise dynamique est positivement corrélée à l'acceptation. L'analyse Mann-Whitney nous permet de constater que la confiance dans le signal est significativement

plus élevée lorsque les participants acceptent ($Mdn = 6$) que lorsqu'ils refusent ($Mdn = 4$) ($W = 2213, p < .001$).

Relations entre la confiance situationnelle (perception du risque) et l'acceptation. **La confiance situationnelle est négativement corrélée avec l'acceptation.** L'analyse Mann-Whitney entre le premier item de l'échelle de perception du risque et l'acceptation permet de constater que la perception du risque est significativement plus faible lorsque les participants acceptent ($Mdn = 2$) que lorsqu'ils refusent ($Mdn = 3$) ($W = 8162, p < .001$).

Relations entre la confiance apprise, la confiance situationnelle et l'acceptation. Nous regardons si la perception du risque est en interaction avec la confiance dans l'IA et dans la confiance dans le signal ou si ce sont trois facteurs indépendants. Pour ce faire nous comparons deux modèles :

- Modèle 1 (3 facteurs indépendants + interactions) : $\text{Acceptation} \sim (\text{CAD} * \text{CAI}) * \text{CS}$
- Modèle 2 (3 facteurs indépendants + 1 interaction) : $\text{Acceptation} \sim (\text{CAD} * \text{CAI}) + \text{CS}$

L'AIC du second modèle ($\text{AIC} = 251$) n'est pas significativement plus faible que l'AIC du modèle 1 ($\text{AIC} = 260$) ($p > .01$). Notons cependant que son pouvoir prédictif est plus élevé (85% contre 78%). **Il ne semble pas y avoir d'interaction entre la confiance apprise (initiale et dynamique) et la confiance situationnelle.** Les résultats présentés dans cette section permettent de soutenir en partie H1.

7.2.2. Relations entre les modalités de transparence de l'IA et l'acceptation

Nous cherchons à savoir si la transparence et la criticité de la situation ont un effet sur l'acceptation et si ces deux variables interagissent. Pour cela nous utilisons une régression logistique binomiale et nous considérons les deux modèles suivants :

- Modèle 1 (2 facteurs indépendants) : $\text{Acceptation} \sim \text{Transparence} + \text{Criticité}$
- Modèle 2 (2 facteurs indépendants + interaction) : $\text{Acceptation} \sim \text{Transparence} * \text{Criticité}$

Nous considérons le modèle 2 comme meilleur car sa capacité prédictive est supérieure (72% contre 71%) et que ce modèle a un AIC significativement plus faible (respectivement $\text{AIC} = 253$ et $\text{AIC} = 230, p < .001$). De cette analyse ressort trois résultats majeurs :

Relations entre les modalités de transparence et l'acceptation. Premièrement, nous voyons un effet principal de la transparence sur l'acceptation. **La transparence sur l'environnement (« risque ») a un effet négatif sur l'acceptation** (Figure 7.5). A l'inverse lorsqu'elle est la seule communiquée par l'IA, **la transparence analytique a un effet positif sur l'acceptation.**

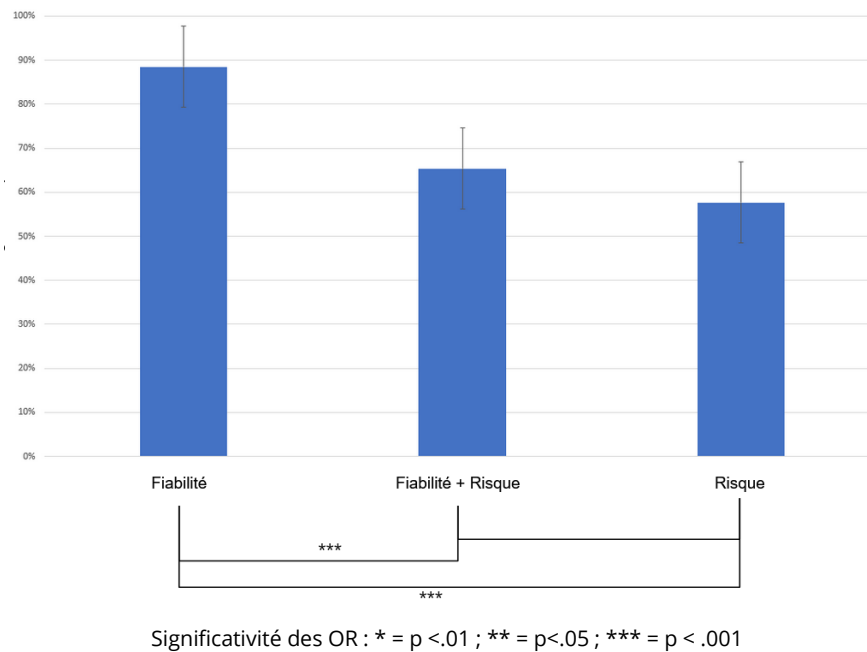
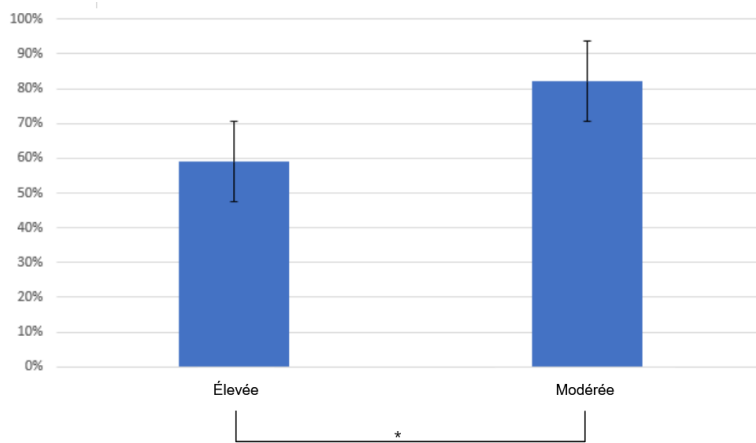


Figure 7.5 : Taux d'acceptation en fonction des modalités de transparence

Nous observons que la probabilité d'acceptation est significativement plus élevée lorsque la transparence de l'IA apporte uniquement des informations sur la « Fiabilité », par rapport aux autres modalités qui mentionne des éléments sur le risque (respectivement pour « Fiabilité + Risque » : $OR = 0.02, p < .001$; et pour « Risque » : $OR = 0.01, p < .001$).

Relations entre la criticité de la situation et l'acceptation. **La criticité de la situation a un effet négatif sur l'acceptation.** Le modèle issu de la régression logistique binomiale prévoit qu'une criticité moins élevée (« Modérée » comparé à « Élevée »), réduit significativement la probabilité d'acceptation ($OR = 0.09, p < .05$) (Figure 7.6).



Significativité des OR * = $p < .01$; ** = $p < .05$; *** = $p < .001$

Figure 7.6 : Taux d'acceptation en fonction de la criticité de la situation

Relations entre les modalités de transparence, la criticité de la situation et l'acceptation. Le modèle montre une interaction entre la transparence et la criticité de la situation sur l'acceptation. **La transparence sur l'environnement (« risque »), dans une situation à criticité élevée, modère négativement l'effet de la transparence analytique (« fiabilité ») sur l'acceptation.** Nous observons que la probabilité d'acceptation est significativement plus faible lorsque la transparence de l'IA apporte uniquement des informations sur la « Fiabilité » dans des situations à criticité élevée, par rapport à une IA transparente sur le risque dans situations à criticité modérée (respectivement « Fiabilité + Risque » : « Modérée » : OR = 50, $p = .001$; et pour « Risque » : « Modérée » : OR = 144, $p < .001$) (Figure 7.7 et Figure 7.8). Ces résultats permettent de soutenir H2.

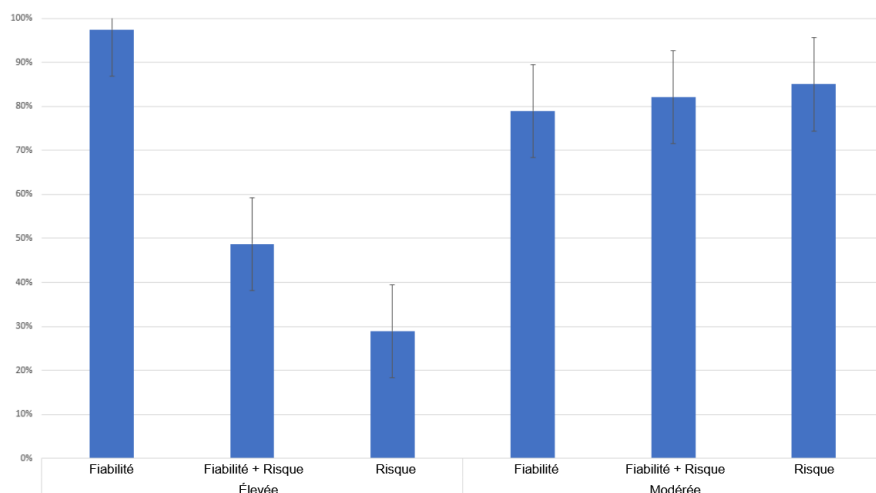


Figure 7.7 : Taux d'acceptation en fonction de la transparence et de la criticité de la situation

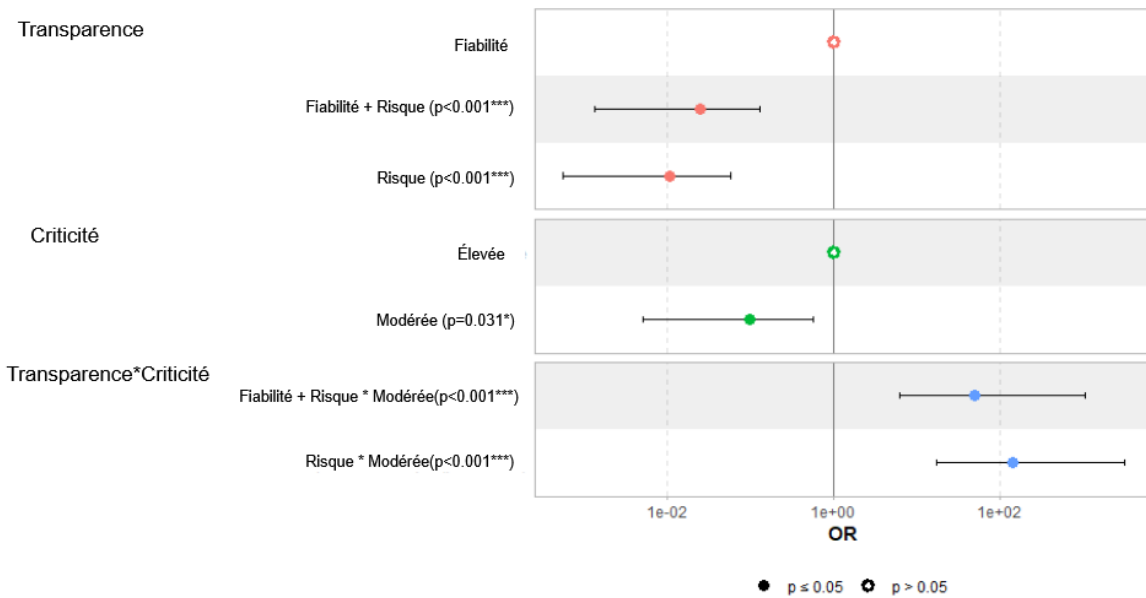


Figure 7.8 : Régression logistique sur l'effet de la transparence et de la situation sur l'acceptation (référence transparence « F » et criticité « Élevée »)

7.2.3. Relations entre les modalités de transparence de l'IA et les composantes de confiance

Relations entre les modalités de transparence et la confiance apprise initiale. Pour explorer l'effet de la transparence et de la criticité de la situation sur la composante de confiance apprise initiale nous utilisons un test de Kruskal-Wallis. Les résultats de celui n'indiquent pas d'effet significatif. **La transparence analytique et environnement n'ont pas d'effets sur la confiance apprise initiale.**

Pour analyser le rôle de la transparence et de la criticité de la situation sur les composantes de confiance apprise dynamique (confiance dans le signal) et de confiance situationnelle (questionnaire de perception du risque) nous utilisons des régressions (logistiques). Pour chacune des composantes de confiance nous comparons deux modèles :

- Modèle 1 (2 facteurs indépendants) : Composantes de confiance \sim Transparence + Criticité
- Modèle 2 (2 facteurs indépendants + interaction) : Composantes de confiance \sim Transparence * Criticité

Sur chacune des trois composantes de confiance investiguées, le second modèle est celui avec l'AIC le plus faible (Tableau 7.5). Nous regardons plus en détail ces modèles.

Tableau 7.5 : Prédicativité des modèles sur les composantes de la confiance en fonction de la transparence et de la criticité de la situation

Confiance		Modèle	AIC	<i>p-value</i> ¹
Confiance dans le signal (CAD)	~	Transparence + Criticité	804	<.01**
		Transparence * Criticité	811	
Perception du risque (CS)	~	Transparence + Criticité	670	<.001***
		Transparence * Criticité	684	

¹Différence entre les modèles 1 et 2

Relations entre les modalités de transparence et la confiance apprise dynamique. Premièrement, la transparence sur l'environnement (« risque ») a un effet négatif sur la confiance apprise dynamique (CAD) et la transparence analytique (« fiabilité ») a un effet positif sur la confiance apprise dynamique. (Figure 7.9).

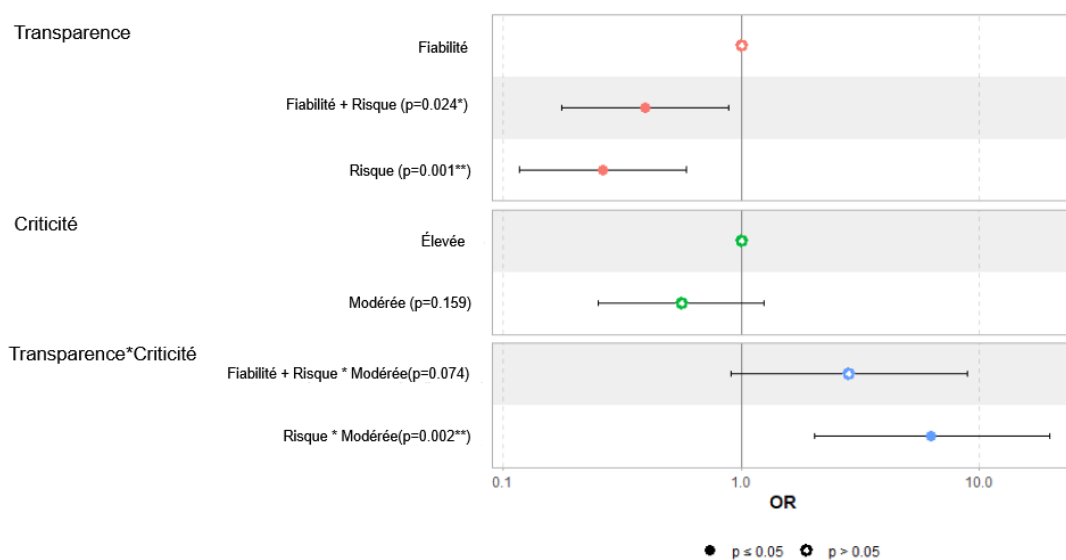


Figure 7.9 : Régression logistique ordinaire sur l'effet de transparence et de la situation sur la confiance dans le signal (référence transparence « F » et criticité « Élevée »)

Nous constatons que lorsque l'IA est transparente uniquement sur sa « Fiabilité » il est significativement plus probable d'avoir une confiance dans le signal (CAD) plus élevée

comparée aux modalités où l'IA communique sur les risques (respectivement pour « Fiabilité + Risque » : $OR = 0.39, p < .05$; et pour « Risque » : $OR = 0.26, p = .001$)

Relations entre les modalités de transparence et la confiance situationnelle (perception du risque). **La transparence sur l'environnement (« risque ») a un effet positif sur la confiance situationnelle (CS) et la transparence analytique (« fiabilité ») a un effet négatif sur la confiance situationnelle (CS).** Nous constatons que lorsque l'IA est transparente uniquement sur sa « Fiabilité » il est significativement plus probable d'avoir une perception du risque (CS) plus faible comparée aux modalités où l'IA communique sur les risques (respectivement pour « Fiabilité + Risque » : $OR = 4.73, p < .001$; et pour « Risque » : $OR = 6.02, p = .001$) (Figure 7.10).

Relations entre la criticité de la situation et la confiance situationnelle (perception du risque). **La criticité de la situation a un effet négatif sur la confiance situationnelle.** Nous constatons que lorsque la criticité de la situation est « Modérée » la probabilité de percevoir du risque (CS) est plus élevée que par rapport à une criticité « Élevée » ($OR = 3.21, p = .005$) (Figure 7.10).

Relations entre la criticité de la situation, les modalités de transparence et la confiance apprise et situationnelle (perception du risque). Pour explorer l'effet de la transparence et de la criticité de la situation sur la composante de confiance apprise initiale nous utilisons un test de Kruskal-Wallis. Les résultats de celui n'indiquent pas d'effet significatif. **La transparence et la criticité de la situation n'ont pas d'effet sur la confiance apprise initiale.** En revanche, **la transparence sur l'environnement (« risque »), dans une situation à criticité modérée, a un effet positif sur la confiance apprise dynamique (CAD) et la transparence analytique (« fiabilité »), dans une situation à criticité élevée, a un effet négatif sur la confiance apprise dynamique (CAD).** Par rapport aux modalités de référence « Fiabilité » et criticité « Élevée », le modèle prédit une confiance dans le signal (CAD) plus élevée que lorsque l'IA est transparente uniquement sur le « Risque » et avec une situation de criticité « Modérée » ($OR = 6.26, p < .01$) (Figure 7.9). Cette interaction est inversée sur la perception du risque (CS). **La transparence sur l'environnement (« risque »), dans une situation à criticité modérée, a un effet négatif sur la confiance situationnelle et la transparence analytique (« fiabilité »), dans une situation à criticité élevée, a un effet positif sur la confiance situationnelle (CS).** La

perception du risque (CS) est plus élevée lorsque l'IA est transparente sur la « Fiabilité » avec une criticité « Élevée », que dans les deux modalités où l'IA communique des risques et avec une situation de criticité « Modérée » (respectivement pour « Fiabilité + Risque » : « Modérée » : OR = 0.11, $p < .001$; et pour « Risque » : « Modérée » : OR = 0.13, $p < .001$) (Figure 7.10).

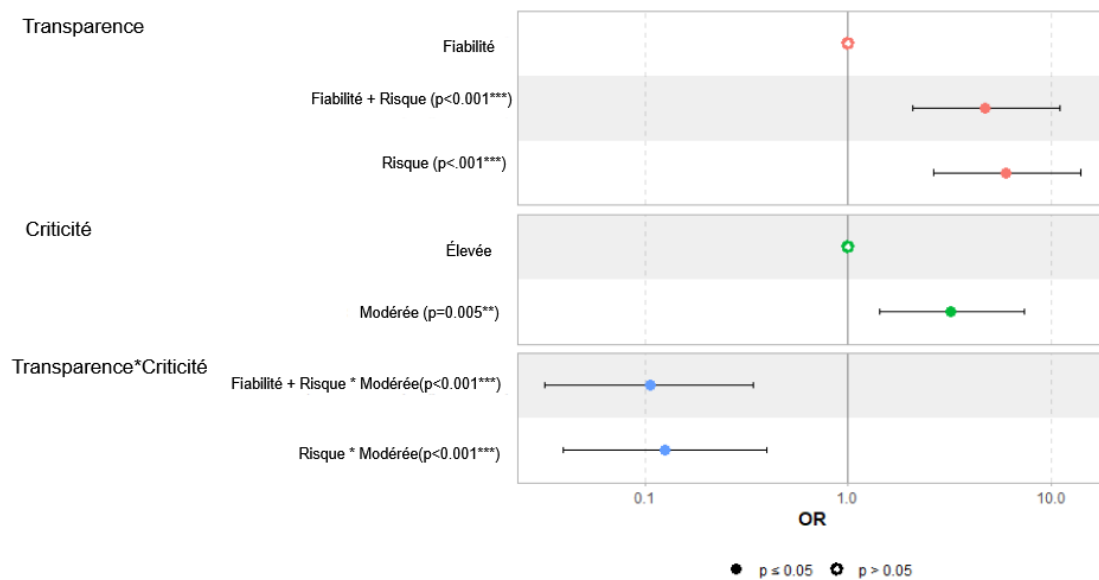


Figure 7.10 : Régression logistique ordinal sur l'effet de la transparence et de la situation sur la perception du risque (référence transparence « F » et criticité « Élevée »)

Relations entre les modalités de transparence, la criticité de la situation et la confiance situationnelle (charge mentale). Afin de s'assurer que le niveau de transparence n'infère pas sur la charge mentale et vérifier un potentiel effet de la situation, nous utilisons une anova à deux facteurs. **La transparence n'a pas d'effet sur la charge mentale.** Nous ne constatons pas d'effet des modalités de transparence et de criticité de la situation sur la charge mentale ($F(2,1,2) = 0.799, p = .45$), ni d'effet simple de la transparence ($p = .51$), ni de la situation ($p = .57$). Les résultats présentés dans cette partie permettent de soutenir H3.

7.3. Discussion

Cette seconde expérimentation a permis de montrer l'existence d'un lien entre la confiance apprise dynamique et l'acceptation et un lien entre la confiance situationnelle et l'acceptation. En revanche, nous ne retrouvons pas de lien entre la confiance dispositionnelle et l'acceptation. Nous ne trouvons pas non plus de lien entre la confiance apprise initiale et

l'acceptation. Enfin, nous ne trouvons pas d'effet d'interaction entre la confiance situationnelle, la confiance apprise (dynamique et initiale) et l'acceptation. Les résultats présentés montrent également que les modalités de transparence ont un effet sur l'acceptation. Une IA transparente sur les risques liés à sa proposition verra sa proposition moins acceptée par l'opérateur. Cette transparence a également un effet sur les composantes de confiance apprise dynamique et de confiance situationnelle.

Relations entre les composantes de la confiance et l'acceptation. Nous ne constatons pas d'effet de la confiance dispositionnelle sur l'acceptation malgré ce qui aurait pu être attendu (Hoff & Bashir, 2015 ; Jessup, 2018 ; Franke et al., 2019). **La confiance dispositionnelle semble être un déterminant mineur de l'acceptation de l'opérateur.** Malgré cela, il semble important de continuer à explorer cette facette de la confiance de par son intérêt pour l'acceptation de la technologie et comme variable contrôle.

La confiance apprise initiale n'est également pas liée à l'acceptation. En revanche, la confiance apprise dynamique permet de prédire la prise de décision de l'opérateur. Plus l'opérateur a une confiance apprise dynamique élevée dans l'IA, plus il aura tendance à accepter sa proposition. Son inverse est également vrai. Comme évoqué par Hoff et Bashir (2015), **la confiance apprise dynamique est déterminant de l'acceptation.** La confiance apprise dynamique est un meilleur prédicteur des comportements de confiance que la confiance apprise initiale. Ceci est explicable par le fait que, tout comme le comportement, la CAD est le reflet de la confiance de l'opérateur envers l'IA à un moment donné. Le comportement et la CAD se cristallisent dans l'interaction en fonction de la situation. Même si l'opérateur fait confiance à l'IA (i.e. la CAI est élevée : on le considère fiable quand il répond à nos attentes), il est possible que dans une situation précise, l'opérateur préfère agir de son propre chef et refuser sa proposition. Par exemple, l'IA communique une faible fiabilité de sa proposition, ce qui dans la situation amène l'opérateur à baisser momentanément sa confiance apprise dynamique et refuser la proposition, mais pas sa confiance apprise initiale. Sa proposition est momentanément moins adaptée par rapport à la situation. Une autre explication serait que l'opérateur se considère comme plus expert (Schaffer et al., 2019).

La confiance situationnelle (CS) permet également de prédire si l'opérateur va accepter la proposition de l'IA. **Plus la CS (perception du risque) est élevée, moins l'opérateur**

aura tendance à accepter la proposition de l'IA. Si le participant perçoit une situation comme risquée, il aura tendance à refuser la proposition de l'IA. Nous voyons ici le rôle qu'à la confiance situationnelle dans la prise de décision de l'opérateur.

Dans le contexte présenté ici, où une IA fiable fait une proposition, la perception du risque n'interagit pas avec la confiance apprise dynamique ou initiale. **La CAD et la CS influencent l'acceptation de façon indépendante.** La fiabilité de l'IA dans cette expérimentation était élevée (90%), et ses propositions étaient cohérentes avec l'analyse des données utilisées (disponible dans l'interface, voir Figure 7.2) par l'IA pour faire sa proposition. Le fait que le risque soit pris en compte par l'IA et qu'il soit fiable explique peut-être que le risque n'a pas d'effet modérateur sur la confiance dans le signal et sa transformation en comportement. Si nous ne retrouvons pas d'effet de modération comme proposé par Chancey et al. (2017), tout comme Hoesterey et Onasch (2022), nous voyons ici que la perception du risque vient influencer l'acceptation. Cette influence permet de considérer la perception du risque comme constitutif de la confiance situationnelle telle que décrite par Hoff & Bashir (2015). Notons également que l'acceptation peut être considérée comme l'élément le plus important à prendre en compte dans la coopération humain-machine (Bolton, 2022). Si le risque influence ce comportement, il est nécessaire de continuer à prendre en compte cet élément dans la coopération humain-machine.

Toutes les composantes de la confiance ne sont pas corrélées à l'acceptation, ce qui nous permet de mieux comprendre notre premier verrou visant à savoir quelles composantes sont des déterminants de la prise de décision de l'opérateur. **Nous constatons notamment que la composante de la confiance apprise dynamique et de la confiance situationnelle sont des déterminants de la prise de décision de l'opérateur.**

Relations entre les modalités de transparence, la criticité et l'acceptation. De façon concomitante avec les résultats de Pynadath et al. (2022), lorsque l'IA communique des informations sur les risques associés à sa proposition (transparence sur la dimension « environnement »), le taux d'acceptation est réduit. **La dimension sur l'environnement a un effet négatif sur l'acceptation.** Le risque et la fiabilité sont deux informations que Chen et al. (2018) considèrent de niveau 3 mais qui appartiennent à deux dimensions différentes pour Lyons (2013). Nous constatons ici que les informations contenues dans le niveau 3 n'ont pas

toutes le même effet sur l'acceptation. Comme le suggérait van de Merwe et al. (2022) il est important de considérer la transparence selon les informations communiquées en surcroît des modèles existants. Plus que le niveau de transparence, c'est ce sur quoi porte l'information communiquée qui joue sur l'acceptation de l'opérateur. Dans un environnement où la prise de risque est minimale, même si une IA est fiable, l'opérateur tiendra plus compte des risques liés à la proposition que de la fiabilité. En d'autres termes, et au regard du modèle de Lyons (2013) la transparence sur la dimension « environnement » aura plus d'influence que la transparence sur la dimension analytique. Comme Zang et al. (2022) nous voyons qu'une IA transparente sur sa haute fiabilité amène les participants à accroître leur confiance. En revanche, nous montrons que **la transparence sur une fiabilité haute et l'acceptation est modéré par la transparence sur la criticité de la situation.**

Nous observons également que la criticité de la situation joue un rôle dans l'acceptation de l'opérateur. Le modèle issu de nos données ne va pas dans le sens attendu car les opérations à criticité « Modérée » ont une probabilité plus grande d'être refusées. Une explication est que les participants étaient des non-experts. Leur manque d'expertise ne leur permettait pas de différencier la criticité de la situation lorsque le risque n'était pas affiché. L'interaction que l'on observe entre la transparence sur le risque de la dimension « environnement » et la criticité de la situation permet d'abonder dans ce sens. Une IA transparente sur le risque avec une situation de criticité modérée aura plus de chance de voir sa proposition acceptée que lorsqu'elle n'est pas transparente sur les risques et que la criticité de la situation est élevée. L'accès aux risques de la proposition permet de calibrer la confiance du participant et l'aide à savoir lorsqu'il est préférable d'accepter ou de refuser. Le fait que l'IA soit transparente sur la dimension « environnement » permet de compenser le manque d'expertise de l'opérateur en améliorant ses connaissances sur la situation.

Ces résultats nous permettent de valider notre seconde hypothèse. Cela améliore notre compréhension du lien entre la transparence et l'acceptation, en lien avec notre second verrou. **Les modalités de transparence des CPS vont influencer la prise de décision de l'opérateur**, et les modalités de transparence ayant des informations sur le risque ont un effet plus important sur la prise de décision.

Relations entre les modalités de transparence, la criticité et les composantes de la confiance. Nous voyons qu'en fonction de la composante de confiance observée, la sensibilité à la transparence sera plus ou moins forte. Cela permet de conforter le modèle proposé par Hoff & Bashir (2015) qui structurait la confiance en plusieurs composantes plus ou moins éloignées de l'interaction. Ceci expliquerait pourquoi la transparence n'a pas influencé la confiance apprise initiale (CAI) mais plutôt la confiance apprise dynamique (CAD). **La CAD s'inscrit dans l'interaction (Hoff & Bashir, 2015) et elle est dès lors plus sensible à la transparence de l'IA.** Cette sensibilité est à rapprocher du constat de Chen et al. (2021) sur le fait que la transparence impacte moins la confiance que l'acceptation. Sur la CAD, une IA transparente sur les risques provoque une baisse de cette composante.

Nous constatons également un effet de la transparence sur la confiance situationnelle (CS). **Nous constatons qu'une IA transparente sur les risques (dimension de l'environnement) augmente la perception du risque.** Le modèle de SAT définit qu'une IA qui communique sur les risques permet à l'opérateur d'améliorer sa conscience de la situation (Chen et al., 2014 ; Chen et al., 2018). Ici, ce serait le lien entre la conscience de la situation et la perception du risque qui expliquerait cette hausse. Comme la CAD, la CS est dans l'interaction. La CS est donc sensible aux variations de transparence d'une IA. Nous constatons également un effet de la criticité de la situation sur la perception du risque, mais le modèle issu de nos données va à l'inverse de nos attentes. Une criticité « modérée » amènerait à une perception du risque probablement plus élevée qu'une criticité « élevée ». Le manque d'expertise des participants a probablement conduit à ce que les propositions de criticité « modérée » soient interprétées comme représentant un risque. Enfin, nous constatons l'existence d'une interaction entre la transparence et la criticité de la situation sur la CAD et la CS.

Une IA transparente sur les risques lors d'une proposition à criticité « Modérée » vient augmenter la confiance et baisser la perception du risque. Le fait d'être transparent sur une absence de risque augmente la confiance dans la proposition, dans la perception de l'IA et réduit le risque perçu. Il y a un côté « rassurant » lorsque l'IA communique que sa proposition n'engendrera pas d'événements graves. **La situation joue un rôle important dans l'effet de la transparence sur un opérateur.** Une même modalité de transparence pour deux situations

aux enjeux différents aura des effets différents sur l'opérateur. Enfin, comme attendu (Chen, 2021) les changements de transparence ne sont pas associés à des changements de charge mentale.

Ces résultats valident en partie notre troisième hypothèse. La transparence a un effet sur la confiance apprise dynamique et sur la confiance situationnelle, mais pas sur la confiance apprise initiale. Cette différenciation permet d'éclairer notre second verrou sur le lien entre la transparence du CPS et les composantes de la confiance. **La transparence a un effet sur les composantes de confiance qui sont dans l'interaction, mais pas sur les composantes en dehors de l'interaction.**

Limites. L'une des limites de notre étude est liée au fait que nos participants étaient des novices. Il serait intéressant de pouvoir réaliser une réplique de celle-ci avec un public expert. Le niveau d'expertise pourrait être associé à une plus forte remise en question de la proposition de l'IA, et à un effet différent de la transparence sur le risque. Une perspective applicative pourrait être d'utiliser la transparence sur le risque dans les dispositifs de formation des novices afin de leur faire acquérir des compétences expertes.

Si la confiance permet de prédire la prise de décision, c'est plus précisément sa composante « confiance apprise dynamique » qui est la plus utile. Mesurer cette composante de la confiance permettrait de savoir si un opérateur va accepter ou refuser une proposition. L'absence de lien entre la CAI, la confiance dispositionnelle et l'acceptation est explicable par le fait que ce sont des mesures stables dans le temps. Or, la prise de décision est-elle relative à une proposition à instant T et dans un contexte précis. L'acceptation se base plus sur la CAD et la CS. Comme nous avons pu le montrer, la transparence joue un rôle essentiel dans ces deux composantes de confiance et dans l'acceptation. C'est pourquoi avoir une transparence qui est la plus adaptée possible à la situation est une voie à privilégier pour la coopération humain-machine, notamment dans le cas où la machine est force de proposition.

7.4. Conclusion

Cette expérimentation avait pour but d'explorer l'effet des modalités de transparence analytique (sur la fiabilité) et environnement (sur le risque) sur la prise de décision de niveau

tactique d'un opérateur assisté par une IA. Pour ce faire, les participants étaient chargés de répondre aux propositions d'une IA sur des changements de calendrier de maintenance.

Notre première hypothèse n'est pas complètement validée par nos résultats. **Une partie des composantes de la confiance (confiance apprise dynamique et confiance situationnelle) seraient corrélées avec l'acceptation** (Figure 7.11). Ce résultat permet de mieux comprendre le premier verrou. Toutes les composantes de la confiance ne prédisent pas autant la prise de décision que la confiance apprise dynamique et la confiance situationnelle.

Notre deuxième hypothèse est validée : **La transparence et la criticité de la situation sont corrélées avec l'acceptation d'une proposition**. Notre troisième hypothèse est partiellement validée : **La transparence et la criticité de la situation sont corrélées avec les composantes de la confiance (confiance apprise, dynamique et situationnelle)** (Figure 7.11). Cela permet d'éclairer le second verrou de recherche. La transparence sur la criticité de la situation va corriger la prise de décision du participant en fonction du risque de la situation. Cet effet de la transparence se retrouve également sur les composantes de la confiance.

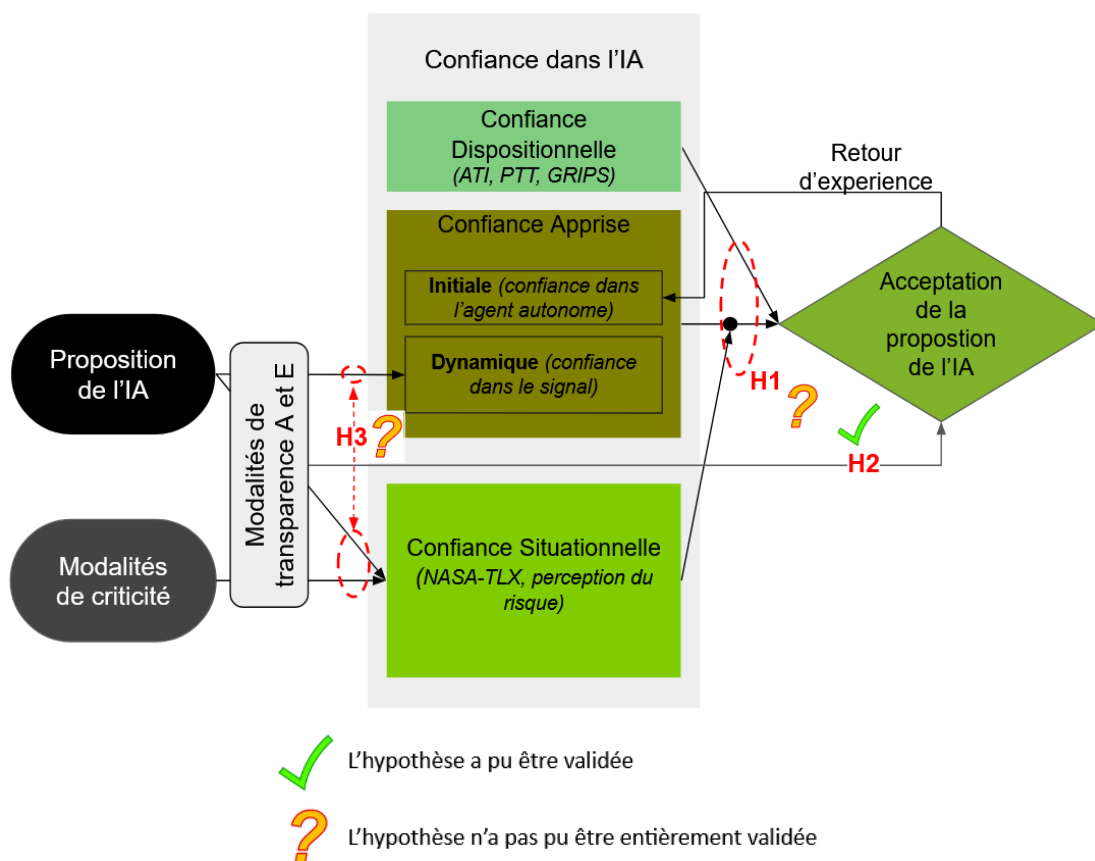


Figure 7.11 : Validation et synthèse des hypothèses initiales

La transparence a un effet direct sur l'acceptation d'un opérateur. En milieu industriel, il semble important que les systèmes cyber-physique (CPS) soient transparents sur les risques liés à leur proposition. De cette façon, l'opérateur pourra prendre la décision d'accepter ou de refuser en ayant une conscience de la situation adaptée. Cela pourra également éviter les phénomènes de surconfiance et de sous-confiance. La transparence sur le risque voit également ses effets modérés en fonction du type de situation, il est tout aussi important d'être transparent sur le risque quand la situation n'en comporte pas (promouvant l'usage du CPS) que lors d'une situation à risque (modérant l'usage du CPS).

Le sujet de la transparence dans la coopération humain-machine nécessite encore des recherches pour améliorer la compréhension de son rôle et de ses effets. Les prochaines études devraient s'intéresser de façon précise aux effets de la transparence en la découpant selon le sujet sur lequel porte l'information transmise en plus des dimensions actuellement usitées dans le domaine.

Pour faire suite à cette expérimentation nous souhaitons explorer la possibilité de faire varier les dimensions de la transparence sur un seul type de situation. L'objectif serait de valider l'utilisation de la transparence comme capacité adaptative du CPS dans le cadre d'une coopération humain-machine.

Chapitre 8 : Adaptation dynamique de la transparence pour calibrer la confiance

Ce troisième chapitre expérimental a pour objectif d’explorer la possibilité d’utiliser la transparence de façon dynamique pour aider à la coopération humain-machine. Ce chapitre permettra de répondre à notre troisième verrou scientifique portant sur **la compréhension des effets d’une transparence adaptative** (section 5.4).

La proposition de l’intelligence artificielle (IA) de la même façon que l’expérimentation précédente, sera *de niveau tactique* sur la planification de la maintenance. Contrairement aux deux expérimentations précédentes nous nous intéressons ici à une coopération où les *situations sont dépendantes* entre elles. L’objectif était de proposer une *situation d’interaction sur le long-terme* afin de simuler un environnement plus écologique.

Nous regardons ici l’effet de la transparence sur l’acceptation, et plus particulièrement si les changements de transparence influencent les changements de comportement (Akash et al., 2020a ; Akash et al., 2020b). Nous explorons également l’effet des retours d’expérience sur la prise de décision de l’opérateur (Javaid & Estivill-Castro, 2021). Comme souligné par Parasuraman et Riley (1997), il peut arriver que l’opérateur soit en situation de surconfiance ou de sous-confiance. Par rapport à la prise de décision assistée par un système cyber-physique (CPS) cela reviendrait à accepter lorsqu’il est préférable de refuser (surconfiance) et refuser lorsqu’il faudrait accepter (sous-confiance). L’implantation d’une capacité de transparence adaptative chez les CPS semble être une voie prometteuse pour calibrer la confiance d’un opérateur (Baker et al., 2018 ; Dizaji & Hu, 2021) et éviter la surconfiance et la sous-confiance. La sensibilité à une transparence adaptative pourrait dépendre de la confiance dispositionnelle et nécessite donc d’être investiguée (Kortschot et al., 2022).

8.1. Méthode

8.1.1. Participants

Cette expérimentation a eu lieu à l’ENSM (école nationale supérieure maritime) au Havre en mars 2022. Le recrutement a été fait en interne grâce à l’envoi de mail et il n’y avait pas de rétribution prévue pour les participants. Cette expérimentation a eu lieu dans le cadre

d'un contrat de prestation de service entre le Lab-Sticc et l'ENSM, c'est pourquoi les enseignants ont incité leurs étudiants à s'inscrire. Les participants étaient conviés à une expérimentation sur la maintenance prévisionnelle dans laquelle ils devaient coopérer avec une intelligence artificielle. Il y avait 16 participants ($M = 25,31$ ans, $E-T = 9,85$). Tous les participants ont déjà navigué ($M = 37,31$ mois ; $E-T = 118,09$). Sur les 16 participants, 15 sont des élèves de l'ENSM et 1 participant est un ancien chef mécanicien.

8.1.2. Protocole expérimental

Dans cette expérimentation les participants sont amenés à résoudre 60 situations. Chaque situation correspond à une prise de décision, accepter ou refuser, sur une proposition émanant de l'IA Seanatic (Figure 8.1). Les situations sont toutes de criticité élevée (cf. Tableau 7.2). Les 60 situations portent sur le fait de décaler une maintenance d'un équipement avec une forte gravité s'il y a un défaut de maintenance et ce avec une forte probabilité de survenue (i.e. un changement de filtre). Nous choisissons cette situation pour que les participants soient dans un cas où accepter la proposition représente un risque important. Selon les sujets experts interrogés lors de la création des situations expérimentales, cette situation est celle qui serait la plus probablement refusée en milieu réel au vu des risques encourus (section 7.1.2).

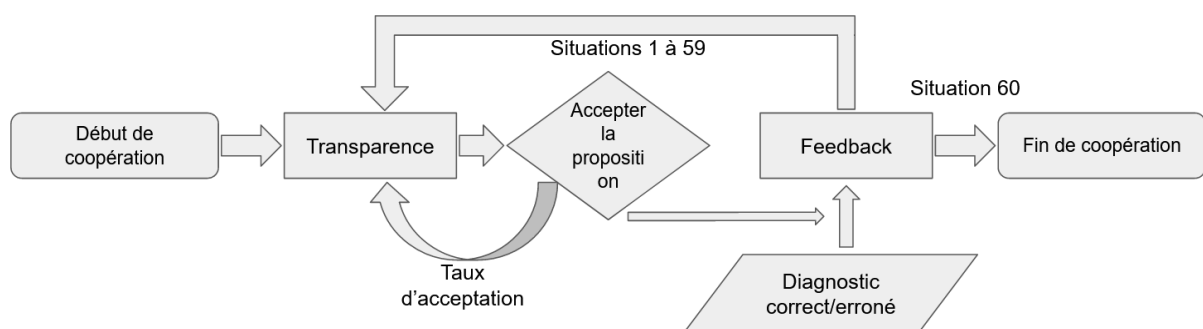


Figure 8.1 : Protocole expérimental

Pour que la fiabilité de l'IA (i.e. 90%) corresponde à une réalité écologique, cinq des situations se basent sur un diagnostic erroné (i.e. le filtre s'encrasse plus tôt que prévu). Ces cinq situations ne peuvent pas être différenciées des autres et sont programmées pour arriver au même moment pour tous les participants (lors des situations 8, 21, 36, 48 et 57). L'apparition de ces situations au cours de l'expérimentation a été choisie sur la base de deux

critères. Le premier critère est que les participants soient face à une proposition basée sur un diagnostic erroné toutes les 10 situations en moyenne. Deuxièmement, un critère de répartition dans le sens où nous voulons éviter que les participants soient confrontés de manière successive à plusieurs situations avec diagnostic erroné.

Entre chaque situation une ellipse temporelle est simulée. Cette ellipse tient compte du temps de vie de l'équipement et du calendrier de gestion de maintenance assistée par ordinateur (GMAO). Si les participants refusent, la date GMAO est utilisée, et l'ellipse est de 14 jours avant la prochaine situation. Si les participants acceptent, le temps de vie de l'équipement est prolongé, et l'ellipse est de 18 jours avant la prochaine situation.

La transparence de l'IA varie selon le taux d'acceptation du participant. L'IA peut être transparente soit : sur la fiabilité de sa proposition (**dimension analytique**), soit sur les risques liés à cette proposition (**dimension environnement**), soit sur les deux informations en même temps (Figure 8.2).

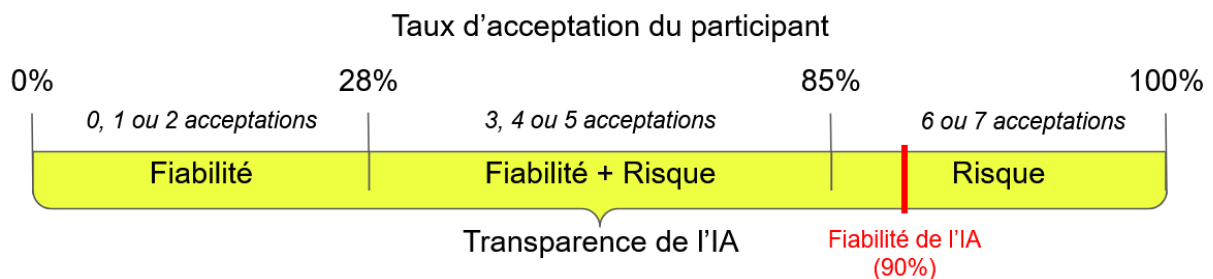


Figure 8.2 : Transparence en fonction du taux d'acceptation

Le taux d'acceptation est calculé, en pourcentage, selon la décision du participant sur les 7 situations précédentes. Ces taux peuvent être considérés comme des paliers. Nous choisissons de calculer le taux d'acceptation sur 7 situations afin qu'il soit proche de la fiabilité de l'IA (i.e. 90%) (Figure 8.2). A chaque fois que le participant accepte la proposition de l'IA, cela augmente le taux d'acceptation de 14%. Nous voulons également un nombre équilibré de taux d'acceptation pour chaque condition. Chaque condition de transparence de l'IA est ainsi activée pour 2 ou 3 taux d'acceptation. Si sur ses 7 dernières décisions le participant accepte 6 fois ou plus (taux de plus de 85%), nous considérons cela comme de la surconfiance. En conséquence l'IA communique uniquement sur les risques de la situation. A l'inverse, si le participant accepte 2 fois (taux de moins de 28%) ou moins nous considérons cela comme de

la sous-confiance. En conséquence, l'IA communique uniquement sur sa fiabilité. Si le participant accepte 3, 4 ou 5 fois sur les 7 situations précédentes (taux compris entre 28% et 85%), l'IA communique sa fiabilité et les risques de sa proposition (Figure 8.2).

Après son choix, le participant reçoit un feedback (Tableau 8.1).

Tableau 8.1 : Feedback de l'IA en fonction de la justesse du diagnostic de l'IA et du choix du participant

Proposition de l'IA (situation à criticité élevée)	Choix du participant (repousser ou non un changement de filtre)	Feedback de l'IA (conséquence)
Basée sur un diagnostic correct (55 propositions)	Accepter	Feedback positif « <i>Le filtre a tenu sans conséquences négatives jusqu'à la date proposée par Seanatic</i> »
	Refuser	Feedback négatif « <i>Après inspection suite à son changement, le filtre aurait pu tenir sans conséquences négatives jusqu'à la date proposée par Seanatic</i> »
Basée sur un diagnostic erroné (5 propositions)	Accepter	Feedback négatif : « <i>Le filtre était encrassé lors de votre maintenance</i> »
	Refuser	Feedback positif « <i>Le filtre présentait les traces d'un début d'encrassement lors de votre maintenance</i> »

Ce feedback textuel peut être positif ou négatif en fonction de la décision du participant (i.e. accepter ou refuser la proposition) et de la justesse du diagnostic de l'IA (i.e. diagnostic correct ou erroné). Il porte sur les conséquences de la prise de décision du participant sur l'état de l'équipement lors de la réalisation de la maintenance. Le rôle de ce feedback est de rendre la situation de coopération écologique en créant une boucle de rétroaction entre la décision et la construction de la confiance de l'opérateur envers l'IA. Ce feedback permet aux participants d'avoir les conséquences de leur prise de décision, et donc d'inférer sur s'ils peuvent faire confiance à l'IA. Selon Hoff et Bashir (2015), la confiance apprise initiale (CAI)

se construit sur les expériences passées avec le CPS et sur le résultat de la coopération (i.e. le feedback). La CAI, au même titre que les autres composantes de la confiance, vient ensuite influencer la prise de décision de l'opérateur. La présence de ce feedback permet de calibrer, à posteriori, la confiance du participant dans l'IA, en ayant des éléments permettant de définir si cette IA fait des propositions qui sont sans danger pour les opérations de maintenance.

Déroulé de la passation. Le participant était accueilli dans la salle de simulateur de l'ENSM (Figure 8.3). Il lui était ensuite présenté l'objectif de l'étude et son déroulement avant de lui demander de remplir la feuille de consentement libre et éclairée (Annexe K).



Figure 8.3 : Installation expérimentale

Suite à cela il répondait aux questionnaires pré-test portant sur l'affinité envers les technologies (Annexe D.1), la propension à faire confiance aux technologies (Annexe D.2), la propension au risque (Annexe D.3) et un questionnaire de confiance apprise initiale (CAI) (Annexe E.1). Après une présentation de la maintenance prévisionnelle et du logiciel, le participant pouvait commencer le scénario. A la fin du scénario un second questionnaire portant sur sa CAI lui était présenté (Annexe E.1). Pour finir, le participant répondait à des questions socio-démographiques (Annexe L).

8.1.3. Matériel

Pour la réalisation de cette expérimentation les participants ont accès à un PC, un écran et une souris. Les questionnaires sont remplis grâce à l'outil Limesurvey. Les phases de coopération sont réalisées grâce au logiciel Seanatic (6.1.3.1 : Interface, Chapitre 7). L'outil Seanatic est simulé via un powerpoint qui reprend l'interface de dialogue (Figure 8.4) et l'interface des données (Figure 8.5). Les fonctionnalités du powerpoint ont été créées sous VBA (accepter/refuser des propositions, changer de pages, etc.). L'interface de dialogue affiche différentes informations selon le taux d'acceptation du participant. Ce taux d'acceptation vient modifier la modalité de transparence activée. L'IA est capable d'utiliser trois modalités de transparence : transparente sur la fiabilité (**dimension analytique**, en vert dans la Figure 8.4), sur le risque (**dimension environnement**, en rouge dans la Figure 8.4) ou sur les deux à la fois (comme sur la Figure 8.4).

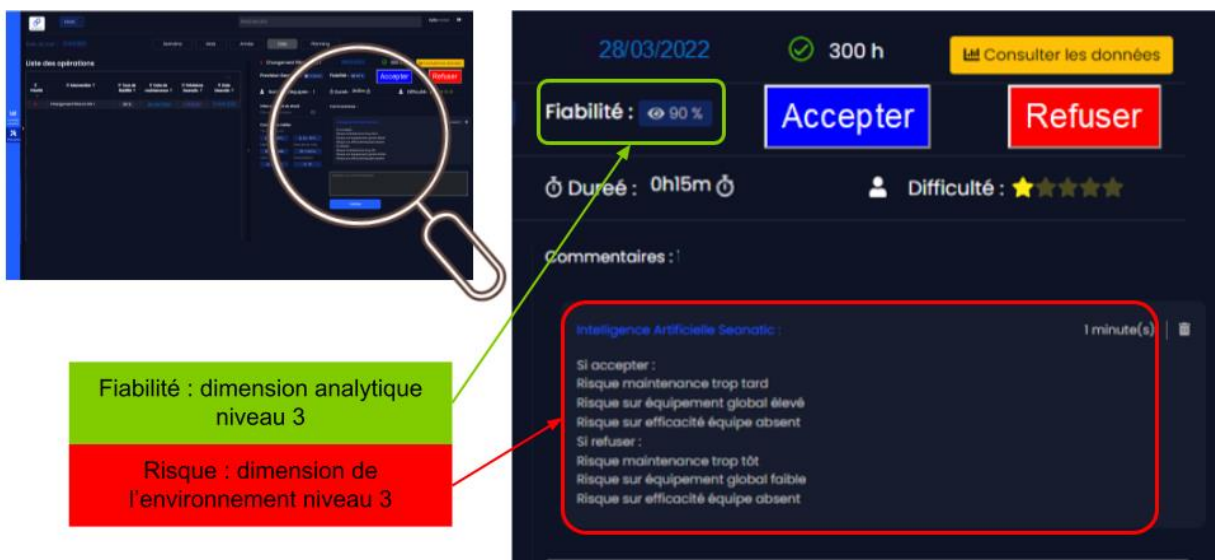


Figure 8.4 : Interface de dialogue Seanatic

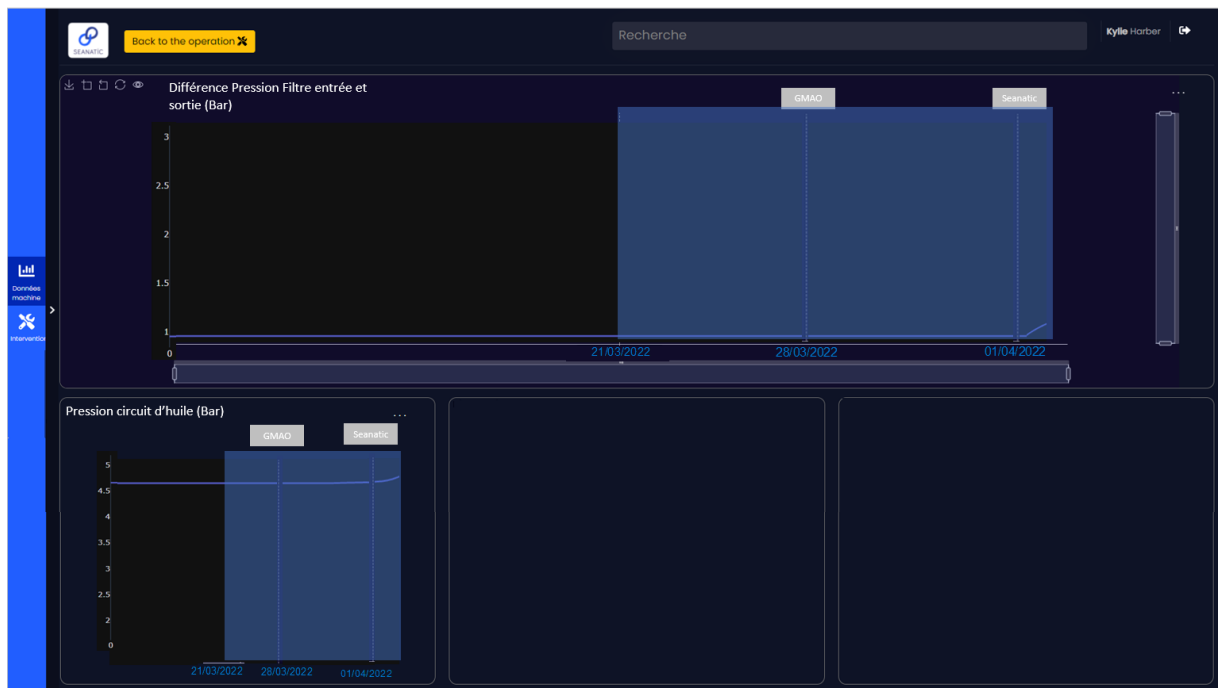


Figure 8.5 : Données utilisées par l'IA Seanatic

8.1.4. Observables, méthode d'analyse des résultats, et hypothèses

Nous mesurons une variable de façon objective :

- **L'acceptation** - mesurée à chaque situation suite à la décision du participant (accepter ou refuser)

Nous mesurons quatre variables de façon subjective :

- **La confiance dispositionnelle (CD) :**
 - L'affinité envers la technologie du participant grâce à l'échelle ATI (traduction française de Franke et al., 2019) (Annexe D.1)
 - La propension du participant à faire confiance à la technologie grâce à l'échelle PTT (traduction française de Jessup, 2018) (Annexe D.2)
 - La propension générale au risque du participant grâce à l'échelle GriPS (traduction française de Zhang et al., 2019) (Annexe D.3)
- **La confiance apprise initiale (CAI)** du participant - mesurée au début puis à la fin avec l'échelle IMOTRIS (traduction française de Mayer et al., 1995, Lyons & Guznov, 2019) (Annexe E.1)

Nous répondons à **troisième verrou** au travers l’hypothèse exploratoire suivante (Figure 8.6) :

H. Une IA avec une capacité de transparence adaptative modifierait l’acceptation des opérateurs ? (Akash et al., 2020a ; Akash et al., 2020b ; Okamura & Yamada, 2020) (**Verrou 3**)

Nous explorons en premier lieu le lien entre la confiance dispositionnelle, la confiance apprise initiale et la coopération avec une IA à transparence adaptative et vérifions l’existence de groupes différenciés (Kortschot et al., 2022). Nous explorons également l’effet des changements de feedbacks sur les changement d'acceptation (Hoff & Bashir, 2015).

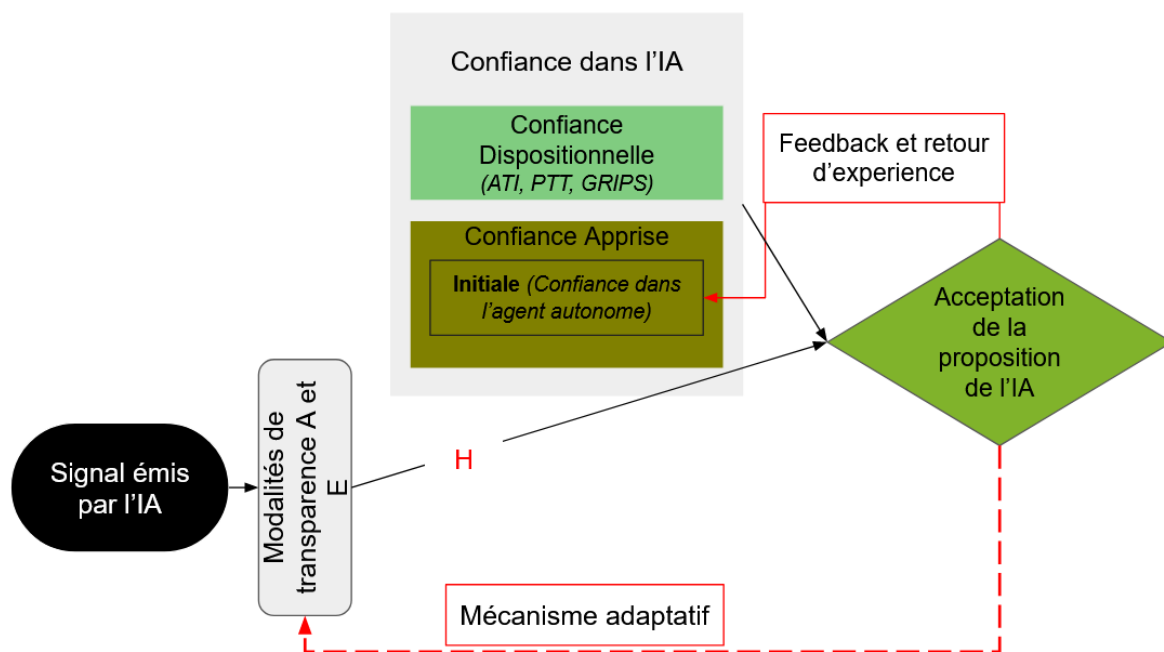


Figure 8.6 : Positionnement des hypothèses, des observables et des variables dans le cadre de la prise de décision assistée par une IA

La transparence est codée comme suit :

Transparence
Fiabilité = F
Fiabilité + Risque = F+R
Risque = R

Les changements d’acceptation (passer d’une acceptation à un refus), de transparence (comme passer d’une transparence « F+R » à « R ») et de feedback (comme passer d’un feedback positif à un feedback négatif) sont nommés « Delta ». Ces différents deltas sont

explicités dans la Figure 8.7. L'absence de delta est illustrée par « none ». Nous entendons par delta un changement d'état soit de la transparence, soit du feedback, soit du type de réponse du participant (i.e. changement d'acceptation).

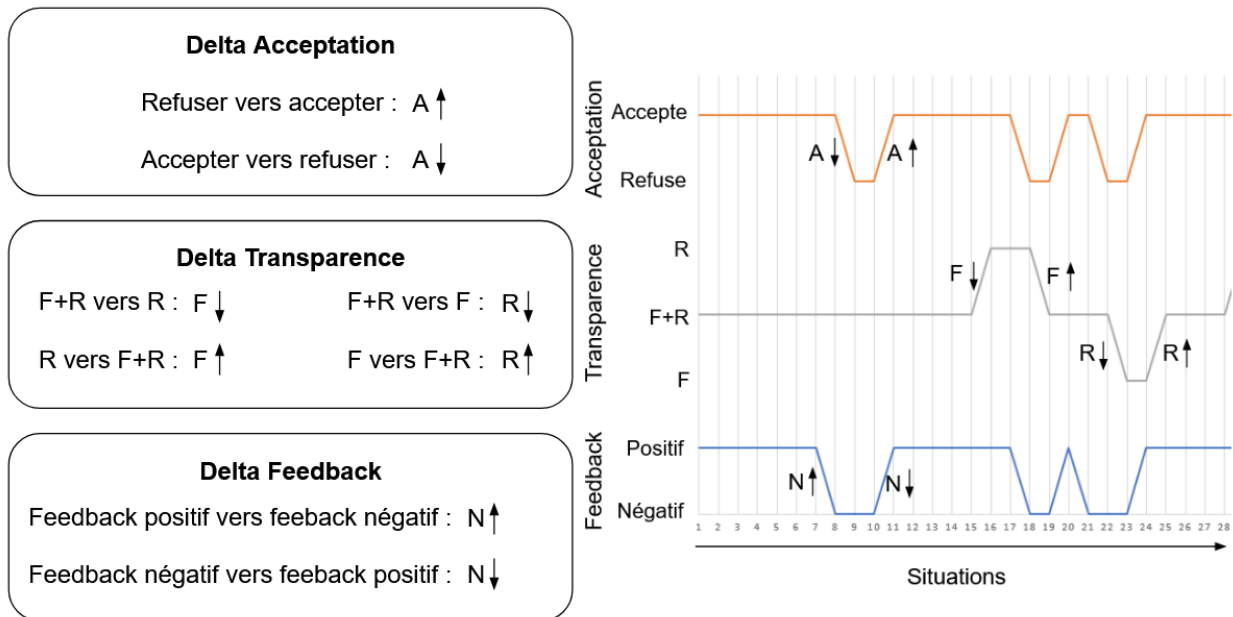


Figure 8.7 : Codage des deltas et application à un chronogramme d'un participant

Nous utilisons des T-Test de Student pour analyser la confiance dispositionnelle. Nous utilisons une régression logistique multinomiale pour analyser les effets des deltas de transparence et de feedback sur le delta d'acceptation.

8.2. Résultats

8.2.1. Relations entre la confiance dispositionnelle, la confiance apprise initiale et la coopération avec une IA à transparence adaptative

Au sein de notre effectif de participants, nous voyons émerger deux types de comportements face aux propositions de l'intelligence artificielle. Une partie des participants (9 d'entre eux) ont tendance à conserver la même décision (accepter les propositions) tout au long de l'expérimentation (Figure 8.8, haut). L'autre partie des participants (7) modifie leurs décisions au gré des situations, entre accepter et refuser, suite aux propositions de l'IA (Figure 8.8, bas). Nous scindons les participants en deux groupes sur la base de 10% de refus. Ces 10% correspond à la marge d'erreur de l'IA (dont la fiabilité est 90%). Le premier groupe est «

constant » (participants 5, 6, 8, 9, 10, 11, 13, 15, 16) et le second « **versatile** » (participants 1, 2, 3, 4, 7, 12 et 14). Dans le groupe « versatile » le taux d'acceptation varie entre 10% et 27%.

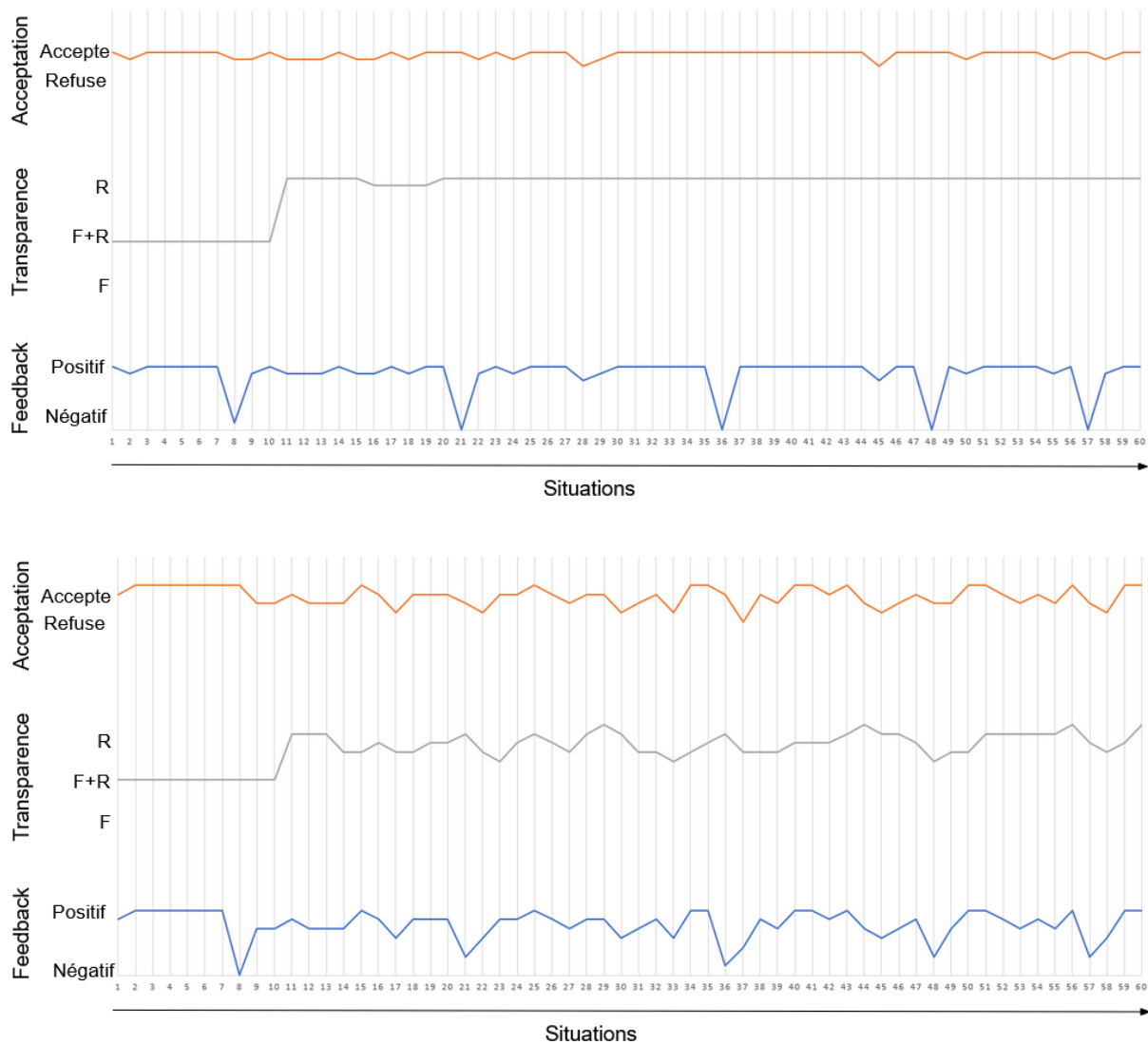


Figure 8.8 : Chronogramme moyen des participants du groupe « Constant » (haut) et « Versatile » (bas)

Confiance dispositionnelle et acceptation. Pour comprendre ce qui peut expliquer ces deux types de comportement nous nous intéressons aux différences potentielles qui existent du point de vue de la confiance dispositionnelle. **La confiance dispositionnelle semble avoir une influence sur l'utilisation d'une IA à transparence adaptative.** Nous utilisons des tests t-student afin de comparer les deux groupes. Nous constatons que la moyenne du groupe « versatile » ($M = 4.41$, $E-T = 0.98$) est plus élevée sur les échelles d'affinité envers la technologie que le groupe « constant » ($M = 4.39$, $E-T = 0.74$). Cette différence n'est pas significative ($t(14) = 0.041$, $p = .484$). Nous constatons également que la propension à faire

confiance à la technologie du groupe « versatile » ($M = 3.62$, $E-T = 0.82$) et la propension au risque ($M = 3.32$, $E-T = 1.10$) sont plus faibles que ceux du groupe « constant » (respectivement $M = 3.74$, $E-T = 0.30$ et $M = 3.83$, $E-T = 1.53$). Ces différences ne sont pas significatives (respectivement $t(14) = 0.414$, $p = .452$ et $t(14) = 0.748$, $p = .467$).

Confiance apprise initiale et acceptation. Les échelles de confiance remplies par les participants montrent qu'en moyenne le groupe « versatile » a une confiance plus faible ($M = 3.71$, $E-T = 1.08$) avant l'interaction que le groupe « constant » ($M = 3.82$, $E-T = 0.64$) (Figure 8.9). Cette différence n'est pas significative ($t(14) = 0.258$, $p = .8$). Nous constatons également qu'après l'interaction c'est le groupe « versatile » qui a une confiance plus élevée ($M = 4.04$, $E-T = 1.24$) par rapport au groupe « constant » ($M = 3.6$, $E-T = 1.28$). Cette différence n'est pas significative ($t(14) = 0.709$, $p = .46$). Nous constatons également que le groupe « versatile » a une augmentation de la confiance ressentie entre avant et après l'interaction ($M = 3.71$, $E-T = 1.08$ et $M = 4.04$, $E-T = 1,24$). A l'inverse, le groupe « constant » a une baisse de la confiance ressentie entre avant et après l'interaction ($M = 3.82$, $E-T = 0.64$ et $M = 3.6$, $E-T = 1.28$) (Figure 8.9). Ces différences ne sont pas significatives (respectivement $t(6) = 1.211$, $p = .27$ et $t(8) = 0.865$, $p = .41$).

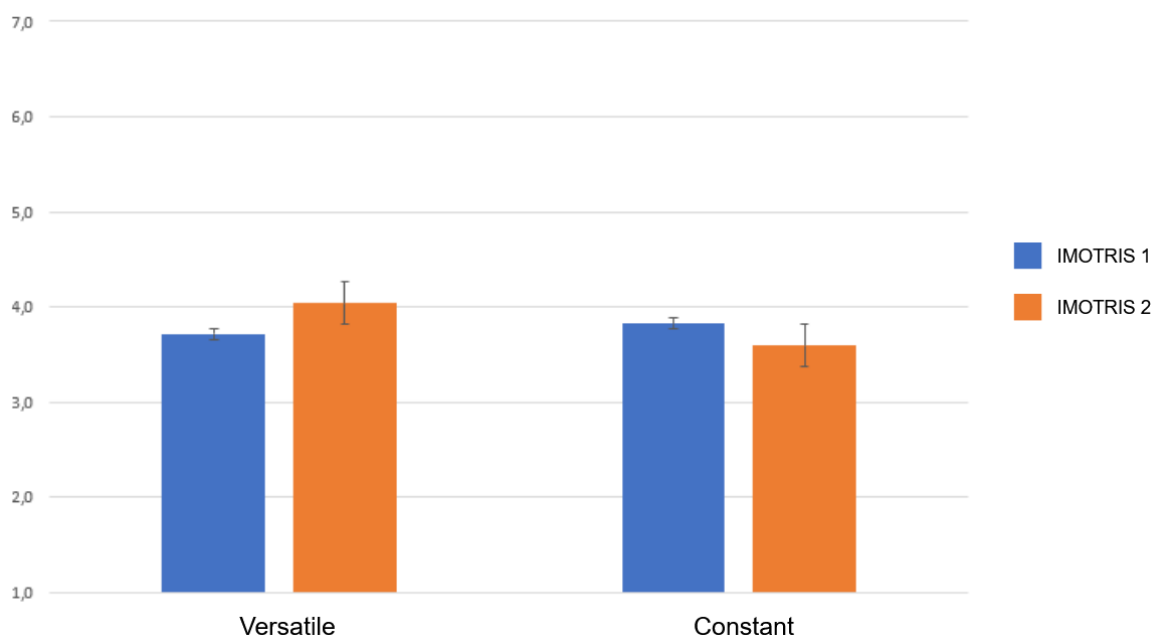


Figure 8.9 : Moyennes des groupes « versatile » et « constant » sur les échelles IMOTRIS 1 et 2

Les participants du groupe « constant » ayant toujours accepté, ce groupe ne rentre pas dans notre cadre d'intérêt. Ils ne permettent pas de comprendre les effets des deltas de

transparence ou de feedback sur un changement de comportement. Nous les excluons des prochaines analyses, et nous nous concentrerons sur le groupe « versatile » (pour rappel l'effectif de ce groupe est de 7 participants).

8.2.2. Relations entre changement de transparence, changement de feedback et acceptation

Nous utilisons ici une régression multinomiale par étape afin de conserver le meilleur modèle. Celui-ci inclut deux effets principaux liés aux deltas de transparence et aux deltas de feedback, mais pas d'interaction entre ces deltas.

Relations entre les changements de transparence et les changement d'acceptation. Nous constatons un effet principal d'un delta de transparence (Figure 8.10 et Tableau 8.2). **Lorsque l'IA n'est plus transparente sur la dimension analytique (F↓), cela a un effet négatif sur l'acceptation.**

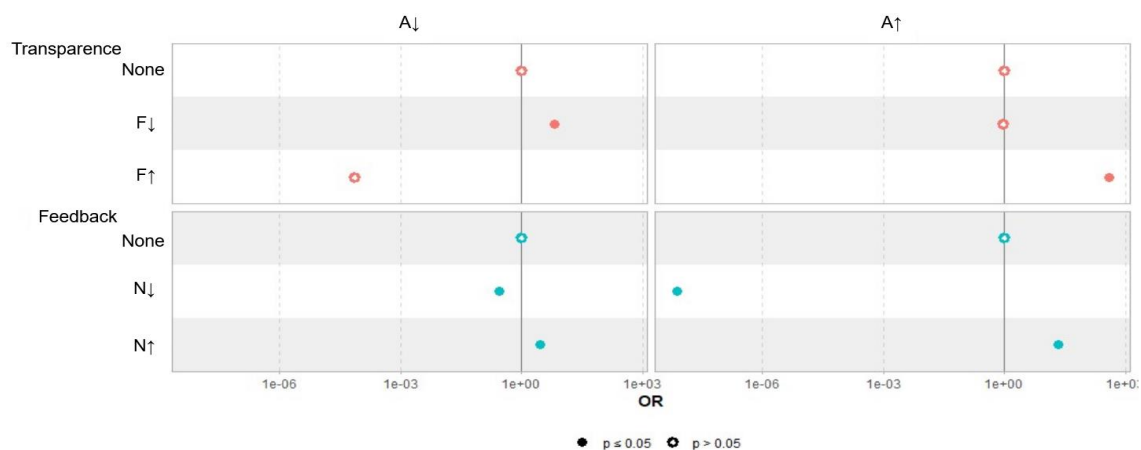


Figure 8.10 : Effet du delta transparence et du delta feedback sur le delta d'acceptation (Régression logistique multinomiale)

L'utilisation de la régression logistique multinomiale nous permet de constater que lorsque l'IA n'est plus transparente sur sa fiabilité (i.e. F↓), il y a plus de probabilité que le participant change sa décision et se mette à refuser (i.e. A↓) par rapport aux valeurs de références (absence de delta de transparence et de delta d'acceptation ($OR = 6,26, p < .001$)). A l'inverse, **lorsque l'IA est de nouveau transparente sur la dimension analytique (F↑), cela a un effet positif sur l'acceptation.** Lorsque l'IA est de nouveau transparente sur sa fiabilité (i.e. F↑), il y a plus de probabilité que les participants se mettent à accepter (i.e. A↑) par rapport aux valeurs de référence ($OR = 372, p < .001$).

Tableau 8.2 : Effet du delta transparence et du delta feedback sur le delta d'acceptation

Modalités	OR ¹	95%IC ²	<i>p</i> -value
A↓			
Delta Transparence (réf. None)			
F↓	6.26	2.92, 13.4	<.001***
Delta Feedback (réf. None)			
N↓	0.28	0.09, 0.83	0.022*
N↑	2.79	1.22, 6.39	0.015*
A↑			
Delta Transparence (réf. None)			
F↓	0.90	0.10, 7.83	0.9
F↑	372	44.1, 3,140	<.001***
Delta Feedback (réf. None)			
N↑	21.6	8.40, 55.6	<.001***

¹OR = Odd Ratio, ²IC = intervalle de confiance

Relations entre les changements de feedback et les changement d'acceptation. Nous constatons également un effet principal du delta de feedback. **Un feedback qui devient négatif (N↑) a un effet positif sur l'acceptation.** Lorsque le feedback passe de positif à négatif (i.e. N↑), les probabilités que le participant se mette à accepter (i.e. A↑) augmentent par rapport aux valeurs de référence ($OR = 21.6, p < .001$). Pareillement, **ce changement de feedback a également un effet positif sur le refus.** Les probabilités que le participant se mette à refuser (i.e. A↓) augmentent par rapport aux valeurs de références ($OR = 2.79, p = .015$). **Les probabilités que le participant se mette à accepter sont cependant plus élevées que les probabilités qu'il se mette à refuser (Tableau 8.2).**

8.3. Discussion

L'objectif de cette expérimentation était d'utiliser la transparence de façon adaptative afin de calibrer la confiance d'un opérateur. Nous avons constaté l'apparition de deux groupes aux réactions différentes face à une IA à transparence adaptative. Nous avons constaté que le groupe « versatile » modifie ses prises de décision avec une IA à la transparence adaptative, quand le groupe « constant » ne l'est pas. Cela implique que si l'usage de la transparence de façon adaptative est prometteur, il est nécessaire de prendre en compte la différence inter-individuelle. Nous avons pu montrer qu'il était possible de faire varier la transparence d'une IA pour modifier la prise de décision de l'opérateur. Une IA qui n'est plus transparente sur sa dimension analytique va amener un refus de sa proposition. Une IA dont le feedback devient négatif va amener le participant à changer de décision par rapport à sa décision précédente.

Relations entre la confiance dispositionnelle et une IA à transparence adaptative.

Nous avons constaté qu'une IA avec une transparence adaptative ne provoquait pas la même réaction chez tous les participants. Ce résultat est semblable à ce qu'avait pu observer Pynadatah et al. (2022) et Kortschot et al. (2022). **Les résultats présentés ici ne permettent cependant pas de statuer avec certitude sur l'effet de la confiance dispositionnelle sur la coopération avec un CPS à transparence adaptative.** En effet, dû au faible nombre de participants, les analyses statistiques ne permettent pas d'identifier si les différences relevées sont dues au hasard, ou si elles correspondent à une réalité généralisable. Nous pouvons constater que les différences entre les groupes vont dans le même sens. Ainsi, les participants qui ont le plus eu de changement dans leurs comportements sont également ceux qui avaient le plus d'affinité envers la technologie, et donc le plus de connaissances sur les potentielles limites. Ces participants avaient également une propension à faire confiance à la technologie plus faible, ce qui expliquerait le fait d'avoir un taux d'acceptation plus faible que les participants constants. Ce groupe avait également une propension au risque plus faible, ce qui pourrait également expliquer leur réaction face à des situations qui étaient à criticité élevée, ces situations pouvant amener à prendre un risque. Ce constat est d'autant plus cohérent dans les situations où l'IA communiquait exclusivement sur le risque. Ce lien entre la personnalité et la confiance est en accord avec le constat de Bhat, Lyons, Shi et Yang (2022)

qui ont également trouvé un lien entre la personnalité et la dynamique de la confiance entre un individu et la machine.

Nous constatons également que les participants réceptifs à une transparence adaptative avaient la moyenne de confiance la plus basse avant de débiter l'expérimentation, et la moyenne la plus haute à la fin de l'expérimentation. Ce groupe de participants a peut-être acquis une meilleure connaissance des limites de l'IA en variant leurs réponses, et dès lors peut se permettre d'y placer une plus grande confiance. Cette même confiance qui était plus faible au début les a peut-être poussés à refuser plus facilement lorsque l'IA n'était plus transparente sur sa fiabilité. Cela pourrait avoir pour effet que leur modèle mental s'est amélioré et que leur confiance a pu être calibrée. Une étude avec un plus grand nombre de participants pourrait étayer ces hypothèses.

Ces résultats permettent d'apporter une réflexion sur notre troisième verrou. **Nous constatons que tous les opérateurs ne réagissent pas de la même façon à un CPS avec une capacité de transparence adaptative.** Nous voyons donc que la solution d'utiliser la transparence de façon adaptative ne convient pas pour tous les opérateurs.

Relations entre les changements de transparence, les changements de feedback et les changements d'acceptation. Les résultats liés aux changements de comportement de l'opérateur nous permettent de constater une corrélation avec les situations où l'IA modifie sa transparence. Lorsque l'IA ne communique plus sur sa fiabilité, les opérateurs « versatile » modifient leur comportement et se mettent à refuser. A l'inverse, lorsque l'IA communique à nouveau sur sa fiabilité, les participants modifient leur comportement et se mettent à accepter. Conformément aux études précédentes (Akash et al., 2020b ; Okamura & Yamada, 2020) **un opérateur qui coopère avec une IA en capacité d'adapter sa transparence modifie son comportement en fonction des changements de transparence.** Pour les participants du groupe « versatile », c'est la perte ou l'ajout de la transparence qui étaient utilisés pour calibrer leur décision et l'accorder en fonction des informations à leur disposition. En occultant la transparence sur la dimension analytique (i.e. la fiabilité) cela accentue le poids de la dimension « environnement » sur la décision de l'opérateur. A l'inverse, **lorsque l'IA communique sur deux dimensions, l'opérateur prend en compte ces deux dimensions pour prendre sa décision.** L'utilisation de l'IA pouvant modifier leur transparence de façon

dynamique est de ce fait, une voie intéressante à explorer de sorte que les IA mettent l'accent sur une dimension ou sur une autre s'ils perçoivent que la confiance de l'opérateur n'est pas appropriée.

Concernant l'effet des feedbacks, le modèle issu de la régression logistique multinomiale prévoit qu'un changement de feedback, positif vers négatif, augmente les probabilités de changements de comportements de l'opérateur. Il existe deux cas de figure :

- Premier cas de figure, le **changement de feedback fait passer l'opérateur d'un refus à une acceptation**. Dans ce cas l'opérateur a refusé une proposition basée sur un diagnostic correct de l'IA. Son refus fait passer le feedback de positif à négatif. L'opérateur se met alors accepter voyant qu'il est préférable de choisir ce comportement pour avoir un feedback positif. Dans cette configuration, l'opérateur a probablement voulu tester l'IA en refusant une proposition, et a constaté qu'il était préférable de faire confiance à l'IA.
- Second cas de figure, le **changement de feedback fait passer l'opérateur d'une acceptation à un refus**. Ici, les opérateurs ont accepté une proposition basée sur un diagnostic erroné. L'erreur de diagnostic de l'IA fait passer le feedback de positif à négatif. Pour les opérateurs cela peut amener à une perte de confiance dans l'IA. En conséquence, ils considèrent comme probable que l'IA refasse une erreur, et se mette à refuser sa proposition. Contrairement à Salem et al. (2015) nous constatons un effet de l'erreur de l'IA sur la prise de décision de l'opérateur. Nous pensons que cela est dû au contexte maritime qui a pour objectif d'éviter toute erreur de maintenance (Simon et al., 2021a). Dans cette situation l'importance d'une erreur est telle qu'elle vient modifier la prise de décision de l'opérateur, et ce malgré la fiabilité élevée de l'IA.

Les résultats présentés ici montrent que ce sont les changements de transparence et les changements de feedback qui amènent des changements de prise de décision. Concernant notre troisième verrou, nous pouvons donc dire qu'il **est possible d'utiliser la transparence adaptative pour calibrer la confiance**.

Limites. Nous relevons plusieurs limites à cette étude. Premièrement, l'algorithme d'adaptation est assez simple et ne se base que sur le taux d'acceptation. Se pose donc la question de ce qu'il se passerait sur une plus longue durée (et avec une IA qui adapterait sa

transparence moins régulièrement), ou au contraire sur des échéances plus courtes (et avec une IA adaptant sa transparence plus régulièrement). L'idée serait d'explorer la différence entre un algorithme se basant sur la dernière situation ou sur toutes les situations. Également, le faible échantillon ne permet pas tirer de conclusions sur le rôle de la confiance dispositionnelle dans la coopération avec une IA à transparence adaptative. La fiabilité très haute (i.e. 90%) de l'IA a pu pousser certains participants à accepter toutes les propositions, limitant l'effet d'une transparence adaptative. Il serait opportun de moduler cette fiabilité pour constater quels sont les effets sur l'utilisation de la transparence de façon adaptative. Enfin, nous avons fait varier la transparence de l'IA sur deux dimensions. Il semble nécessaire d'utiliser d'autres dimensions de la transparence pour observer leur potentiel en tant que capacité adaptative, et ce par rapport à d'autres déclencheurs.

Cette étude montre que la confiance dispositionnelle semble jouer un rôle important dans la coopération avec un CPS à transparence adaptative. Nous montrons également qu'il est possible d'utiliser les dimensions d'un même niveau de transparence pour calibrer la confiance. L'utilisation du taux d'acceptation comme déclencheur des mécanismes d'adaptation de la transparence du CPS est une voie à explorer. Ces résultats pourraient permettre, dans un contexte industriel, d'éviter que les opérateurs se trouvent dans une situation de complaisance en les faisant se questionner sur leur choix d'accepter ou de refuser les propositions d'une IA.

8.4. Conclusion

Cette troisième expérimentation a permis de montrer que la calibration de la confiance grâce à une IA à transparence adaptative semble dépendante de la confiance dispositionnelle. La confiance dispositionnelle (l'attitude envers la technologie, la propension à faire confiance à la technologie et la propension au risque) sous-tend le changement, ou non, de prise de décision face à un CPS à transparence adaptative. Il est possible d'utiliser la transparence de façon adaptative. Une IA modulant sa transparence selon l'acceptation de l'opérateur permet de calibrer la confiance de celui-ci. Les changements de transparence peuvent conduire à des variations de comportement, et ce indépendamment d'un changement de feedback. Cependant, nous avons vu que l'utilisation de la transparence de façon adaptative n'est pas généralisable à tous les opérateurs.

Une réponse à notre question exploratoire émerge par nos résultats : **Les changements de transparence modifient l'acceptation de la proposition** (Figure 8.11). Ces résultats permettent d'éclairer le troisième verrou. La transparence peut être utilisée comme une capacité adaptative du CPS afin de calibrer la confiance de l'opérateur. Cependant, son efficacité semble dépendre des facteurs individuels.

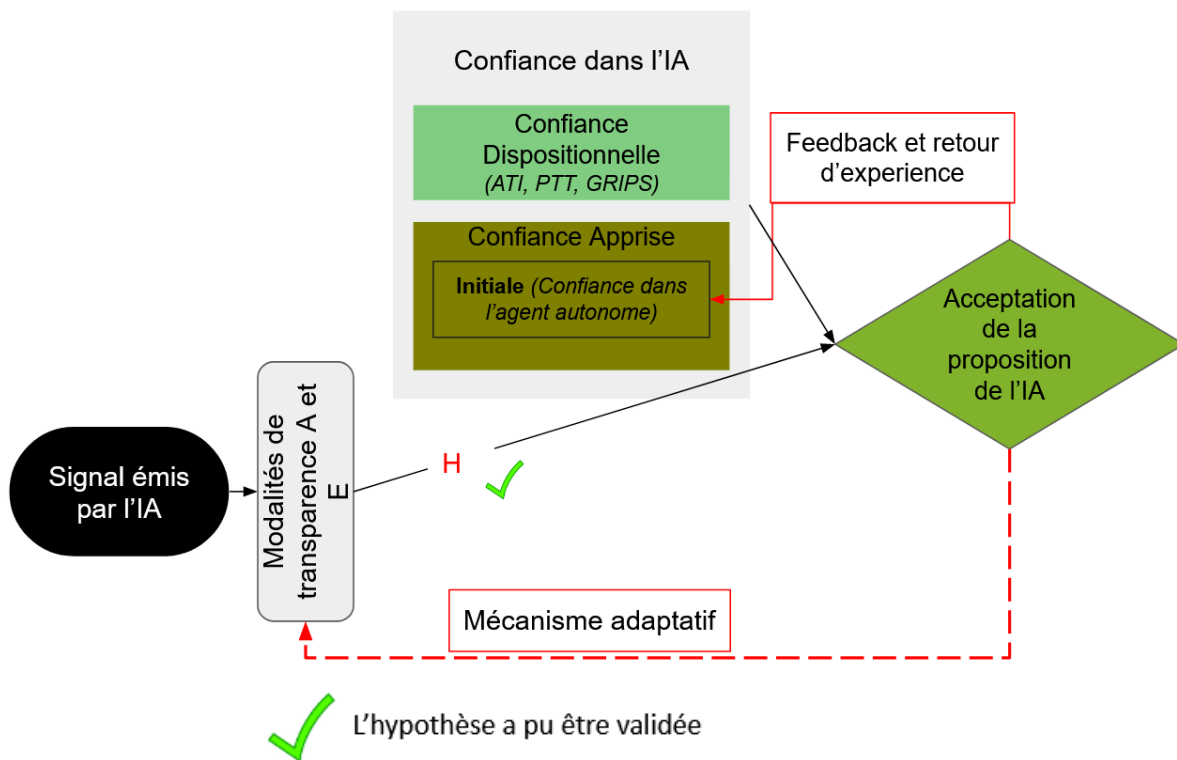


Figure 8.11 : Validation et synthèse de l'hypothèse initiale

Les composantes de la confiance sont cependant hétérogènes en termes d'explicabilité de l'acceptation et ne peuvent pas toutes être considérées comme des déterminants de la prise de décision de l'opérateur. Nous avons également pu montrer que la transparence influence la prise de décision de l'opérateur. De plus, la transparence du CPS a un effet sur les différentes composantes de la confiance (la confiance apprise dynamique et la confiance situationnelle). Enfin, nous avons constaté que, sous certaines conditions, les modifications de transparence du CPS peuvent être utilisées pour aider à la calibration de la confiance.

9.1. Apports théoriques, méthodologiques et pratiques de la thèse

9.1.1. Apports théoriques

Apports sur les relations entre les composantes de la confiance et les déterminants de la prise de décision de l'opérateur. D'un point de vue théorique nous avons apporté des éléments améliorant la compréhension du lien entre les composantes de la confiance et la prise de décision. Nous constatons que **toutes les composantes ne sont pas des déterminants de la prise de décision de l'opérateur**. Premièrement, la confiance dispositionnelle ne permet pas de prédire l'acceptation d'une proposition par l'opérateur. Contrairement à ce qui est avancé par Hoff et Bashir (2015), nous ne constatons pas de lien entre cette composante de la confiance et l'acceptation. S'il a déjà été constaté que les facteurs individuels, tels que l'âge, le genre ou la personnalité (analysé avec le Big Five par exemple (Kox, Seigling & Kerstholt, 2022)), sont en lien avec l'adoption d'une technologie, nous ne constatons pas de lien avec la réponse d'un opérateur à la proposition du CPS. Nous pensons que cela est dû à un manque de sensibilité de la confiance dispositionnelle. La confiance dispositionnelle est assez sensible pour prédire un taux d'acceptation global mais n'est pas assez sensible pour prédire l'acceptation à un instant T d'une proposition. Cette explication est appuyée par les résultats de notre troisième expérimentation. Celle-ci nous permet de constater que la confiance dispositionnelle est un indice sur le comportement qu'adoptera l'opérateur avec la technologie sur plusieurs propositions, mais pas de savoir s'il acceptera, de façon micro, sur une seule proposition (Figure 9.2).

Nos résultats permettent également d'étayer la proposition faite par Hoff et Bashir (2015) d'une **confiance structurée sous la forme de composantes et différenciées selon leur**

« **proximité** » de l'interaction. Les coefficients de corrélations entre la confiance apprise dynamique et l'acceptation, sont plus élevées qu'entre la confiance apprise initiale et l'acceptation (Figure 9.2). Nous l'expliquons de la même façon que précédemment, et en accord avec la conceptualisation de la confiance de Hoff et Bashir (ibid), par le fait que la confiance apprise dynamique est dans l'interaction, quand la confiance apprise initiale se trouve en amont de l'interaction. La confiance apprise initiale est une confiance plus générale envers le CPS, au-delà de la proposition qu'il peut émettre. La confiance apprise dynamique est, elle, un déterminant de la prise de décision de l'opérateur. Elle est la confiance de l'opérateur dans le CPS dans une interaction spécifique.

Nos résultats montrent que la confiance situationnelle est fortement corrélée avec l'acceptation. **Cette composante de la confiance est également un déterminant de la prise de décision de l'opérateur** (Figure 9.2).

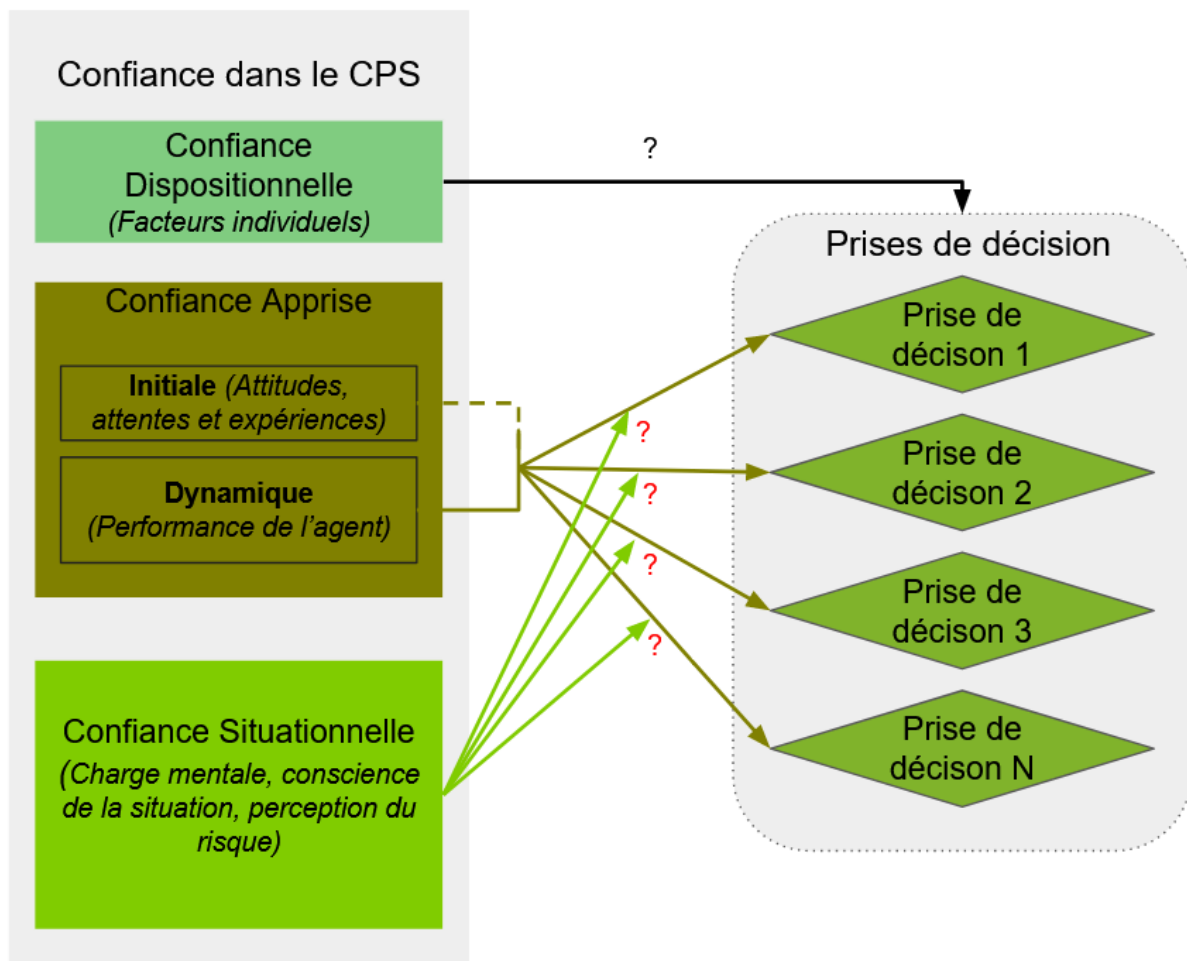


Figure 9.2 : Composantes de la confiance et déterminants de la prise de décision de l'opérateur

Pour Hoff et Bashir (ibid) cette confiance situationnelle est l'agrégat de multiples facteurs tels que la charge mentale, la conscience de la situation ou la perception du risque. En fonction des facteurs utilisés, cette composante de la confiance pourra être plus ou moins prédictive de l'acceptation. Dans nos expérimentations il était attendu que la charge mentale ne varie pas. Cela explique que cet indicateur ne soit pas corrélé à l'acceptation (Sato et al., 2020). En revanche, la perception du danger et la perception du risque étaient corrélées avec l'acceptation dans nos différentes expérimentations.

Nous constatons indirectement un effet modérateur de la confiance situationnelle sur le lien entre la confiance apprise et l'acceptation. Dans l'expérimentation Humanism nous constatons une interaction entre les deux dimensions de transparence (i.e. analytique et état de l'opérateur). La transparence état de l'opérateur représentait la capacité du CPS de transmettre des informations sur le risque. Ces informations venaient modérer positivement le lien entre fiabilité et acceptation lorsque les risques étaient faibles, et au contraire, modérer négativement ce lien lorsque les risques étaient élevés. Dans l'expérimentation Seanatic nous observons la même interaction entre les dimensions de transparence (i.e. analytique et environnement). La transparence sur l'environnement représentait la capacité du CPS de transmettre des informations sur les risques. Lorsque le risque était élevé, un effet modérateur négatif de la transparence « environnement » sur le lien entre la transparence analytique et l'acceptation a été observé. Ce rôle modérateur des informations sur le risque sur des informations de fiabilité d'un CPS est à rapprocher des résultats de Chancey (2016) et Hoesterey et Onnasch (2022). Leurs études montrent un effet modérateur du risque entre la confiance et divers comportements de confiance (acceptation et échantillonnage d'information auprès d'un CPS). Dans nos expérimentations, le fait d'avoir accès à des informations sur le risque vient réduire l'acceptation. Le rôle du risque dans la thématique de recherche de la confiance et de la coopération humain-machine est un sujet au cœur de l'actualité scientifique et nécessite encore d'être approfondi (Atchley et al., 2023). **Nos résultats ne permettent pas de statuer avec certitude si le risque vient modifier la confiance apprise dynamique, s'il vient en tant que modérateur ou s'il a un effet direct sur l'acceptation.** Il est probable que le contexte de coopération et les capacités du CPS expliquent la difficulté de statuer sur cette question. Le contexte jouant un rôle majeur dans

cette relation de modération. Ce rôle majeur du contexte expliquerait les différences de résultats qui ont pu être observés (Atchley et al., *ibid*).

Apports sur les relations entre la transparence, l'acceptation et les composantes de la confiance. Les résultats issus des expérimentations présentées dans ce manuscrit permettent d'améliorer notre compréhension théorique sur le rôle de la transparence dans la prise de décision assistée par un CPS. **Nos résultats confirment qu'il existe un lien entre la transparence et la prise de décision de l'opérateur** (Mercado et al., 2016 ; Stowers et al., 2020 ; Bhaskara et al., 2021). Contrairement à de précédents résultats (Mercado et al., 2016 ; Stowers et al., 2020), une plus grande transparence du CPS n'amène pas forcément à plus d'acceptation. Les résultats de notre première étude montrent que si le CPS augmente sa transparence (sur la dimension analytique) pour montrer une compétence moindre, alors l'opérateur est moins enclin à accepter une proposition. De la même façon, un CPS plus transparent sur une situation délétère (dans Humanism sur la dimension des états de l'opérateur) induit une baisse d'acceptation de ses propositions. A contrario, si la situation est positive, et ce malgré une faible fiabilité du CPS, les opérateurs acceptent plus la proposition. **La transparence joue ici son rôle de calibrateur de confiance en montrant les limites du CPS, ou la dangerosité de la situation, à l'opérateur.** L'opérateur peut alors prendre en compte ces informations lors de sa prise de décision (Yang et al., 2017 ; Zhang et al., 2020). Des résultats similaires sont observables dans notre seconde expérimentation où l'on constate que lorsque le CPS communique uniquement sur sa fiabilité (dimension analytique), le taux d'acceptation de la proposition est élevé. En revanche, lorsque le CPS communique uniquement des informations sur une situation à risque (dimension de l'environnement), alors le taux d'acceptation est bas. Lorsque le CPS communique sur ces deux dimensions, le taux d'acceptation se situe entre les taux d'acceptation observés lorsque le CPS communique sur l'une ou l'autre de ces dimensions. Les résultats de ces deux expérimentations permettent de montrer que les dimensions de la transparence interagissent entre elles. **L'opérateur utilise toutes les informations transmises par le CPS pour prendre sa décision, l'amenant à peser le pour et le contre avec les données mises à disposition.** De plus, lorsque le CPS est transparent sur sa fiabilité, et que celle-ci est haute, la confiance de l'opérateur augmente (Wright et al., 2020). Comme l'étude de Monsaigneon et al. (2020), nous voyons qu'il existe une interaction entre la fiabilité du CPS et le risque

d'une situation. Lorsque la situation présente des risques et que le CPS ne semble pas fiable, alors l'opérateur accepte moins. En revanche, nous voyons que lorsque la situation est favorable et que le CPS ne semble pas fiable, l'opérateur va accepter la proposition. A l'inverse, nous constatons que lorsque le CPS est fiable, mais que la situation est risquée, alors l'acceptation baisse. Cet effet des risques sur la fiabilité des CPS a été vu pour des prises de décision de niveau opérationnel et de niveau tactique.

Toujours du point de vue théorique, nos résultats montrent que **la transparence a un effet sur les composantes de la confiance.** Premièrement, la transparence a un effet sur la confiance apprise initiale (CAI). Les résultats de notre première expérimentation indiquent que seule la dimension de l'opérateur influence la CAI. Dans le cas où la situation est positive, la CAI augmentait. En revanche, la transparence sur la dimension analytique n'a pas influencé la CAI car le CPS ne communiquait pas de fiabilité, ou communiquait une fiabilité faible. En conséquence, la confiance des opérateurs ne variait pas. Du point de vue de l'opérateur, que le CPS ne communique pas sa fiabilité ou qu'il communique une fiabilité faible, a le même effet de calibration. Les résultats de la seconde expérimentation n'ont pas permis de retrouver un effet de la transparence sur la CAI. Ces résultats divergeant questionnent sur le type d'information transparente ayant un effet sur la CAI. Selon Hoff et Bashir (2015), la transparence n'influence pas la CAI. C'est pourquoi nous pensons que dans nos expérimentations, la transparence sur l'opérateur pouvait être assimilée à un feedback. Dans l'expérimentation Humanism, la transparence sur la dimension opérateur montrait un risque, ou une absence de risque, certain (le taux de charge du second opérateur). Dans l'expérimentation Seanatic la transparence sur l'environnement indiquait un risque probable (les risques liés à la situation). Un CPS transparent sur des risques certains a pu être interprété comme un retour d'expérience sur la situation et amener l'opérateur à reconsidérer la confiance qu'il avait dans le CPS. A l'inverse, un CPS transparent sur des probabilités de risques n'amenait pas l'opérateur à considérer cela comme un retour d'expérience sur la situation et n'a pas influencé la CAI.

La confiance apprise dynamique (CAD) est également influencée par les modalités de transparence utilisées par le CPS. Les résultats de l'expérimentation Humanism montrent un effet similaire des modalités de transparence sur la CAD à la CAI. La seconde

expérimentation montre en revanche que la CAD est plus sensible et varie dès que le CPS communique sur la dimension « environnement ». Cette différence de sensibilité est explicable par le fait que, selon Hoff et Bashir (2015), la CAD est dans l'interaction. Lorsque le CPS communique de nouvelles informations, l'opérateur les intègre et modifie sa confiance pour la calibrer avec ces nouvelles informations à disposition.

Enfin, la transparence a un effet sur la confiance situationnelle. Comme attendu (Bhaskara et al., 2020 ; Chen, 2021 ; van de Merwe et al., 2022), la transparence a un effet sur la conscience de la situation (SA) et la perception du risque, et une absence d'effet sur la charge mentale. Il est normal que la transparence sur les dimensions de l'opérateur et sur l'environnement influence cette confiance car le CPS transmet des informations sur la situation. Ces informations modifient la SA et la perception du risque. Concernant la charge mentale, nous pensons que le processus est à rapprocher de celui d'une utilisation d'une interface écologique (Garabet & Burns, 2004). L'augmentation de l'attention de l'opérateur nécessaire pour appréhender plus d'information est compensée par la création d'un meilleur référentiel commun (Common Frame Of Reference).

Apports sur la calibration de la confiance avec un CPS à transparence adaptative. En milieu réel, les situations de coopération ne sont pas indépendantes les unes aux autres. Pour aider les opérateurs à calibrer leur confiance durant une coopération où se succède plusieurs situations, l'utilisation de la transparence de façon dynamique a déjà montré des résultats encourageants (Akash et al., 2020a ; Akash et al., 2020b ; Okamura & Yamada, 2020). Nos résultats permettent de confirmer les résultats obtenus par ces précédentes expérimentations. **Un changement dynamique de transparence permet de modifier le comportement et donc la prise de décision de l'opérateur.** Dans notre expérimentation, lorsque le CPS ne communiquait plus sur sa fiabilité, les opérateurs se montraient moins confiants, et refusaient la proposition. A l'inverse, lorsque le CPS était de nouveau transparent sur sa fiabilité, les opérateurs se montraient à nouveau confiants et acceptaient la proposition. Nous retrouvons ici des résultats similaires à ceux obtenus dans la première expérimentation du projet Seanatic. Cela donne une perspective par rapport aux différents travaux sur la transparence (Bhaskara et al., 2020 ; van de Merwe et al., 2022). Les travaux sur la transparence peuvent être ré-utilisés dans un contexte plus écologique et de façon

dynamique. Nous n'avons pas pu observer si ce feedback a également influencé la confiance apprise initiale comme proposé par Hoff et Bashir (2015). En revanche, nous avons constaté un effet du feedback sur les prises de décisions des opérateurs.

Enfin nous observons que **la confiance dispositionnelle joue probablement un rôle important dans la sensibilité de l'opérateur face à un CPS avec une capacité de transparence adaptative**. Ces résultats sont cohérents avec ce que d'autres études ont montré (Pynadath et al., 2022 ; Kortschot et al., 2022). Si nos résultats ne sont pas significatifs, la confiance dispositionnelle et la confiance apprise initiale vont dans le sens attendu. L'affinité envers la technologie est associée à une meilleure conscience des limites des technologies par les opérateurs et est donc inversement corrélés avec le taux d'acceptation. La propension à la confiance est quant à elle corrélée avec un plus haut taux d'acceptation. La propension au risque est également corrélée avec plus d'acceptation. Faire confiance étant se mettre dans une position de prise de risque (Mayer et al., 1995), il est cohérent que cette propension au risque soit corrélée avec l'acceptation.

9.1.2. Apports méthodologiques

Apports sur les questionnaires. Nous montrons que la mesure de la confiance peut passer par plusieurs questionnaires différents. Le modèle proposé par Hoff et Bashir (2015) identifie la confiance comme structurée en trois composantes. Pour la confiance apprise nous montrons qu'en fonction du questionnaire utilisé nous pouvons différencier la confiance apprise initiale et dynamique. Il n'est pas possible de mesurer la confiance dispositionnelle et situationnelle, de façon exhaustive. Ces deux composantes sont basées sur de multiples facteurs (comme : la charge mentale et la conscience de la situation pour la confiance situationnelle) ce qui rend la tâche trop importante. C'est pourquoi nous avons restreint ce que nous observons aux éléments qui sont les plus important dans la coopération (Parasuraman et al., 2008).

Apports sur l'approche mixte. Nous avons, notamment dans l'expérimentation Humanism, utilisé un protocole mixte. C'est-à-dire que nous avons utilisé des méthodologies complémentaires pour mesurer les composantes de la confiance. Cette approche mixte peut être lourde à mettre en œuvre, notamment pour le participant, mais elle permet de croiser les résultats et d'améliorer la compréhension des effets provoqués dans l'expérimentation.

Cette approche mixte permet également de pouvoir analyser une même situation sous plusieurs angles. Cela permet d'apporter une meilleure compréhension des effets que peuvent avoir les variables manipulées.

Apports sur les modalités de transparence. Afin de manipuler le concept de transparence nous avons utilisé ce que nous appelons des modalités. Ces modalités proviennent de l'articulation de deux modèles que sont le Human-Robot Transparency Model (Lyons, 2013) et le SAT (Chen et al., 2018). Ces modalités permettent d'utiliser la transparence à un niveau plus fin et nous permet d'apporter un nouvel éclairage sur la façon dont un opérateur appréhende la transparence d'un CPS. L'utilisation de ces modalités permet de manipuler avec plus de finesse l'information qui est transmise par le CPS. Cette manipulation plus fine permet de mieux comprendre les effets sur les composantes de la confiance et sur l'acceptation. Ces modalités permettent également de tester des interactions entre différentes dimensions et différents niveaux et améliorer la compréhension des effets de la transparence.

Apports sur les déclencheurs de l'adaptation. Nous montrons qu'il est possible d'utiliser la prise de décision de l'opérateur comme déclencheur de l'adaptation de la transparence du CPS. Nous montrons également qu'un algorithme relativement simple, réalisé sur VBA, permet de créer des protocoles avec des déclencheurs. Nous montrons qu'avec simplement ces deux éléments, il est possible de proposer des expérimentations aux résultats pouvant contribuer à la compréhension de la calibration de la confiance.

9.1.3. Apports pratiques

Apports sur la transparence, les risques et la prise de décision assistée par un CPS. Les travaux présentés dans ce manuscrit montre l'importance qu'ont les informations sur le risque dans la prise de décision assistée par un CPS. Dans nos deux cas de figures, que la prise de décision soit de niveau opérationnelle ou stratégique, un CPS transparent sur les risques influence la prise de décision de l'opérateur. L'intégration d'un CPS, pouvant faire office d'assistant à la prise de décision de l'opérateur, dans un environnement de travail, nécessite, de penser à la façon dont ce CPS doit être transparent sur les risques et par quelles modalités. Il semble également important de comprendre quels risques doivent être communiqués à

l'opérateur. Dans l'expérimentation Seanatic nous nous sommes, par exemple, basés sur les experts métiers afin de savoir quels risques ils souhaitaient avoir.

Apports sur la transparence adaptative et calibration de la confiance. Nous montrons qu'il est possible d'utiliser la transparence de façon adaptative pour aider les opérateurs à calibrer leur confiance. Intégrer cette capacité au CPS pour moduler les informations en fonction des besoins des opérateurs est une des prochaines étapes pour améliorer la coopération humain-CPS.

9.2. Limites et perspectives

9.2.1. Limites et perspectives théoriques

Limites et perspectives sur les composantes de la confiance et la prise de décision.

Les travaux présentés dans cette thèse permettent d'améliorer la compréhension du rôle de la confiance dans la prise de décision. Nos résultats soutiennent en partie la modélisation de la confiance telle que proposée par Hoff et Bashir (2015). Nous voyons que les composantes qui se situent dans l'interaction, au sens du moment de l'interaction, sont plus prédictives que les composantes plus éloignées du moment de l'interaction. Cela confirme que la confiance est hétérogène. Par conséquent, il existe plusieurs méthodes pour appréhender, observer et comprendre la confiance. Nous apportons également des résultats permettant de penser que la confiance situationnelle est modératrice du lien entre la confiance apprise et l'acceptation. Nous retrouvons ici des résultats se rapprochant de la proposition de Chancey et al. (2017) proposant que le risque soit modérateur du lien entre confiance et acceptation. Les futurs travaux devraient s'intéresser à mieux comprendre l'effet de la perception du risque et de la conscience de la situation sur la prise de décision en les appliquant en milieu réel (Hoesterey & Onnasch, 2022). De plus, Hoff et Bashir (2015) considèrent de nombreux éléments dans la confiance situationnelle, il serait intéressant d'explorer le rôle de ces autres facteurs. Nous pensons notamment à l'allocation des fonctions ou à l'expertise de l'opérateur. Pour Bolton (2022), c'est le comportement, représenté dans ce manuscrit par la prise de décision, qui doit être au centre des attentions dans la coopération humain-machine. Mieux comprendre le rôle et les interactions entre les composantes de la confiance permettrait de mieux expliquer les comportements.

Limites et perspectives sur la transparence, l'acceptation et les composantes de la confiance. Concernant la transparence, la proposition de van de Merwe et al. (2022) était de dépasser les modèles existants et d'analyser ses effets selon les informations transmises. Nous avons proposé de combiner les modèles de la transparence de Chen et al. (2018) et Lyons (2013) afin d'analyser la transparence de façon micro tout en gardant une typologie existante. Utiliser des modèles déjà existant permet de faciliter la comparaison avec les expérimentations précédentes tout en améliorant la compréhension des effets de la transparence de façon plus micro. Nous pensons que la capacité de transparence d'un CPS doit être décomposée selon les informations qu'elle transmet, mais également sur la valence de ces informations. Cette valence joue un rôle sur le risque perçu par l'opérateur et celui-ci vient ensuite modérer la confiance. Une information positive sur la dimension « environnement » ou « état de l'opérateur » n'a pas le même effet qu'une information négative. Pareillement sur la dimension analytique, être transparent sur une faible fiabilité n'a pas le même effet qu'être transparent sur une fiabilité élevée. L'objectif est d'améliorer la compréhension des effets d'un CPS transparent sur la prise de décision de l'opérateur et sur les composantes de la confiance déterminantes de sa prise de décision. Si la transparence a bien vocation à calibrer la confiance (Theodorou et al., 2017), pour comprendre les mécanismes sous-jacents, les modèles actuels ne suffisent pas.

Limites et perspectives sur la calibration de la confiance avec un CPS à transparence adaptative. Nous montrons qu'il est possible de calibrer la confiance chez certains opérateurs grâce à la transparence. Cependant, nos résultats ne permettent pas de statuer sur le rôle de la confiance dispositionnelle des opérateurs dans la calibration de la confiance. Améliorer nos connaissances sur ce point permettra d'aider à l'intégration de ces CPS, et de créer une coopération qui prend en compte les besoins et les envies des opérateurs. Également, nous ne savons pas si les changements de transparence ont entraîné une évolution du modèle mental des opérateurs pour qu'il soit conforme à la réalité.

9.2.2. Limites et perspectives méthodologiques

Limites et perspectives sur les questionnaires. Nous avons proposé, au travers nos expérimentations, d'utiliser différentes échelles pour mesurer les composantes. La limite majeure de la multiplication de longues échelles est le temps nécessaire pour les remplir. Il

semble important de créer un questionnaire rapide pour la confiance apprise initiale et un second pour la confiance apprise dynamique. Les questionnaires de confiance sont, de plus, régulièrement remis en question sur leur validité et sur le concept ou l'objet qu'ils mesurent (Chita-Tegmark et al., 2022). Les questionnaires que nous avons utilisés sont sûrement perfectibles, et en conséquence, un travail de recherche les concernant est à considérer.

Limites et perspectives sur l'approche mixte. Nous avons également tenté d'utiliser une approche mixte pour comprendre et étudier les composantes de la confiance, grâce à des questionnaires, des comportements et pour notre première expérimentation, des données physiologiques. Ce positionnement méthodologique n'est cependant pas au centre de ce manuscrit et nous ne proposons ici que des ébauches de réponses. A notre sens, l'amélioration des méthodes de quantification des composantes est une des étapes majeures pour améliorer la compréhension du rôle de la confiance. Cette amélioration de la compréhension passe par la multiplication des croisements entre les méthodologies pour, à terme, pouvoir utiliser les méthodes les plus appropriées en fonction du contexte.

Limites et perspectives sur les modalités de transparence. Les expérimentations présentées dans cette thèse montrent l'intérêt de créer des protocoles où la transparence est catégorisée selon des dimensions et des niveaux. Cela permet de comprendre avec plus de précision les effets de la transparence. La suite à donner à nos expérimentations est l'utilisation de nouvelles modalités avec d'autres dimensions (Lyons, 2013) et niveaux (Chen et al., 2018). En effet, une des limites de cette thèse se trouve dans l'impossibilité de pouvoir tester de manière exhaustive toutes les configurations de transparence. L'exhaustivité n'est d'ailleurs probablement pas la manière la plus efficace pour la définition des modalités à explorer. Nous avons commencé ce travail en utilisant les analyses du Cognitive Work Analysis (Stanton et al., 2017) pour définir quelles informations le CPS devait communiquer pour faciliter sa compréhension. Une seconde manière de faire serait de coupler l'approche prescriptive utilisée dans ce manuscrit avec la méthodologie descriptive proposée par Johnson et al. (2014). Cette proposition vise à analyser les formes que prennent la coopération et l'interdépendance humain-machine dans une situation donnée pour décrire comment le CPS doit communiquer. L'objectif étant de coupler une méthodologie théorique

à une méthodologie pratique pour affiner la compréhension des effets de la transparence et l'adapter à un contexte.

Limites et perspectives sur les déclencheurs de l'adaptation. Nos résultats ont permis de confirmer les recherches précédentes sur la possibilité d'utiliser les comportements de l'opérateur comme déclencheur de l'adaptation de l'interface (Akash et al., 2020b ; Okamura & Yamada, 2020). Notre algorithme informatique était simple mais permettait tout de même d'arriver à une solution que nous avons jugé satisfaisante. Des algorithmes spécifiques prenant en compte plus d'éléments, comme le poste dans l'entreprise (Akiki, Bandara & Yu, 2013), est une suite logique à explorer. Nous retrouvons ici l'intérêt d'un travail en amont portant sur les méthodes de quantification de la confiance. L'utilisation d'autres métriques, tels que le langage (Li, Erickson, Cross & Lee, 2023), est une piste de recherche avec de nombreuses possibilités. L'objectif serait d'adapter la transparence selon d'autres états de l'opérateur et ces états devraient être recueillis avec le plus de précision et de fiabilité possible. L'utilisation d'une approche mixte permettrait aussi, à terme, de diversifier les déclencheurs utilisés par un CPS pour modifier sa transparence.

Limites et perspectives sur la généralisation des résultats. Les différentes conclusions apportées dans cette thèse sont basées sur des données récoltées chez un public non-expert. Les participants des expérimentations provenant pour la plupart de l'école d'ingénieur ENSIBS. D'un point de vue méthodologique, il serait intéressant de comparer nos résultats avec un public expert, ayant déjà été confronté aux CPS de types cobot ou IA, ou à défaut travaillant dans des postes de superviseur industriel ou de chef mécanicien. Comme évoqué par Hoff et Bashir (2015) l'expertise joue un rôle dans la confiance et la prise de décision. Cet élément est donc important à intégrer dans de futurs travaux pour éprouver la généralisation des résultats.

9.2.3. Limites et perspectives pratiques

Limites et perspectives sur la transparence, les risques et la prise de décision assistée par un CPS. Les travaux présentés dans ce manuscrit montrent l'importance de la transparence dans le cadre d'une coopération humain-machine. En fonction des modalités de transparence du CPS, l'opérateur sera plus ou moins prompt à adopter des comportements

de confiance. Ces comportements peuvent être d'accepter une proposition (la « compliance »), mais également de se fier au CPS (la « reliance »). La transparence permet de calibrer la confiance, et à l'opérateur de choisir d'accepter ou de refuser une proposition, parce qu'il connaît les forces et les faiblesses du CPS. Favoriser une confiance sur la base d'éléments tangibles permettra d'éviter des situations à risques. De la même façon qu'avec une interface écologique, connaître les propriétés et les contraintes du CPS permettra d'améliorer sa supervision et son utilisation. Pour l'industrie 4.0, l'intégration de CPS transparent améliore leur « savoir-coopérer » et permettra de favoriser la création d'un référentiel commun. Un meilleur « savoir-coopérer » des CPS améliorera la coopération entre l'humain et la machine (Pacaux-Lemoine, 2020). Les opérateurs seront mieux armés pour comprendre et utiliser les propositions des CPS.

Limites et perspectives sur les déclencheurs de l'adaptation. La transparence est un levier pour que le CPS adapte sa communication en fonction de l'opérateur. Les résultats présentés dans cette thèse ont pour vocation à garder, voir remettre, l'humain au centre de l'industrie et de la coopération humain-machine. Cette promesse d'une industrie anthropocentrée était au cœur de l'industrie 4.0 (Fantini et al., 2020), mais semble se déplacer vers ce que certains appellent désormais l'industrie 5.0 (Xu, Lu, Vogel-Heuser & Wang, 2021). Nonobstant cela, l'utilisation de la transparence permettra d'aider l'émergence d'une coopération humain-machine où la machine est une aide pour l'opérateur, et non une contrainte. Pour arriver à cela, les approches prescriptives et descriptives, sur la base d'une analyse de l'activité, seront complémentaires pour l'implantation de la transparence en milieu réel. Les changements de transparence du CPS en fonction de divers éléments est un levier à utiliser pour adapter le dialogue humain-CPS. De la même façon que deux personnes font évoluer leurs façons de converser grâce au développement d'un référentiel commun, (en développant des codes communs, Hoc, 2001), l'utilisation de la transparence de façon dynamique permettra de favoriser l'émergence d'une coopération symbiotique.

Perspectives générales. Pour conclure, nous assistons actuellement à une démocratisation de l'utilisation des outils d'intelligence artificielle dans notre quotidien avec, par exemple, l'outil ChatGPT. L'utilisation de ces outils pose la question de la confiance accordée à ces systèmes informatisés et la façon de les utiliser au sein de la société. Le

fonctionnement de ChatGPT se base sur la probabilité de survenue d'un mot en fonction des précédents mots et du contexte de l'énoncé. Les individus interprètent cela comme une réponse à une question alors que ce n'en est techniquement pas une. Le rôle de cette IA conversationnelle est « simplement » de continuer un énoncé initial (ScienceEtonnante, 2023). Le nombre d'étudiants utilisant cet outil sans remettre en question les résultats semble de plus en plus élevé. Un ChatGPT plus transparent (tel qu'en mode « *Playground* » où apparaît les probabilités utilisées pour continuer l'énoncé) serait-il dès lors associé à une confiance plus appropriée dans la réponse fournie ? Il est fort probable qu'à l'avenir ce genre de CPS fasse partie intégrante de notre façon de vivre, d'apprendre et d'enseigner. La transparence semble être une des meilleures voies(x) pour aider à une utilisation appropriée de ces outils.

Bibliographie :

- Adam, M., Gregor, S., Hevner, A., & Morana, S. (2021). Design Science Research Modes in Human-Computer Interaction Projects. *AIS Transactions on Human-Computer Interaction*, 13(1), 1-11. <https://doi.org/10.17705/1thci.00139>
- Ahmed, I., Jeon, G., & Piccialli, F. (2022). From Artificial Intelligence to Explainable Artificial Intelligence in Industry 4.0 : A Survey on What, How, and Where. *IEEE Transactions on Industrial Informatics*, 18(8), 5031-5042. <https://doi.org/10.1109/TII.2022.3146552>
- Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations : A theoretical analysis and review of empirical research. *Psychological bulletin*, 84(5), 888.
- Akash, K., Jain, N., & Misu, T. (2020a). Toward Adaptive Trust Calibration for Level 2 Driving Automation. *Proceedings of the 2020 International Conference on Multimodal Interaction*, 538-547.
- Akash, K., McMahon, G., Reid, T., & Jain, N. (2020b). *Human Trust-based Feedback Control : Dynamically varying automation transparency to optimize human-machine interactions*.
- Akiki, P., Bandara, A., & Yu, Y. (2013). RBUIS : Simplifying Enterprise Application User Interfaces through Engineering Role-Based Adaptive Behavior. In *EICS 2013—Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems* (p. 12). <https://doi.org/10.1145/2494603.2480297>
- Alarcon, G. M., Capiola, A., Hamdan, I. A., Lee, M. A., & Jessup, S. A. (2023). Differential biases in human-human versus human-robot interactions. *Applied Ergonomics*, 106, 103858. <https://doi.org/10.1016/j.apergo.2022.103858>
- Al-Tarawneh, H. A. (2011). The Main Factors beyond Decision Making. *Journal of Management Research*, 4(1). <https://doi.org/10.5296/jmr.v4i1.1184>
- Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Bonelli, S., Golfetti, A., Pozzi, S., Imbert, J.-P., Granger, G., Benhacene, R., & Babiloni, F. (2016). Adaptive Automation Triggered by EEG-Based Mental Workload Index : A Passive Brain-Computer Interface Application in Realistic Air Traffic Control Environment. *Frontiers in Human Neuroscience*, 10. <https://doi.org/10.3389/fnhum.2016.00539>
- Asimov, I. (1950). *I, Robot*. Doubleday. <https://books.google.fr/books?id=W-qRyAEACAAJ>
- Atchley, A., Barr, H. M., O'Hear, E., Weger, K., Mesmer, B., Gholston, Sampson, & Tenhundfeld, N. (2023). Trust in systems : Identification of 17 unresolved research questions and the highlighting of inconsistencies. *Theoretical Issues in Ergonomics Science*, 0(0), 1-25. <https://doi.org/10.1080/1463922X.2023.2223251>

- B**addeley, A. D., & Logie, R. H. (1999). Working Memory : The Multiple-Component Model. In A. Miyake & P. Shah (Éds.), *Models of Working Memory* (1^{re} éd., p. 28-61). Cambridge University Press. <https://doi.org/10.1017/CBO9781139174909.005>
- Baheti, R., & Gill, H. (2011). Cyber-physical systems. *The impact of control technology*, 12(1), 161-166.
- Bainbridge, L. (1983). Ironies of automation. In *Analysis, design and evaluation of man-machine systems* (p. 129-135). Elsevier.
- Bates, D., Maechler, M., & Bolker, B. (2012). lme4 : Linear mixed-effects models using S4 classes [Software]. Retrieved from CRAN. R-project.org/package=lme4.
- Baxter, G., Rooksby, J., Wang, Y., & Khajeh-Hosseini, A. (2012). The ironies of automation : Still going strong at 30? *Proceedings of the 30th European Conference on Cognitive Ergonomics - ECCE '12*, 65. <https://doi.org/10.1145/2448136.2448149>
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin*, 91(2), 276.
- Bhaskara, A., Duong, L., Brooks, J., Li, R., McInerney, R., Skinner, M., Pongracic, H., & Loft, S. (2021). Effect of automation transparency in the management of multiple unmanned vehicles. *Applied Ergonomics*, 90, 103243. <https://doi.org/10.1016/j.apergo.2020.103243>
- Bhaskara, A., Skinner, M., & Loft, S. (2020). Agent Transparency : A Review of Current Theory and Evidence. *IEEE Transactions on Human-Machine Systems*, 50(3), 215-224. <https://doi.org/10.1109/THMS.2020.2965529>
- Blaauwgeers, E., Dubois, L., & Ryckaert, L. (2013). Real-time risk estimation for better situational awareness. *IFAC Proceedings Volumes*, 46(15), 232-239.
- Bolton, M. L. (2022). Trust is Not a Virtue : Why We Should Not Trust Trust. *Ergonomics in Design*, 10648046221130172. <https://doi.org/10.1177/10648046221130171>
- Browne, D., Norman, M., & Riches, D. (1990). Chapter 2—Why Build Adaptive Systems? In D. Browne, P. Totterdell, & M. Norman (Éds.), *Adaptive User Interfaces* (p. 15-57). Academic Press. <https://doi.org/10.1016/B978-0-12-137755-7.50007-8>
- Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational psychologist*, 38(1), 53-61.
- Burns, C. M., & Hajdukiewicz, J. (2004). *Ecological Interface Design*. CRC Press.
- C**alvary, G., & Coutaz, J. (2002). Plasticité des Interfaces : Une nécessité. *Information-Interaction-Intelligence, Actes des deuxièmes Assises nationales du GDR I*, 3, 247-261.
- Canito, A., Mota, D., Marreiros, G., Corchado, J. M., & Martins, C. (2021). Contextual Adaptive Interfaces for Industry 4.0. In S. Rodríguez González, A. González-Briones, A. Gola, G. Katranas, M. Ricca, R. Loukanova, & J. Prieto (Éds.), *Distributed Computing and Artificial Intelligence, Special Sessions, 17th International Conference* (p.

- 149-157). Springer International Publishing. https://doi.org/10.1007/978-3-030-53829-3_14
- Cardin, O. (2019). Classification of cyber-physical production systems applications : Proposition of an analysis framework. *Computers in Industry*, 104, 11-21. <https://doi.org/10.1016/j.compind.2018.10.002>
- Cegarra, J., & Chevalier, A. (2008). The use of Tholos software for combining measures of mental workload : Toward theoretical and methodological improvements. *Behavior Research Methods*, 40(4), 988-1000.
- Chancey, E. T. (2016). *The effects of alarm system errors on dependence : Moderated mediation of trust with and without risk*. Old Dominion University.
- Chancey, E., Bliss, J., Yamani, Y., & Handley, H. (2017). Trust and the Compliance-Reliance Paradigm : The Effects of Risk, Error Bias, and Reliability on Trust and Dependence. *Human Factors The Journal of the Human Factors and Ergonomics Society*, 59, 333-345. <https://doi.org/10.1177/0018720816682648>
- Chauvin, C., Coppin, G., & Chéné, H. (2010). Analysis of the dynamics of common ground : A methodological proposal. *Proceedings of the 28th Annual European Conference on Cognitive Ergonomics*, 209-212.
- Chen, J., Mishler, S., & Hu, B. (2021). Automation Error Type and Methods of Communicating Automation Reliability Affect Trust and Performance : An Empirical Study in the Cyber Domain. *IEEE Transactions on Human-Machine Systems*.
- Chen, J. Y. C. (2021). Agent Transparency. In *Smart and Intelligent Systems*. CRC Press.
- Chen, J. Y. C., Lakhmani, S. G., Stowers, K., Selkowitz, A. R., Wright, J. L., & Barnes, M. (2018). Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical Issues in Ergonomics Science*, 19(3), 259-282. <https://doi.org/10.1080/1463922X.2017.1315750>
- Chen, J. Y., Procci, K., Boyce, M., Wright, J., Garcia, A., & Barnes, M. (2014). *Situation Awareness-Based Agent Transparency* (ARL-TR-6905). ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD HUMAN RESEARCH AND ENGINEERING DIRECTORATE. <https://apps.dtic.mil/docs/citations/ADA600351>
- Chen, T., Campbell, D., Gonzalez, L. F., & Coppin, G. (2015). Increasing Autonomy Transparency through capability communication in multiple heterogeneous UAV management. *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2434-2439. <https://doi.org/10.1109/IROS.2015.7353707>
- Chita-Tegmark, M., Law, T., Rabb, N., & Scheutz, M. (2021). Can You Trust Your Trust Measure? *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 92-100. <https://doi.org/10.1145/3434073.3444677>
- Collet, C., Salvia, E., & Petit-Boulanger, C. (2014). Measuring workload with electrodermal activity during common braking actions. *Ergonomics*, 57(6), 886-896.

Correia, A., & Lindley, S. (2022). Collaboration in relation to Human-AI Systems : Status, Trends, and Impact. *2022 IEEE International Conference on Big Data (Big Data)*, 3417-3422. <https://doi.org/10.1109/BigData55660.2022.10020416>

Davis, S. (2019). *Individual Differences in Operators' Trust in Autonomous Systems : A Review of the Literature*.

Day, M.-Y., Cheng, T.-K., & Li, J.-G. (2018). AI robo-advisor with big data analytics for financial services. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 1027-1031.

de Visser, E. J., Pak, R., & Shaw, T. H. (2018). From “automation” to ‘autonomy’ : The importance of trust repair in human-machine interaction. *Ergonomics*, *61*(10), 1409-1427. <https://doi.org/10.1080/00140139.2018.1457725>

Di Nardo, M., Forino, D., & Murino, T. (2020). The evolution of man-machine interaction : The role of human in Industry 4.0 paradigm. *Production & Manufacturing Research*, *8*(1), 20-34. <https://doi.org/10.1080/21693277.2020.1737592>

Dizaji, L. G., & Hu, Y. (2021). Building And Measuring Trust In Human-Machine Systems. *2021 IEEE International Conference on Autonomous Systems (ICAS)*, 1-5. <https://doi.org/10.1109/ICAS49788.2021.9551131>

Edwards, W. (1954). The theory of decision making. *Psychological bulletin*, *51*(4), 380.

Efendic, E. (2017). *The Impact of Multiple Affective Reactions on Decision Making : Combination of Affect and the Mediating Mechanisms of Affective Influence*. 213.

Endsley, M. R. (1995a). Measurement of Situation Awareness in Dynamic Systems. *Human Factors*, *37*(1), 65-84. <https://doi.org/10.1518/001872095779049499>

Endsley, M. R. (1995b). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, *37*(1), 32-64. <https://doi.org/10.1518/001872095779049543>

Endsley, M. R. (2006). Expertise and situation awareness. *The Cambridge handbook of expertise and expert performance*, 633-651.

Endsley, M. R. (2015). Situation Awareness Misconceptions and Misunderstandings. *Journal of Cognitive Engineering and Decision Making*, *9*(1), 4-32. <https://doi.org/10.1177/1555343415572631>

Endsley, M. R. (2021). A Systematic Review and Meta-Analysis of Direct Objective Measures of Situation Awareness : A Comparison of SAGAT and SPAM. *Human Factors*, *63*(1), 124-150. <https://doi.org/10.1177/0018720819875376>

Ermisch, J., Gambetta, D., Laurie, H., Siedler, T., & Noah Uhrig, S. C. (2009). Measuring people's trust. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*(4), 749-769.

- F**antini, P., Pinzone, M., & Taisch, M. (2020). Placing the operator at the centre of Industry 4.0 design : Modelling and assessing human activities within cyber-physical systems. *Computers & Industrial Engineering*, *139*, 105058.
- Franke, T., Attig, C., & Wessel, D. (2019). A Personal Resource for Technology Interaction : Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human–Computer Interaction*, *35*(6), 456-467. <https://doi.org/10.1080/10447318.2018.1456150>
- G**arabet, A., & Burns, C. M. (2004). Collaboration with Ecological Interface Design. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(3), 543-546. <https://doi.org/10.1177/154193120404800357>
- Gorecky, D., Schmitt, M., Loskyll, M., & Zühlke, D. (2014). Human-machine-interaction in the industry 4.0 era. *2014 12th IEEE International Conference on Industrial Informatics (INDIN)*, 289-294. <https://doi.org/10.1109/INDIN.2014.6945523>
- Göritzlehner, R., Borst, C., Ellerbroek, J., Westin, C., van Paassen, M. M., & Mulder, M. (2014). Effects of transparency on the acceptance of automated resolution advisories. *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2965-2970. <https://doi.org/10.1109/SMC.2014.6974381>
- Greef, T. de, Lafeber, H., Oostendorp, H. van, & Lindenberg, J. (2009). Eye movement as indicators of mental workload to trigger adaptive automation. *International Conference on Foundations of Augmented Cognition*, 219-228.
- Greenberg, S., & Witten, I. H. (1985). Adaptive personalized interfaces—A question of viability. *Behaviour & Information Technology*, *4*(1), 31-45. <https://doi.org/10.1080/01449298508901785>
- Guerin, C., Rauffet, P., Chauvin, C., & Martin, E. (2019). Toward production operator 4.0 : Modelling Human-Machine Cooperation in Industry 4.0 with Cognitive Work Analysis. *IFAC-PapersOnLine*, *52*(19), 73-78. <https://doi.org/10.1016/j.ifacol.2019.12.111>
- Guesmia, M., Chattia, M. A., Vorgerda, L., Joardera, S., Aina, Q. U., Ngoa, T., Zumora, S., Suna, Y., Jia, F., & Muslimb, A. (2021). *Input or Output : Effects of Explanation Focus on the Perception of Explainable Recommendation with Varying Level of Details*.
- Guznov, S., Lyons, J., Pfahler, M., Heironimus, A., Woolley, M., Friedman, J., & Neimeier, A. (2020). Robot Transparency and Team Orientation Effects on Human–Robot Teaming. *International Journal of Human–Computer Interaction*, *36*(7), 650-660. <https://doi.org/10.1080/10447318.2019.1676519>
- H**abib, L., Pacaux-Lemoine, M.-P., Berdal, Q., & Trentesaux, D. (2021). From Human-Human to Human-Machine Cooperation in Manufacturing 4.0. *Processes*, *9*(11), Article 11. <https://doi.org/10.3390/pr9111910>

- Hancock, P. A., Billings, D. R., Schaefer, K., Chen, J., Visser, E. D., & Parasuraman, R. (2011). A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Hum. Factors*. <https://doi.org/10.1177/0018720811417254>
- Harper, K., Guarino, S., & Zacharias, G. (2000). Adaptive human/machine interface for advanced air traffic management. *AIAA Guidance, Navigation, and Control Conference and Exhibit*, 4063.
- Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9), 904-908. <https://doi.org/10.1177/154193120605000909>
- Hart, S., G., & Staveland, L., E. (1988). Development of NASA-TLX (Task Load Index) : Results of Empirical and Theoretical Research. *Advances in Psychology*, 52, 139-183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Helldin, T. (2014). *Transparency for future semi-automated systems* [PhD diss., Orebro University.]. https://www.his.se/PageFiles/11241/tove_helldin_14_april.pdf
- Helldin, T., Falkman, G., Riveiro, M., & Davidsson, S. (2013). Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving. *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '13*, 210-217. <https://doi.org/10.1145/2516540.2516554>
- Hirsch-Kreinsen, H. (2014). Wandel von Produktionsarbeit–„Industrie 4.0 «. *WSI-Mitteilungen*, 67(6), 421-429.
- Hoc, J.-M. (1996). *Supervision et controle de processus : La cognition en situation dynamique*. Presses Universitaires de Grenoble.
- HOC, J.-M. (2001). Towards a cognitive approach to human-machine cooperation in dynamic situations. *International Journal of Human-Computer Studies*, 54(4), 509-540. <https://doi.org/10.1006/ijhc.2000.0454>
- Hoesterey, S., & Onnasch, L. (2022). The effect of risk on trust attitude and trust behavior in interaction with information and decision automation. *Cognition, Technology & Work*. <https://doi.org/10.1007/s10111-022-00718-y>
- Hoff, K. A., & Bashir, M. (2015). Trust in automation : Integrating empirical evidence on factors that influence trust. *Human factors*, 57(3), 407-434.
- Hollnagel, E. (2005). *Designing for joint cognitive systems* (p. 51). <https://doi.org/10.1049/ic:20050450>
- Hozdić, E. (2015). Smart factory for industry 4.0 : A review. *International Journal of Modern Manufacturing Technologies*, 7(1), 28-35.
- Hudon, A., Demazure, T., Karran, A., Léger, P.-M., & Sénécal, S. (2021). Explainable Artificial intelligence (XAI) : How the Visualization of AI Predictions Affects User Cognitive Load and Confidence. *Fred D. Davis, René Riedl, Jan vom Brocke, Pierre-Majorique Léger, Adriane B. Randolph, Gernot Müller-Putz (Eds.)*, 263.

Hudson, P. (2014). *The Industrial Revolution*. Bloomsbury Publishing.

Inkulu, A. K., Bahubalendruni, M. V. A. R., Dara, A., & K., S. (2021). Challenges and opportunities in human robot collaboration context of Industry 4.0—A state of the art review. *Industrial Robot: the international journal of robotics research and application*, 49(2), 226-239. <https://doi.org/10.1108/IR-04-2021-0077>

Innocent, P. R. (1982). Towards self-adaptive interface systems. *International Journal of Man-Machine Studies*, 16(3), 287-299. [https://doi.org/10.1016/S0020-7373\(82\)80063-1](https://doi.org/10.1016/S0020-7373(82)80063-1)

Jarrahi, M. H. (2018). Artificial intelligence and the future of work : Human-AI symbiosis in organizational decision making. *Business horizons*, 61(4), 577-586.

Javid, M., & Estivill-Castro, V. (2021). Explanations from a Robotic Partner Build Trust on the Robot's Decisions for Collaborative Human-Humanoid Interaction. *Robotics*, 10(1), Article 1. <https://doi.org/10.3390/robotics10010051>

Jayaraman, S. K., Creech, C., Tilbury, D. M., Yang, X. J., Pradhan, A. K., Tsui, K. M., & Robert Jr, L. P. (2019). Pedestrian trust in automated vehicles : Role of traffic signal and AV driving behavior. *Frontiers in Robotics and AI*, 6, 117.

Jessup, S. A. (2018). *Measurement of the Propensity to Trust Automation*. 65.

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics*, 4(1), 53-71. https://doi.org/10.1207/S15327566IJCE0401_04

Jiang, J., Karran, A. J., Coursaris, C. K., Léger, P.-M., & Beringer, J. (2022). A Situation Awareness Perspective on Human-AI Interaction : Tensions and Opportunities. *International Journal of Human-Computer Interaction*, 1-18. <https://doi.org/10.1080/10447318.2022.2093863>

Johannsen, G. (2009). Human-machine interaction. *Control Systems, Robotics and Automation*, 21, 132-162.

Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C. M., Van Riemsdijk, M. B., & Sierhuis, M. (2014). Coactive Design : Designing Support for Interdependence in Joint Activity. *Journal of Human-Robot Interaction*, 3(1), 43. <https://doi.org/10.5898/JHRI.3.1.Johnson>

Kaber, D. B., & Riley, J. M. (1999). Adaptive automation of a dynamic control task based on secondary task workload measurement. *International Journal of Cognitive Ergonomics*, 3, 169-187. https://doi.org/10.1207/s15327566ijce0303_1

Kadir, B. A., Broberg, O., & Souza da Conceição, C. (2018). Designing human-robot collaborations in industry 4.0 : Explorative case studies. *DS 92: Proceedings of the DESIGN 2018 15th International Design Conference*, 601-610.

Kagermann, H., & Wahlster, W. (2022). Ten years of Industrie 4.0. *Sci*, 4(3), 26.

- Kagermann, H., Wahlster, W., & Helbig, J. (2013). Securing the future of German manufacturing industry : Recommendations for implementing the strategic initiative INDUSTRIE 4.0. *Final report of the Industrie, 4(0)*.
- Kaplan, A. D., Kessler, T. T., Brill, J. C., & Hancock, P. A. (2023). Trust in Artificial Intelligence : Meta-Analytic Findings. *Human Factors, 65(2)*, 337-359. <https://doi.org/10.1177/00187208211013988>
- Karpinsky, N. D., Chancey, E. T., Palmer, D. B., & Yamani, Y. (2018). Automation trust and attention allocation in multitasking workspace. *Applied Ergonomics, 70*, 194-201. <https://doi.org/10.1016/j.apergo.2018.03.008>
- Kaufman, A. R., Persoskie, A., Twesten, J., & Bromberg, J. (2018). A review of risk perception measurement in tobacco control research. *Tobacco Control, tobaccocontrol-2017-054005*. <https://doi.org/10.1136/tobaccocontrol-2017-054005>
- Khavas, Z. R. (2021). A Review on Trust in Human-Robot Interaction. *arXiv:2105.10045 [cs]*. <http://arxiv.org/abs/2105.10045>
- Kinateder, M., Kuligowski, E., Reneke, P., & Peacock, R. (2015). Risk perception in fire evacuation behavior revisited : Definitions, related concepts, and empirical evidence. *Fire Science Reviews, 4, 1*. <https://doi.org/10.1186/s40038-014-0005-z>
- Kintz, J. R., Banerjee, N. T., Zhang, J. Y., Anderson, A. P., & Clark, T. K. (2022). Estimation of Subjectively Reported Trust, Mental Workload, and Situation Awareness Using Unobtrusive Measures. *Human Factors, 00187208221129371*. <https://doi.org/10.1177/00187208221129371>
- Kizilcec, R. F. (2016). How Much Information? : Effects of Transparency on Trust in an Algorithmic Interface. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 2390-2395*. <https://doi.org/10.1145/2858036.2858402>
- Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *The quarterly journal of economics, 133(1)*, 237-293.
- Kluy, L., & Roesler, E. (2021). Working with Industrial Cobots : The Influence of Reliability and Transparency on Perception and Trust. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 65(1)*, 77-81. <https://doi.org/10.1177/1071181321651110>
- Koh, L., Orzes, G., & Jia, F. (Jeff). (2019). The fourth industrial revolution (Industry 4.0) : Technologies disruption on operations and supply chain management. *International Journal of Operations & Production Management, 39(6/7/8)*, 817-828. <https://doi.org/10.1108/IJOPM-08-2019-788>
- Kohler, D., & Weisz, J.-D. (2016). Industrie 4.0 : Comment caractériser cette quatrième révolution industrielle et ses enjeux ? *Annales des Mines - Réalités industrielles, vembre 2016(4)*, 51-56. <https://doi.org/10.3917/rindu1.164.0051>
- Komogortsev, O., Holland, C., & Camou, J. (2011). Adaptive eye-gaze-guided interfaces : Design & performance evaluation. *CHI '11 Extended Abstracts on Human*

- Kortschot, S. W., Jamieson, G. A., & Prasad, A. (2022). Detecting and Responding to Information Overload With an Adaptive User Interface. *Human Factors*, 64(4), 675-693. <https://doi.org/10.1177/0018720820964343>
- Kostenko, A. S. (2017). *Évaluation multidimensionnelle et dynamique de la maîtrise de la situation par l'opérateur : Création d'un indicateur temps réel de charge mentale pour l'activité de supervision de drones* [Phdthesis, Université de Bretagne Sud]. <https://tel.archives-ouvertes.fr/tel-01677692>
- Kox, E., Siegling, L., & Kerstholt, J. (2022). Trust Development in Military and Civilian Human-Agent Teams: The Effect of Social-Cognitive Recovery Strategies. *International Journal of Social Robotics*, 14. <https://doi.org/10.1007/s12369-022-00871-4>
- Kramer, A. F. (2020). Physiological metrics of mental workload : A review of recent progress. *Multiple-task performance*, 279-328.
- Kunze, A., Summerskill, S. J., Marshall, R., & Filtner, A. J. (2019). Automation transparency : Implications of uncertainty communication for human-automation interaction and interfaces. *Ergonomics*, 62(3), 345-360. <https://doi.org/10.1080/00140139.2018.1547842>
- Lai, V., Chen, C., Liao, Q. V., Smith-Renner, A., & Tan, C. (2021). *Towards a Science of Human-AI Decision Making : A Survey of Empirical Studies* (arXiv:2112.11471). arXiv. <https://doi.org/10.48550/arXiv.2112.11471>
- Lallemand, C., & Gronier, G. (2015). *Méthodes de design UX : 30 méthodes fondamentales pour concevoir et évaluer les systèmes interactifs*. Editions Eyrolles.
- Lassalle, J., Rauffet, P., Leroy, B., Guérin, C., Chauvin, C., Coppin, G., & Saïd, F. (2017). COmmunication and WORKload analyses to study the COLlective WORK of fighter pilots : The COWORK2 method. *Cognition, Technology & Work*, 19(2), 477-491. <https://doi.org/10.1007/s10111-017-0420-8>
- Lassalle, J., Simon, L. (2020). Tâche 1.1 SEANATIC: Spécifications des besoins-recommandations psycho-ergonomiques et cognitives pour la conception des IHS de maintenance maritime prédictive. [Rapport ADEME L1.1_1]. Université Bretagne Sud
- Law, T., & Scheutz, M. (2021). Chapter 2 - Trust : Recent concepts and evaluations in human-robot interaction. In C. S. Nam & J. B. Lyons (Éds.), *Trust in Human-Robot Interaction* (p. 27-57). Academic Press. <https://doi.org/10.1016/B978-0-12-819472-0.00002-2>
- Lean, Y., & Shan, F. (2012). Brief review on physiological and biochemical evaluations of human mental workload. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 22(3), 177-187.

- Lee, E. A. (2006). Cyber-physical systems-are computing foundations adequate. *Position paper for NSF workshop on cyber-physical systems: research motivation, techniques and roadmap*, 2, 1-9.
- Lee, J. D., & See, K. A. (2004). Trust in Automation : Designing for Appropriate Reliance. *Human Factors*, 46(1), 50-80. https://doi.org/10.1518/hfes.46.1.50_30392
- Li, M., Erickson, I. M., Cross, E. V., & Lee, J. D. (2023). It's Not Only What You Say, But Also How You Say It : Machine Learning Approach to Estimate Trust from Conversation. *Human Factors*, 187208231166624. <https://doi.org/10.1177/00187208231166624>
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.
- Lindholm, M., Reiman, A., & Väyrynen, S. (2020). On Future Occupational Safety and Health Challenges : A Systematic Literature Review. *International Journal of Occupational and Environmental Safety*, 4(1), 108-127.
- Liu, B. (2021). In AI We Trust? Effects of Agency Locus and Transparency on Uncertainty Reduction in Human–AI Interaction. *Journal of Computer-Mediated Communication*, zrab013. <https://doi.org/10.1093/jcmc/zrab013>
- Liu, H., Hirayama, T., Saiki, L. Y. M., & Murase, H. (2020). What Timing for an Automated Vehicle to Make Pedestrians Understand Its Driving Intentions for Improving Their Perception of Safety? *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 1-6. <https://doi.org/10.1109/ITSC45102.2020.9294696>
- Loft, S., Bhaskara, A., Lock, B. A., Skinner, M., Brooks, J., Li, R., & Bell, J. (2021). The Impact of Transparency and Decision Risk on Human–Automation Teaming Outcomes. *Human Factors*, 00187208211033445. <https://doi.org/10.1177/00187208211033445>
- Loiselet, A., & Hoc, J.-M. (2001). La gestion des interférences et du référentiel commun dans la coopération : Implications pour la conception. *Psychologie française*, 46(2), 167-179.
- Longo, F., Nicoletti, L., & Padovano, A. (2017). Smart operators in industry 4.0 : A human-centered approach to enhance operators' capabilities and competencies within the new smart factory context. *Computers & industrial engineering*, 113, 144-159.
- Loup-Escande, É., Burkhardt, J.-M., & Richir, S. (2013). Anticiper et évaluer l'utilité dans la conception ergonomique des technologies émergentes : Une revue. *Le travail humain*, 76(1), 27-55. <https://doi.org/10.3917/th.761.0027>
- Lu, Y., Adrados, J. S., Chand, S. S., & Wang, L. (2021). Humans Are Not Machines—Anthropocentric Human–Machine Symbiosis for Ultra-Flexible Smart Manufacturing. *Engineering*, 7(6), 734-737. <https://doi.org/10.1016/j.eng.2020.09.018>
- Lu, Y., Morris, K. C., & Frechette, S. (2016). Current standards landscape for smart manufacturing systems. *National Institute of Standards and Technology, NISTIR*, 8107, 39.

- Luster, M. S., & Pitts, B. J. (2021). Trust in Automation : The Effects of System Certainty on Decision-Making. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 65(1), 32-36. <https://doi.org/10.1177/1071181321651079>
- Lyons, J. B. (2013, mars 15). Being Transparent about Transparency : A Model for Human-Robot Interaction. *2013 AAAI Spring Symposium Series*. 2013 AAAI Spring Symposium Series. <https://www.aaai.org/ocs/index.php/SSS/SSS13/paper/view/5712>
- Lyons, J. B., & Guznov, S. Y. (2019). Individual differences in human-machine trust : A multi-study look at the perfect automation schema. *Theoretical Issues in Ergonomics Science*, 20(4), 440-458.
- Lyons, J. B., Koltai, K. S., Ho, N. T., Johnson, W. B., Smith, D. E., & Shively, R. J. (2016). Engineering trust in complex automated systems. *Ergonomics in Design*, 24(1), 13-17.
- M**aas, V. S., & Yin, H. (2021). Finding partners in crime? How transparency about managers' behavior affects employee collusion. *Accounting, Organizations and Society*, 101293. <https://doi.org/10.1016/j.aos.2021.101293>
- Madonna, M., Monica, L., Anastasi, S., & Di Nardo, M. (2019). *EVOLUTION OF COGNITIVE DEMAND IN THE HUMAN-MACHINE INTERACTION INTEGRATED WITH INDUSTRY 4.0 TECHNOLOGIES*. 13-19. <https://doi.org/10.2495/SAFE190021>
- Marsh, S., & Dibben, M. R. (2003). The role of trust in information science and technology. *Annual Review of Information Science and Technology*, 37(1), 465-498. <https://doi.org/10.1002/aris.1440370111>
- Märtins, J., Kehl, R., Westmattmann, D., & Schewe, G. (2022). A Multi-Dimensional Theory of Transparency, Trust, and Risk in Technology Adoption. *ICIS 2022 Proceedings*. https://aisel.aisnet.org/icis2022/is_implement/is_implement/7
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model Of Organizational Trust. *Academy of Management Review*, 20(3), 709-734. <https://doi.org/10.5465/amr.1995.9508080335>
- McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G. S., & Darzi, A. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788), 89-94.
- McNeese, N. J., Demir, M., Cooke, N. J., & Myers, C. (2018). Teaming With a Synthetic Teammate : Insights into Human-Autonomy Teaming. *Human Factors*, 60(2), 262-273. <https://doi.org/10.1177/0018720817743223>
- Mehler, B., Reimer, B., Coughlin, J. F., & Dusek, J. A. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation research record*, 2138(1), 6-12.
- Mercado, J. E., Rupp, M. A., Chen, J. Y. C., Barnes, M. J., Barber, D., & Procci, K. (2016). Intelligent Agent Transparency in Human-Agent Teaming for Multi-UxV

- Management. *Human Factors*, 58(3), 401-415. <https://doi.org/10.1177/0018720815621206>
- Meshkati, N. (1988). Heart rate variability and mental workload assessment. In *Advances in psychology* (Vol. 52, p. 101-115). Elsevier.
- Meuwissen, M., & Bollen, L. (2021). *Transparency versus Explainability in AI*. <https://doi.org/10.13140/RG.2.2.27466.90561>
- Miller, G. A. (1956). The magical number seven, plus or minus two : Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81-97. <https://doi.org/10.1037/h0043158>
- Millot, P. (1988). *Supervision des procédés automatisés et ergonomie*. Hermès.
- Millot, P. (1999). *Systèmes homme-machine et automatique*.
- Millot, P., & Mandiau, R. (1995). *Man-machine cooperative organizations : Formal and pragmatic implementation methods*. 213-228.
- Millot, P., & Pacaux-Lemoine, M.-P. (2013). A Common Work Space for a mutual enrichment of Human-machine Cooperation and Team-Situation Awareness. *IFAC Proceedings Volumes*, 46(15), 387-394. <https://doi.org/10.3182/20130811-5-US-2037.00061>
- Monostori, L. (2014). Cyber-physical Production Systems : Roots, Expectations and R&D Challenges. *Procedia CIRP*, 17, 9-13. <https://doi.org/10.1016/j.procir.2014.03.115>
- Monsaingeon, N., Caroux, L., Langlois, S., Hurgobin, Y., & Lemercier, C. (2020). Driver compliance with automation reliability information regarding hazardous environmental circumstances. *Le travail humain*, 83(4), 343-360. <https://doi.org/10.3917/th.834.0343>
- Moore, J. C., & Whinston, A. B. (1986). A model of decision-making with sequential information-acquisition (Part 1). *Decision Support Systems*, 2(4), 285-307. [https://doi.org/10.1016/0167-9236\(86\)90001-1](https://doi.org/10.1016/0167-9236(86)90001-1)
- Morgan, P., Williams, C., Flower, J., Alford, C., & Parkin, J. (2019). *Trust in an Autonomously Driven Simulator and Vehicle Performing Maneuvers at a T-Junction with and Without Other Vehicles* (p. 363-375). https://doi.org/10.1007/978-3-319-93885-1_33
- Morrison, J. G., Cohen, D., & Gluckman, J. P. (1993). Prospective principles and guidelines for the design of adaptively automated crewstations. *The adaptive function allocation for intelligent cockpits (AFAIC) program: Interim research and guidelines for the application of adaptive automation*, 1-7.
- Moulières-Seban, T. (2017). *Conception de systèmes cobotiques industriels : Approche cognitive : application à la production pyrotechnique au sein d'Ariane Group* [These de doctorat, Bordeaux]. <http://www.theses.fr/2017BORD0774>

- N**aiseh, M., Al-Thani, D., Jiang, N., & Ali, R. (2021). Explainable recommendation : When design meets trust calibration. *World Wide Web*, 24. <https://doi.org/10.1007/s11280-021-00916-0>
- Nguma, H. (2020). Étude de l'effet de synchronie pour la conception d'IHM adaptatives : Effet sur la vigilance, la charge mentale, la conscience de la situation et les performances. In [Http://www.theses.fr](http://www.theses.fr). Université Côte d'Azur. <http://www.theses.fr/2020COAZ2007>
- Nguyen, T., Lim, C. P., Nguyen, N. D., Gordon-Brown, L., & Nahavandi, S. (2019). A Review of Situation Awareness Assessment Approaches in Aviation Environments. *IEEE Systems Journal*, 13(3), 3590-3603. <https://doi.org/10.1109/JSYST.2019.2918283>
- Nurunnabi, M. (2021). Disclosure, Transparency, and International Financial Reporting Standards. In *International Financial Reporting Standards Implementation : A Global Experience* (p. 199-311). Emerald Publishing Limited. <https://doi.org/10.1108/978-1-80117-440-420211005>
- O**kamura, K., & Yamada, S. (2020). Adaptive trust calibration for human-AI collaboration. *PLOS ONE*, 15(2), e0229132. <https://doi.org/10.1371/journal.pone.0229132>
- Orasanu, J., Dismukes, R. K., & Fischer, U. (1993). Decision Errors in the Cockpit. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 37(4), 363-367. <https://doi.org/10.1177/154193129303700423>
- Osofsky, S., Sanders, T., Jentsch, F., Hancock, P., & Chen, J. (2014). *Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems*. 9084, 90840E. <https://doi.org/10.1117/12.2050622>
- Ötting, S. K., Masjutin, L., Steil, J. J., & Maier, G. W. (2020). Let's Work Together : A Meta-Analysis on Robot Design Features That Enable Successful Human–Robot Interaction at Work: *Human Factors*. <https://doi.org/10.1177/0018720820966433>
- P**acaux-Lemoine, M. P., & Debernard, S. (2002). Common work space for human–machine cooperation in air traffic control. *Control Engineering Practice*, 10(5), 571-576. [https://doi.org/10.1016/S0967-0661\(01\)00060-0](https://doi.org/10.1016/S0967-0661(01)00060-0)
- Pacaux-Lemoine, M.-P. (2020). *HUMAN-MACHINE COOPERATION : Adaptability of shared functions between Humans and Machines - Design and evaluation aspects* [Habilitation à diriger des recherches, Université Polytechnique Hauts-de-France]. <https://hal.archives-ouvertes.fr/tel-02959402>
- Parasuraman, R., & Mouloua, M. (1996). *Automation and Human Performance : Theory and Applications*. Routledge.
- Parasuraman, R., & Riley, V. (1997). Humans and Automation : Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(2), 230-253. <https://doi.org/10.1518/001872097778543886>

- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 30(3), 286-297. <https://doi.org/10.1109/3468.844354>
- Parasuraman, R., Sheridan, T., & Wickens, C. (2008). Situation Awareness, Mental Workload, and Trust in Automation : Viable, Empirically Supported Cognitive Engineering Constructs. *Journal of Cognitive Engineering and Decision Making*, 2, 140-160. <https://doi.org/10.1518/155534308X284417>
- Patel, J., Ramaswamy, T., Li, Z., & Pinciroli, C. (2021). Transparency in Multi-Human Multi-Robot Interaction. *arXiv:2101.10495 [cs]*. <http://arxiv.org/abs/2101.10495>
- Pedersen, B. K. M. K., Andersen, K. E., Kösllich, S., Weigelin, B. C., & Kuusinen, K. (2018). Simulations and self-driving cars : A study of trust and consequences. *Companion of the 2018 ACM/IEEE international conference on human-robot interaction*, 205-206.
- Perello-March, J. R., Burns, C. G., Birrell, S. A., Woodman, R., & Elliott, M. T. (2022). Physiological Measures of Risk Perception in Highly Automated Driving. *IEEE Transactions on Intelligent Transportation Systems*, 23(5), 4811-4822. <https://doi.org/10.1109/TITS.2022.3146793>
- Peres, R. S., Jia, X., Lee, J., Sun, K., Colombo, A. W., & Barata, J. (2020). Industrial Artificial Intelligence in Industry 4.0—Systematic Review, Challenges and Outlook. *IEEE Access*, 8, 220121-220139. <https://doi.org/10.1109/ACCESS.2020.3042874>
- Petersen, L., Robert, L., Yang, J., & Tilbury, D. (2019). *Situational Awareness, Driver's Trust in Automated Driving Systems and Secondary Task Performance* (SSRN Scholarly Paper N° 3345543). <https://doi.org/10.2139/ssrn.3345543>
- Petit, J. (2022). *Évaluation du risque ressenti par le passager d'un véhicule autonome naviguant parmi des piétons* [These de doctorat, Nantes Université]. <https://www.theses.fr/fr/2022NANU2005>
- Pokam Meguia, R., Debernard, S., Chauvin, C., & Langlois, S. (2019). The Design of an Interface According to Principles of Transparency. In *Automation Challenges of Socio-technical Systems* (p. 111-150). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119644576.ch4>
- Pokam, R., Debernard, S., Chauvin, C., & Langlois, S. (2019). Principles of transparency for autonomous vehicles : First results of an experiment with an augmented reality human-machine interface. *Cognition, Technology & Work*, 21(4), 643-656. <https://doi.org/10.1007/s10111-019-00552-9>
- Prewett, M. S., Johnson, R. C., Saboe, K. N., Elliott, L. R., & Coovert, M. D. (2010). Managing workload in human-robot interaction : A review of empirical studies. *Computers in Human Behavior*, 26(5), 840-856. <https://doi.org/10.1016/j.chb.2010.03.010>
- Pynadath, D. V., Gurney, N., & Wang, N. (2022). Explainable Reinforcement Learning in Human-Robot Teams : The Impact of Decision-Tree Explanations on Transparency. *2022 31st IEEE International Conference on Robot and Human Interactive*

Communication (RO-MAN), 749-756. <https://doi.org/10.1109/RO-MAN53752.2022.9900608>

- R**ajabiyazdi, F., & Jamieson, G. A. (2020). A Review of Transparency (seeing-into) Models. *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 302-308. <https://doi.org/10.1109/SMC42975.2020.9282970>
- Rajkumar, R. (Raj), Lee, I., Sha, L., & Stankovic, J. (2010). Cyber-physical systems : The next computing revolution. *Proceedings of the 47th Design Automation Conference*, 731-736. <https://doi.org/10.1145/1837274.1837461>
- Rao, A. S., & Georgeff, M. P. (1995). BDI agents : From theory to practice. *Icmas, 95*, 312-319.
- Rasch, R., Kott, A., & Forbus, K. D. (2003). Incorporating AI into military decision making : An experiment. *IEEE Intelligent Systems*, 18(4), 18-26.
- Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE transactions on systems, man, and cybernetics*, 3, 257-266.
- Rasmussen, J., & Vicente, K. J. (1989). Coping with human errors through system design : Implications for ecological interface design. *International Journal of Man-Machine Studies*, 31(5), 517-534. [https://doi.org/10.1016/0020-7373\(89\)90014-X](https://doi.org/10.1016/0020-7373(89)90014-X)
- Rauffet, P., Chauvin, C., Morel, G., & Berruet, P. (2015). Designing sociotechnical systems : A CWA-based method for dynamic function allocation. *Proceedings of the European Conference on Cognitive Ergonomics 2015*, 1-8. <https://doi.org/10.1145/2788412.2788433>
- Rauffet, P., Guerin, C., Chauvin, C., & Martin, E. (2018, octobre). Contribution of Industry 4.0 to the emergence of a joint cognitive and physical production system. *HFES European Chapter*. <https://hal.archives-ouvertes.fr/hal-01893007>
- Rauffet, P., Simon, L., & Guerin, C. (2023, juin 14). *Transparence et Confiance au sein des équipes Humains Systèmes de l'Industrie 4.0*. CIGI QUALITA MOSIM 2023 - Propulser la performance. Interconnectivité et collaboration dans un contexte d'intelligence artificielle. <https://hal.science/hal-04111778>
- Reiman, A., Kaivo-oja, J., Parviainen, E., Takala, E.-P., & Lauraeus, T. (2021). Human factors and ergonomics in manufacturing in the industry 4.0 context – A scoping review. *Technology in Society*, 65, 101572. <https://doi.org/10.1016/j.techsoc.2021.101572>
- Reinecke, K., & Bernstein, A. (2011). Improving performance, perceived usability, and aesthetics with culturally adaptive user interfaces. *ACM Transactions on Computer-Human Interaction*, 18(2), 8:1-8:29. <https://doi.org/10.1145/1970378.1970382>
- Rogalski, J. (1994). Formation Aux Activités Collectives. *Le Travail Humain*, 57(4), 367-386.

- Romero, D., Bernus, P., Noran, O., Stahre, J., & Fast-Berglund, Å. (2016a). The Operator 4.0 : Human Cyber-Physical Systems & Adaptive Automation Towards Human-Automation Symbiosis Work Systems. In I. Nääs, O. Vendrametto, J. Mendes Reis, R. F. Gonçalves, M. T. Silva, G. von Cieminski, & D. Kiritsis (Éds.), *Advances in Production Management Systems. Initiatives for a Sustainable World* (p. 677-686). Springer International Publishing. https://doi.org/10.1007/978-3-319-51133-7_80
- Romero, D., Stahre, J., Wuest, T., Noran, O., Bernus, P., Fast-Berglund, Å., & Gorecky, D. (2016b). Towards an operator 4.0 typology : A human-centric perspective on the fourth industrial revolution technologies. *proceedings of the international conference on computers and industrial engineering (CIE46), Tianjin, China*, 29-31.
- Rosin, F., Forget, P., Lamouri, S., & Pellerin, R. (2022). Enhancing the Decision-Making Process through Industry 4.0 Technologies. *Sustainability*, 14(1), Article 1. <https://doi.org/10.3390/su14010461>
- Rossi, A., Holthaus, P., Dautenhahn, K., Koay, K. L., & Walters, M. L. (2018). Getting to know Pepper : Effects of people's awareness of a robot's capabilities on their trust in the robot. *Proceedings of the 6th international conference on human-agent interaction*, 246-252.
- Roth, G., Schulte, A., Schmitt, F., & Brand, Y. (2020). Transparency for a Workload-Adaptive Cognitive Agent in a Manned–Unmanned Teaming Application. *IEEE Transactions on Human-Machine Systems*, 50(3), 225-233. <https://doi.org/10.1109/THMS.2019.2914667>
- Rothrock, L., Koubek, R., Fuchs, F., Haas, M., & Salvendy, G. (2002). Review and reappraisal of adaptive interfaces : Toward biologically inspired paradigms. *Theoretical Issues in Ergonomics Science*, 3(1), 47-84. <https://doi.org/10.1080/14639220110110342>
- Roundtree, K. A., Goodrich, M. A., & Adams, J. A. (2019). Transparency : Transitioning From Human–Machine Systems to Human-Swarm Systems. *Journal of Cognitive Engineering and Decision Making*, 13(3), 171-195. <https://doi.org/10.1177/1555343419842776>
- Rüßmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P., & Harnisch, M. (2015). Industry 4.0 : The future of productivity and growth in manufacturing industries. *Boston Consulting Group*, 9(1), 54-89.
- S**adler, G., Battiste, H., Ho, N., Hoffmann, L., Johnson, W., Shively, R., Lyons, J., & Smith, D. (2016). Effects of transparency on pilot trust and agreement in the autonomous constrained flight planner. *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, 1-9. <https://doi.org/10.1109/DASC.2016.7777998>
- Saiwaki, N., Togashi, H., Tsujimoto, H., & Nishida, S. (1996). An adaptive interface based on physiological indices. *1996 IEEE International Conference on Systems, Man and Cybernetics. Information Intelligence and Systems (Cat. No.96CH35929)*, 4, 2793-2798 vol.4. <https://doi.org/10.1109/ICSMC.1996.561383>

- Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015). Would You Trust a (Faulty) Robot? : Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 141-148. <https://doi.org/10.1145/2696454.2696497>
- Sanders, T., Kaplan, A., Koch, R., Schwartz, M., & Hancock, P. A. (2019). The relationship between trust and use choice in human-robot interaction. *Human factors*, 61(4), 614-626.
- Sanders, T., Oleson, K. E., Billings, D. R., Chen, J. Y., & Hancock, P. A. (2011). A model of human-robot trust : Theoretical model development. *Proceedings of the human factors and ergonomics society annual meeting*, 55(1), 1432-1436.
- Sarter, N. (2007). Coping with Complexity Through Adaptive Interface Design. *International Conference on Human-Computer Interaction*, 493-498.
- Sato, T., Yamani, Y., Liechty, M., & Chancey, E. T. (2020). Automation trust increases under high-workload multitasking scenarios involving risk. *Cognition, Technology & Work*, 22(2), 399-407. <https://doi.org/10.1007/s10111-019-00580-5>
- Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. *Automation and human performance: Theory and applications*, 37-63.
- Schaefer, K. (2013). *The perception and measurement of human-robot trust*.
- Schemmer, M., Hemmer, P., Nitsche, M., Kühn, N., & Vössing, M. (2022). A Meta-Analysis of the Utility of Explainable Artificial Intelligence in Human-AI Decision-Making. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 617-626. <https://doi.org/10.1145/3514094.3534128>
- Schmidt, K., Rasmussen, J., Brehmer, B., & Leplat, J. (1991). Cooperative work : A conceptual framework. *Distributed decision making: Cognitive models for cooperative work*, 75-110.
- Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2016). Empowerment in veterinary clinics : The role of trust in delegation. *Journal of Trust Research*, 6(1), 76-90. <https://doi.org/10.1080/21515581.2016.1153479>
- Seguin, C., Rioual, Y., Diguët, J.-P., & Gogniat, G. (2022). Data Extraction and Deep Learning Method for Predictive Maintenance in Vessel's Engine Room. *32nd European Safety and Reliability Conference (ESREL 2022)*, 1983-1990.
- Selkowitz, A. R., Lakhmani, S. G., Larios, C. N., & Chen, J. Y. C. (2016). Agent Transparency and the Autonomous Squad Member. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 60(1), 1319-1323. <https://doi.org/10.1177/1541931213601305>
- Selkowitz, A. R., Larios, C. A., Lakhmani, S. G., & Chen, J. Y. C. (2017). Displaying Information to Support Transparency for Autonomous Platforms. In P. Savage-Knepshield & J. Chen (Éds.), *Advances in Human Factors in Robots and Unmanned Systems* (p. 161-173). Springer International Publishing. https://doi.org/10.1007/978-3-319-41959-6_14

- Senoner, J., Netland, T., & Feuerriegel, S. (2022). Using Explainable Artificial Intelligence to Improve Process Quality: Evidence from Semiconductor Manufacturing. *Management Science*, 68(8), 5704-5723. <https://doi.org/10.1287/mnsc.2021.4190>
- Shu, P., Min, C., Bodala, I., Nikolaidis, S., Hsu, D., & Soh, H. (2018). Human trust in robot capabilities across tasks. *Companion of the 2018 ACM/IEEE international conference on human-robot interaction*, 241-242.
- Siegrist, M., & Árvai, J. (2020). Risk perception : Reflections on 40 years of research. *Risk Analysis*, 40(S1), 2191-2206.
- Simon, L., Guérin, C., & Rauffet, P. (2023b, juillet 4). *Utilisation dynamique de la transparence pour la coopération humain-machine*. 12e Colloque EPIQUE Paris. <https://hal.science/hal-04111732>
- Simon, L., Guérin, C., Rauffet, P., & Diguët, J.-P. (2022a). Integrating Transparency to Ecological Interface Design. *Proceedings of the 32nd European Safety and Reliability Conference (ESREL 2022)*, 3301-3308. <https://hal.archives-ouvertes.fr/hal-03601183>
- Simon, L., Guérin, C., Rauffet, P., & Lassalle, J. (2021a). Using cognitive work analysis to develop predictive maintenance tool for vessels. *31st European Safety and Reliability Conference*.
- Simon, L., Lassalle, J. (2022). Tâche 5.1 SEANATIC: Définition des scénarios d'utilisation pour l'évaluation de l'IHS embarquée (simulation/réel) et travail préparatoire des expérimentation et tâche 5.2 SEANATIC: Résultats des évaluations psycho-ergonomiques et cognitives des interfaces Humain-Système déployées (environnement simulé) et recommandations pour évolution [Rapport ADEME L5.1_1 et L5.1_2]. Université Bretagne Sud
- Simon, L., Rauffet, P. (2022). Tâche 4.2 SEANATIC: Développement de métriques sur l'état de l'opérateur et sur la criticité de la situation. Création de règles et d'algorithmes pour filtrer les informations ou modifier leur présentation en fonction de ces métriques. [Rapport ADEME L4.2]. Université Bretagne Sud
- Simon, L., Rauffet, P., & Guérin, C. (2021b). Développement de règles de conception et des mécanismes adaptatifs des interfaces dans le dialogue humain-machine- Application à l'industrie 4.0. *11e Colloque EPIQUE: Doctoriales 2021*.
- Simon, L., Rauffet, P., Guérin, C., & Lassalle, J. (2021c). Exploiter la méthode CWA pour le design écologique d'une interface de maintenance 4.0. *Onzième colloque de Psychologie Ergonomique EPIQUE 2021*, 210.
- Simon, L., Rauffet, P., Guérin, C., & Seguin, C. (2022b, juillet). Trust in an autonomous agent for predictive maintenance: How agent transparency could impact compliance. *13th AHFE Conference (Applied Human Factors and Ergonomics) 2022*. <https://doi.org/10.54941/ahfe1001602>
- Simon, L., Guérin, C., Rauffet, P., Chauvin, C., & Martin, É. (2023a). How Humans Comply With a (Potentially) Faulty Robot : Effects of Multidimensional Transparency. *IEEE Transactions on Human-Machine Systems*, 1-10. <https://doi.org/10.1109/THMS.2023.3273773>

- Sirkin, D., Martelaro, N., Johns, M., & Ju, W. (2017). Toward measurement of situation awareness in autonomous vehicles. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 405-415.
- Sitkin, S. B., & Pablo, A. L. (1992). Reconceptualizing the determinants of risk behavior. *Academy of management review*, 17(1), 9-38.
- Sjöberg, L., Moen, B.-E., & Rundmo, T. (2004). Explaining risk perception. *An evaluation of the psychometric paradigm in risk perception research*, 10(2), 665-612.
- Skraaning, G., & Jamieson, G. A. (2021). Human Performance Benefits of The Automation Transparency Design Principle : Validation and Variation. *Human Factors*, 63(3), 379-401. <https://doi.org/10.1177/0018720819887252>
- Slovic, P. E. (2000). *The perception of risk*. Earthscan publications.
- Sonoda, K., & Wada, T. (2017). Displaying System Situation Awareness Increases Driver Trust in Automated Driving. *IEEE Transactions on Intelligent Vehicles*, 2(3), 185-193. <https://doi.org/10.1109/TIV.2017.2749178>
- Soria-Oliver, M., López, J. S., & Torrano, F. (2017). Relations between mental workload and decision-making in an organizational setting. *Psicologia: Reflexão e Crítica*, 30. <https://doi.org/10.1186/s41155-017-0061-0>
- Stanners, M., & French, H. T. (2005). *An empirical study of the relationship between situation awareness and decision making*. DEFENCE SCIENCE AND TECHNOLOGY ORGANIZATION EDINBURGH (AUSTRALIA) LAND
- Stanton, N. A., Salmon, P. M., Walker, G. H., & Jenkins, D. P. (2017). *Cognitive Work Analysis : Applications, Extensions and Future Directions*. CRC Press.
- Stauder, M., & Kühn, N. (2021). AI for in-line vehicle sequence controlling : Development and evaluation of an adaptive machine learning artifact to predict sequence deviations in a mixed-model production line. *Flexible Services and Manufacturing Journal*, 1-39.
- Steyvers, M., & Kumar, A. (2022). *Three Challenges for AI-Assisted Decision-Making*.
- Steyvers, M., Tejada, H., Kerrigan, G., & Smyth, P. (2022). Bayesian modeling of human–AI complementarity. *Proceedings of the National Academy of Sciences*, 119(11), e2111547119. <https://doi.org/10.1073/pnas.2111547119>
- Stowers, K., Kasdaglis, N., Newton, O., Lakhmani, S., Wohleber, R., & Chen, J. (2016). Intelligent Agent Transparency : The Design and Evaluation of an Interface to Facilitate Human and Intelligent Agent Collaboration. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 60(1), 1706-1710. <https://doi.org/10.1177/1541931213601392>
- Stuck, R. E., Tomlinson, B. J., & Walker, B. N. (2021). The importance of incorporating risk into human-automation trust. *Theoretical Issues in Ergonomics Science*, 0(0), 1-17. <https://doi.org/10.1080/1463922X.2021.1975170>

Szaniawski, K. (1980). Philosophy of decision making. *Acta Psychologica*, 45(1), 327-341. [https://doi.org/10.1016/0001-6918\(80\)90041-4](https://doi.org/10.1016/0001-6918(80)90041-4)

Tao, D., Tan, H., Wang, H., Zhang, X., Qu, X., & Zhang, T. (2019). A systematic review of physiological measures of mental workload. *International journal of environmental research and public health*, 16(15), 2716.

Tattersall, A. J., & Foord, P. S. (1996). An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics*, 39(5), 740-748. <https://doi.org/10.1080/00140139608964495>

Taylor, R. M. (2011). Situational Awareness Rating Technique (Sart) : The Development of a Tool for Aircrew Systems Design. In *Situational Awareness*. Routledge.

Theodorou, A., Wortham, R. H., & Bryson, J. J. (2017). Designing and implementing transparency for real time inspection of autonomous robots. *Connection Science*, 29(3), 230-241. <https://doi.org/10.1080/09540091.2017.1310182>

Tsai, Y.-F., Viirre, E., Strychacz, C., Chase, B., & Jung, T.-P. (2007). Task performance and eye activity : Predicting behavior relating to cognitive workload. *Aviation, space, and environmental medicine*, 78(5), B176-B185.

Ulahannan, A., Jennings, P., Oliveira, L., & Birrell, S. (2020). Designing an Adaptive Interface : Using Eye Tracking to Classify How Information Usage Changes Over Time in Partially Automated Vehicles. *IEEE Access*, 8, 16865-16875. <https://doi.org/10.1109/ACCESS.2020.2966928>

Van de Merwe, K., Mallam, S., & Nazir, S. (2022). Agent Transparency, Situation Awareness, Mental Workload, and Operator Performance : A Systematic Literature Review. *Human Factors*, 00187208221077804. <https://doi.org/10.1177/00187208221077804>

Vicente, K. J., Christoffersen, K., & Perekhita, A. (1995). Supporting operator problem solving through ecological interface design. *IEEE Transactions on Systems, Man, and Cybernetics*, 25(4), 529-545. <https://doi.org/10.1109/21.370186>

Wang, N., Pynadath, D. V., & Hill, S. G. (2016). Trust calibration within a human-robot team : Comparing automatically generated explanations. *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 109-116. <https://doi.org/10.1109/HRI.2016.7451741>

Wei, W., Shao, Z., Chen, W., Qiao, R., Yuan, Y., & Cheng, J. (2021). Heating process and damage evolution of microwave absorption and transparency materials under microwave irradiation. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*, 7(3), 86. <https://doi.org/10.1007/s40948-021-00284-z>

Weigelin, B. C., Mathiesen, M., Nielsen, C., Fischer, K., & Nielsen, J. (2018). Trust in Medical Human-Robot Interactions based on Kinesthetic guidance. *2018 27th IEEE*

- International Symposium on Robot and Human Interactive Communication (ROMAN)*, 901-908. <https://doi.org/10.1109/ROMAN.2018.8525677>
- Weiss, A., Wortmeier, A.-K., & Kubicek, B. (2021). Cobots in Industry 4.0 : A Roadmap for Future Practice Studies on Human–Robot Collaboration. *IEEE Transactions on Human-Machine Systems*, 51(4), 335-345. <https://doi.org/10.1109/THMS.2021.3092684>
- Weistroffer, V. (2014). *Étude des conditions d'acceptabilité de la collaboration homme-robot en utilisant la réalité virtuelle* [Phdthesis, Ecole Nationale Supérieure des Mines de Paris]. <https://pastel.archives-ouvertes.fr/tel-01144569>
- Wellsandt, S., Klein, K., Hribernik, K., Lewandowski, M., Bousdekis, A., Mentzas, G., & Thoben, K.-D. (2021). Towards Using Digital Intelligent Assistants to Put Humans in the Loop of Predictive Maintenance Systems. *IFAC-PapersOnLine*, 54(1), 49-54.
- Williams, D. J., & Noyes, J. M. (2007). How does our perception of risk influence decision-making? Implications for the design of risk information. *Theoretical Issues in Ergonomics Science*, 8(1), 1-35. <https://doi.org/10.1080/14639220500484419>
- Wilson, G. F. (2000). Strategies for Psychophysiological. *Situation awareness analysis and measurement*, 175.
- Wilson, R., Zwickle, A., & Walpole, H. (2018). Developing a Broadly Applicable Measure of Risk Perception. *Risk Analysis*, 39. <https://doi.org/10.1111/risa.13207>
- Woods, D., Patterson, E., & Roth, E. (2002). Can We Ever Escape from Data Overload? A Cognitive Systems Diagnosis. *Cognition, Technology & Work*, 4, 22-36. <https://doi.org/10.1007/s101110200002>
- Wortham, R. H. (2020). *Transparency for Robots and Autonomous Systems : Fundamentals, technologies and applications*. Institution of Engineering and Technology. <https://doi.org/10.1049/PBCE130E>
- Wortham, R. H., & Theodorou, A. (2017). Robot transparency, trust and utility. *Connection Science*, 29(3), 242-248. <https://doi.org/10.1080/09540091.2017.1313816>
- Wright, J. L., Chen, J. Y. C., Barnes, M. J., & Hancock, P. A. (2017). The Effect of Agent Reasoning Transparency on Complacent Behavior : An Analysis of Eye Movements and Response Performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61(1), 1594-1598. <https://doi.org/10.1177/1541931213601762>
- Wright, J. L., Chen, J. Y. C., & Lakhmani, S. G. (2020). Agent Transparency and Reliability in Human–Robot Interaction : The Influence on User Confidence and Perceived Reliability. *IEEE Transactions on Human-Machine Systems*, 50(3), 254-263. <https://doi.org/10.1109/THMS.2019.2925717>
- Wu, N., Phang, J., Park, J., Shen, Y., Huang, Z., Zorin, M., Jastrzębski, S., Févry, T., Katsnelson, J., & Kim, E. (2019). Deep neural networks improve radiologists' performance in breast cancer screening. *IEEE transactions on medical imaging*, 39(4), 1184-1194.

- X**ie, Y., Bodala, I. P., Ong, D. C., Hsu, D., & Soh, H. (2019). Robot capability and intention in trust-based decisions across tasks. *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 39-47
- Xu, X., Lu, Y., Vogel-Heuser, B., & Wang, L. (2021). Industry 4.0 and Industry 5.0—Inception, conception and perception. *Journal of Manufacturing Systems*, *61*, 530-535. <https://doi.org/10.1016/j.jmsy.2021.10.006>.
- Y**ang, X. J., Unhelkar, V. V., Li, K., & Shah, J. A. (2017). Evaluating Effects of User Experience and System Transparency on Trust in Automation. *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 408-416.
- Yang, X., Schemanske, C., & Searle, C. (2021). Toward Quantifying Trust Dynamics : How People Adjust Their Trust After Moment-to-Moment Interaction With Automation. *Human Factors*, 00187208211034716. <https://doi.org/10.1177/00187208211034716>
- Yang, Y., Karakaya, B., Dominiononi, G. C., Kawabe, K., & Bengler, K. (2018). An HMI Concept to Improve Driver's Visual Behavior and Situation Awareness in Automated Vehicle. *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 650-655. <https://doi.org/10.1109/ITSC.2018.8569986>
- Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science : Mental workload in ergonomics. *Ergonomics*, *58*(1), 1-17. <https://doi.org/10.1080/00140139.2014.956151>
- Z**ang, J., & Jeon, M. (2022). The effects of transparency and reliability of in-vehicle intelligent agents on driver perception, takeover performance, workload and situation awareness in conditionally automated vehicles. *Multimodal Technologies and Interaction*, *6*(9), 82.
- Zarte, M., Pechmann, A., & Nunes, I. L. (2020). Principles for Human-Centered System Design in Industry 4.0 – A Systematic Literature Review. In I. L. Nunes (Éd.), *Advances in Human Factors and Systems Interaction* (p. 140-147). Springer International Publishing. https://doi.org/10.1007/978-3-030-51369-6_19
- Zhang, D. C., Highhouse, S., & Nye, C. D. (2019). Development and validation of the General Risk Propensity Scale (GRiPS). *Journal of Behavioral Decision Making*, *32*(2), 152-167. <https://doi.org/10.1002/bdm.2102>
- Zhang, Y., Liao, Q. V., & Bellamy, R. K. E. (2020). Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 295-305. <https://doi.org/10.1145/3351095.3372852>

Webographie :

France 2030 : Un appel à manifestation d'intérêt pour soutenir l'offre de solutions pour l'industrie du futur. (s. d.). Consulté 20 mars 2023, à l'adresse <https://www.economie.gouv.fr/france-2030-appel-manifestation-interet-soutenir-offre-solutions-industrie-futur>

Horizon Europe Calls 2023—Destination 6. A human-centred and ethical development of digital and industrial technologies. (2022, décembre 8). <https://hadea.ec.europa.eu/calls-proposals/horizon-europe-calls-2023-destination-6-human-centred-and-ethical-development-digital-and-industrial-en>

INDUSTRIE : Définition de INDUSTRIE. (s. d.). Consulté 3 avril 2023, à l'adresse <https://www.cnrtl.fr/definition/industrie>

Larousse, É. (s. d.). *Définitions : Confiance - Dictionnaire de français Larousse.* Consulté 29 septembre 2022, à l'adresse <https://www.larousse.fr/dictionnaires/francais/confiance/18082>

MAINTENANCE : Définition de MAINTENANCE. (s. d.). Consulté 31 mai 2023, à l'adresse <https://www.cnrtl.fr/definition/maintenance>

NF EN 13306—Janvier 2018. (s. d.). Consulté 18 décembre 2020, à l'adresse <https://www.boutique.afnor.org/norme/nf-en-13306/maintenance-terminologie-de-la-maintenance/article/828621/fa185755>

PLANIFICATION : Définition de PLANIFICATION. (s. d.). Consulté 11 juin 2021, à l'adresse <https://www.cnrtl.fr/definition/planification>

PRODUCTION : Définition de PRODUCTION. (s. d.). Consulté 11 juin 2021, à l'adresse <https://www.cnrtl.fr/definition/production>

SUPERVISION : Définition de SUPERVISION. (s. d.). Consulté 11 juin 2021, à l'adresse <https://www.cnrtl.fr/definition/supervision>

Médiagraphie :

ScienceEtonnante (Réalisateur). (2023, avril 14). *Ce qui se cache derrière le fonctionnement de ChatGPT*. <https://www.youtube.com/watch?v=7ell8KEbhJo>

Logiciel :

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Annexes

Annexe A : Consentement libre et éclairé Humanism

Formulaire d'information et de consentement libre et éclairé

Vous êtes invité(e) à participer à un projet de recherche. Le présent document vous renseigne sur les modalités de ce projet de recherche. S'il y a des mots ou des paragraphes que vous ne comprenez pas, n'hésitez pas à poser des questions. Pour participer à ce projet de recherche, vous devrez signer le consentement à la fin de ce document et nous vous en remettrons une copie signée et datée.

Titre du projet : ANR-HUMANISM

Sous la direction de : Clément Guérin

Promoteur : Université Bretagne Sud, Laboratoire Lab-STICC, Rue de Saint-Maudé, 56100 Lorient

Investigateur principal : Loïck Simon

Objectif du projet : Comprendre les aspects cognitifs et psychologiques mis en œuvre lors de la réalisation d'une tâche de coopération homme-cobot dans le cadre d'un travail d'opérateur de préparation de commande.

Raison et nature du projet : Votre participation sera requise pour utiliser une simulation sur ordinateur et répondre aux demandes du cobot. Cette simulation sera ponctuée de questionnaires, il n'y a pas de bonnes ou de mauvaises réponses dans ces questionnaires. Votre activité cardiaque ainsi que votre activité oculaire seront observées durant cette passation. La passation durera au maximum 1h. Cette passation se déroule au sein de l'ENSIBS, 17 Boulevard Flandres Dunkerque 1940, 56100 Lorient

Avantages pouvant découler de la participation : Votre participation à ce projet de recherche contribuera à l'avancement des connaissances entourant la collaboration homme-cobot dans le cadre des industries et aidera à l'amélioration des conditions de travail des opérateurs.

Inconvénients et risques pouvant découler de la participation : Votre participation à la recherche ne devrait pas comporter d'inconvénients significatifs, si ce n'est le fait de donner de votre temps.

Droit de retrait sans préjudice de la participation : Il est entendu que votre participation à ce projet de recherche est tout à fait volontaire et que vous restez libre, à tout moment, de mettre fin à votre participation sans avoir à motiver votre décision ni à subir de préjudice de quelque nature que ce soit.

Confidentialité, partage, surveillance et publications : Durant votre participation à ce projet de recherche, le chercheur responsable ainsi que son personnel recueilleront et consigneront dans un dossier de recherche les renseignements vous concernant. Seuls les renseignements nécessaires à la bonne conduite du projet de recherche seront recueillis. Ils peuvent comprendre les informations suivantes : nom, sexe, date de naissance, origine ethnique, photographies, enregistrements vidéo ou audio, habitudes de vie, résultats de tous les tests, examens et procédures que vous aurez à subir lors de ce projet, etc.

Tous les renseignements recueillis au cours du projet de recherche demeureront strictement confidentiels dans les limites prévues par la loi. Afin de préserver votre identité et la confidentialité de ces renseignements, vous ne serez identifié(e) que par un numéro de code. La clé du code reliant votre nom à votre dossier de recherche sera conservée par le chercheur responsable du projet de recherche.

Le chercheur principal de l'étude utilisera les données à des fins de recherche dans le but de répondre aux objectifs scientifiques du projet de recherche décrits dans ce formulaire d'information et de consentement.

Les données du projet de recherche pourront être publiées dans des revues scientifiques ou partagées avec d'autres personnes lors de discussions scientifiques. Aucune publication ou communication scientifique ne renfermera d'information permettant de vous identifier. Dans le cas contraire, votre permission vous sera demandée au préalable.

Les données recueillies seront conservées, pour une période n'excédant pas 5 ans. Après cette période, les données seront détruites.

Enregistrement vidéo : Durant cette passation un enregistrement vidéo sera réalisé via les lunettes. N'apparaîtra dans le champ de la caméra que vos mains. Seuls l'écran de la simulation et vos mains seront dans le champ de la caméra. Nous ne filmerons pas les visages. Nous aimerions pouvoir utiliser les images ces dernières, avec votre permission, à des fins de formation et/ou de présentations scientifiques.

Il n'est cependant pas nécessaire de consentir à ce volet pour participer au présent projet. Si vous refusez, les enregistrements vidéos vous concernant seront détruits à la fin du projet dans le respect de la confidentialité.

Nous autorisez-vous à utiliser vos enregistrements à des fins de formations ou de présentations scientifiques et à les conserver avec vos données de recherche ?

Oui Non

Résultats de la recherche et publication : Si vous le souhaitez, vous serez informé des résultats de la recherche et des publications qui en découleront, le cas échéant. A tout moment, vous pouvez solliciter la chercheuse à ce sujet, elle mettra en œuvre les moyens nécessaires pour vous répondre. Nous préserverons l'anonymat des personnes ayant participé à l'étude.

Consentement libre et éclairé :

Je, _____ (nom en caractères d'imprimerie), déclare avoir lu et/ou compris le présent formulaire et j'en ai reçu un exemplaire. Je comprends la nature et le motif de ma participation au projet. J'ai eu l'occasion de poser des questions auxquelles on a répondu, à ma satisfaction.

Par la présente, j'accepte librement de participer au projet.

Signature de la participante ou du participant : _____

Fait à _____, le _____ 201_

Déclaration de responsabilité des chercheurs de l'étude

Je, _____, chercheur principal de l'étude, déclare que les chercheurs collaborateurs ainsi que mon équipe de recherche sommes responsables du déroulement du présent projet de recherche. Nous nous engageons à respecter les obligations énoncées dans ce document et également à vous informer de tout élément qui serait susceptible de modifier la nature de votre consentement.

Signature : _____

Je, _____, certifie avoir expliqué à la participante ou au participant intéressé(e) les termes du présent formulaire, avoir répondu aux questions qu'il ou qu'elle m'a posées à cet égard et lui avoir clairement indiqué qu'il ou qu'elle reste, à tout moment, libre de mettre un terme à sa participation au projet de recherche décrit ci-dessus. Je m'engage à garantir le respect des objectifs de l'étude et à respecter la confidentialité.

Signature : _____

Fait à _____,

Annexe B : Questionnaire Humanism

B.1 - Poker

Question	Réponses			
Je joue à des jeux de type Poker, Black Jack, etc.	Très fréquemment	Fréquemment	Rarement	Jamais

B.2 - Propension au risque

	1 : Pas du tout d'accord 7 : Tout à fait d'accord
Je pourrais investir dans une entreprise qui a des chances de faire faillite :	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Je prends mes décisions selon l'adage « qui ne tente rien, n'a rien » :	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

B.3 - Propension à la confiance

	1 : Pas du tout d'accord 7 : Tout à fait d'accord
En règle générale je fais confiance aux systèmes technologiques :	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
A priori je ferais confiance à ce Cobot :	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

B.4 - CAI

	1 : Pas du tout d'accord 7 : Tout à fait d'accord
J'ai trouvé le cobot compréhensible :	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
J'ai confiance dans le cobot :	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

B.5 - CAD

	1 : Pas du tout d'accord 7 : Tout à fait d'accord
J'ai trouvé que le cobot n'était pas fiable :	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

B.6 - SAGAT

Questions	Réponses			
Le taux de charge était-il supérieur à celui du cobot ?	Oui	Fort probable	Peu probable	Non

B.7 - ISA

Question :	Réponses				
Comment vous sentez-vous ?	Je m'ennuie	Je suis relaxé	Je me sens bien	J'ai fait un effort	Je suis surchargé

Annexe C : Consentement libre et éclairé Seanatic 1

Formulaire d'information et de consentement libre et éclairé

Vous êtes invité(e) à participer à un projet de recherche. Le présent document vous renseigne sur les modalités de ce projet de recherche. S'il y a des mots ou des paragraphes que vous ne comprenez pas, n'hésitez pas à poser des questions. Pour participer à ce projet de recherche, vous devrez signer le consentement à la fin de ce document et nous vous en remettrons une copie signée et datée.

Titre du projet : SEANATIC

Sous la direction de : Rauffet Philippe et Guérin Clément

Promoteur : Université Bretagne Sud, Laboratoire Lab-STICC, Rue de Saint-Maudé, 56100 Lorient

Investigateur principal : Loïck Simon

Objectif du projet : Comprendre les mécanismes de relation lors d'une coopération humain-machine dans le cadre de la replanification d'une opération de maintenance.

Raison et nature du projet : Votre participation sera requise pour utiliser un pro-giciel sur ordinateur et répondre aux propositions d'une intelligence artificielle. Ces propositions seront ponctuées de questionnaires, il n'y a pas de bonne ou de mauvaises réponses dans ces questionnaires. La passation durera au maximum 1h15. Cette passation se déroulera au sein de l'ENSIBS, 17 Boulevard Flandres Dunkerque 1940, 56100 Lorient ou à distance.

Avantages pouvant découler de la participation : Votre participation à ce projet de recherche contribuera à l'avancement des connaissances portant sur la coopération humain-machine.

Inconvénients et risques pouvant découler de la participation : Votre participation à la recherche ne devrait pas comporter d'inconvénients significatifs, si ce n'est le fait de donner de votre temps.

Droit de retrait sans préjudice de la participation : Il est entendu que votre participation à ce projet de recherche est tout à fait volontaire et que vous restez libre, à tout moment, de mettre fin à votre participation sans avoir à motiver votre décision ni à subir de préjudice de quelque nature que ce soit.

Confidentialité, partage, surveillance et publications : Durant votre participation à ce projet de recherche, le chercheur responsable ainsi que son personnel recueilleront et consigneront dans un dossier de recherche les renseignements vous concernant. Seuls les renseignements nécessaires à la bonne conduite du projet de recherche seront recueillis. Ils peuvent comprendre les informations suivantes : nom, sexe, date de naissance, origine ethnique, photographies, enregistrements vidéo ou audio, habitudes de vie, résultats de tous les tests, examens et procédures que vous aurez à passer lors de ce projet, etc.

Tous les renseignements recueillis au cours du projet de recherche demeureront strictement confidentiels dans les limites prévues par la loi. Afin de préserver votre identité et la confidentialité de ces renseignements, vous ne serez identifié(e) que par un numéro de code. La clé du code reliant votre nom à votre dossier de recherche sera conservée par le chercheur responsable du projet de recherche.

Le chercheur principal de l'étude utilisera les données à des fins de recherche dans le but de répondre aux objectifs scientifiques du projet de recherche décrits dans ce formulaire d'information et de consentement.

Les données du projet de recherche pourront être publiées dans des revues scientifiques ou partagées avec d'autres personnes lors de discussions scientifiques. Aucune publication ou communication scientifique ne renfermera d'information permettant de vous identifier. Dans le cas contraire, votre permission vous sera demandée au préalable.

Les données recueillies seront conservées, pour une période n'excédant pas 5 ans. Après cette période, les données seront détruites.

Enregistrement vidéo : Durant cette passation un enregistrement vidéo de l'écran sera réalisé. Vous n'apparaissez pas dans les enregistrements vidéo.

Résultats de la recherche et publication : Si vous le souhaitez, vous serez informé des résultats de la recherche et des publications qui en découleront, le cas échéant. À tout moment, vous pouvez solliciter le chercheur à ce sujet, il mettra en œuvre les moyens nécessaires pour vous répondre. Nous préserverons l'anonymat des personnes ayant participé à l'étude.

Consentement libre et éclairé :

Je, _____ (nom en caractères d'imprimerie), déclare avoir lu et/ou compris le présent formulaire. Je comprends la nature et le motif de ma participation au projet. J'ai eu l'occasion de poser des questions auxquelles on a répondu, à ma satisfaction.

Par la présente, j'accepte librement de participer au projet.

Signature du ou de la participant.e : _____

Fait à _____, le _____ 2022

Déclaration de responsabilité des chercheurs de l'étude

Je, Loïck Simon, chercheur principal de l'étude, déclare que les chercheurs collaborateurs, ainsi que mon équipe de recherche, sommes responsables du déroulement du présent projet de recherche. Nous nous engageons à respecter les obligations énoncées dans ce document et également à vous informer de tout élément qui serait susceptible de modifier la nature de votre consentement.

Signature : Loïck Simon

Je, Loïck Simon, certifie avoir expliqué à la participante ou au participant intéressé(e) les termes du présent formulaire, avoir répondu aux questions qu'il ou qu'elle m'a posées à cet égard et lui avoir clairement indiqué qu'il ou qu'elle reste, à tout moment, libre de mettre un terme à sa participation au projet de recherche décrit ci-dessus. Je m'engage à garantir le respect des objectifs de l'étude et à respecter la confidentialité.

Signature : Loïck Simon

Fait à Lorient, le _____ 2022.

Annexe D : Échelles de confiance dispositionnelle

D.1 - ATI (Franke et al., 2019)

	Répondez en fonction de votre accord avec ces affirmations sur une échelle de 1 à 6. 1 = « pas du tout d'accord » et 6 = « tout à fait d'accord »
J'aime m'intéresser en détail aux systèmes technologiques	○ ○ ○ ○ ○ ○

J'aime tester les fonctionnalités des nouveaux systèmes technologiques	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Je suis principalement en interaction avec des systèmes technologiques car j'y suis obligé	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Lorsque j'ai un nouveau système technologique en face de moi, je l'essaie de façon approfondie	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
J'aime passer du temps à me familiariser avec un nouveau système technologique	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
C'est suffisant pour moi lorsque le système technologique fonctionne ; Je ne me préoccupe pas du comment ou du pourquoi	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
J'essaie de comprendre comment un système technologique fonctionne exactement	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Connaître les fonctionnalités de base d'un système technologique est suffisant pour moi.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
J'essaie d'utiliser pleinement les capacités d'un système technologique	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

D.2 - PTT (Jessup, 2018)

	Répondez en fonction de votre accord avec ces affirmations sur une échelle de 1 à 5. 1 = « pas du tout d'accord », et 5 = « tout à fait d'accord ».
En général, j'ai confiance dans les systèmes technologiques	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

Les systèmes technologiques m'aident à résoudre tout problème.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Je pense que c'est une bonne idée de compter sur l'aide des systèmes technologiques.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Je n'ai pas confiance dans les informations que je reçois des systèmes technologiques.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Les systèmes technologiques sont fiables	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Je peux compter sur les systèmes technologiques	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

D.3 - GRIPS (Zhang et al., 2019)

	Répondez en fonction de votre accord avec ces affirmations sur une échelle de 1 à 7. 1 = « pas du tout d'accord » et 7 = « tout à fait d'accord ».
Prendre des risques rend la vie plus amusante	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Mes amis auraient tendance à dire que je prends des risques	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
J'aime prendre des risques dans la plupart des aspects de ma vie	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Je prendrais des risques même si cela implique que je pourrais être blessé	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

Prendre des risques est partie intégrante de ma vie	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Je prends souvent des décisions risquées	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Je crois dans le fait qu'il faut prendre des risques	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Je suis attiré, plutôt qu'effrayé, par le risque	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

Annexe E : Échelles de confiance apprises

E.1 - IMOTRIS (CAI) (Lyons & Guznov, 2019 ; Mayer et al., 1995)

	Au vu des premières informations que vous avez sur le système technique que vous allez utiliser (avant interaction) / répondez en fonction de votre accord avec ces affirmations sur une échelle (après interaction), répondez aux affirmations suivantes. Répondez en fonction de votre accord avec ces affirmations sur une échelle de 1 à 7. 1 = « pas du tout d'accord », 4 = « ni d'accord, ni pas d'accord » et 7 = « tout à fait d'accord ».
Si cela ne tenait qu'à moi, je ne laisserais pas ce système technologique avoir d'influence sur les aspects importants de la tâche	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Je serais à l'aise de laisser à ce système technologique l'entière responsabilité de la tâche	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
J'aimerais vraiment disposer d'un bon moyen pour surveiller les décisions de ce système technologique	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

Je serais à l'aise de laisser ce système technologique mettre en oeuvre ses décisions, même si je ne peux pas la surveiller	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Je pourrais compter sur ce système technologique sans hésitation	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Je pense qu'utiliser ce système technologique conduira à des résultats positifs	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Je me sentirais à l'aise pour compter sur ce système technologique dans le futur.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Lorsque la tâche était difficile, je sentais que je pouvais dépendre de ce système technologique	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Si je faisais face à une tâche très difficile dans le futur, je voudrais avoir ce système technologique avec moi	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Je serais à l'aise de laisser ce système technologique prendre toute les décisions	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

E.2 - CAD

	Répondez en fonction de votre accord avec ces affirmations sur une échelle de 1 à 7. 1 = « pas du tout d'accord », 4 = « ni d'accord, ni pas d'accord » et 7 = « tout à fait d'accord ».
J'ai confiance dans la proposition du système technologique	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

Annexe F : Échelles de confiance situationnelle

F.1 - Échelle de perception du risque (Wilson et al., 2018)

	Répondez en fonction de votre accord avec ces affirmations sur une échelle de 1 à 5. 1 = « pas du tout d'accord », et 5 = « tout à fait d'accord ».
Cette replanification est risquée	○ ○ ○ ○ ○
Vous vous sentez concerné à propos de cette replanification	○ ○ ○ ○ ○
Vous vous sentez craintif lorsque vous repensez à cette replanification	○ ○ ○ ○ ○
Vous vous sentez anxieux lorsque vous repensez à cette replanification	○ ○ ○ ○ ○
Vous vous sentez inquiet lorsque vous repensez à cette replanification	○ ○ ○ ○ ○
Vous vous sentez concerné par les effets potentiels que la replanification pourraient avoir sur vous personnellement	○ ○ ○ ○ ○
Si je faisais l'expérience d'une replanification, il est probable que cela m'affecterait négativement	○ ○ ○ ○ ○
Si je faisais l'expérience d'une replanification, il est probable que cela aurait un effet sévère sur moi	○ ○ ○ ○ ○

F.2 - NASA TLX (Hart & Staveland, 1988)

	Les questions suivantes servent à mesurer votre
--	---

	charge mentale durant le scénario. Chaque question peut-être répondu de 0 à 20. 0 = Faible et 20 = Fort.
Jusqu'à quel point les activités mentales et perceptives étaient requises pour faire la tâche?	○ ○ ○ ○ ○ [...]○ ○ ○ ○ ○
Jusqu'à quel point les activités physiques étaient requises pour faire la tâche (ex., pousser, tirer, tourner, contrôler, activer, etc.)	○ ○ ○ ○ ○ [...]○ ○ ○ ○ ○
Jusqu'à quel point avez-vous ressenti la pression du temps due au rythme ou à la vitesse à laquelle la tâche ou les éléments de tâche arrivent?	○ ○ ○ ○ ○ [...]○ ○ ○ ○ ○
Jusqu'à quel point pensez-vous que vous réussissez à atteindre les buts de la tâche tels que définis par l'expérimentateur ou par vous-mêmes?	○ ○ ○ ○ ○ [...]○ ○ ○ ○ ○
Jusqu'à quel point avez-vous eu à travailler (mentalement ou physiquement) pour atteindre votre niveau de performance?	○ ○ ○ ○ ○ [...]○ ○ ○ ○ ○
Jusqu'à quel point vous sentiez-vous non confiant, découragé, irrité, stressé et ennuyé plutôt que confiant, avec plaisir, content, relaxe, satisfait de vous durant la tâche?	○ ○ ○ ○ ○ [...]○ ○ ○ ○ ○

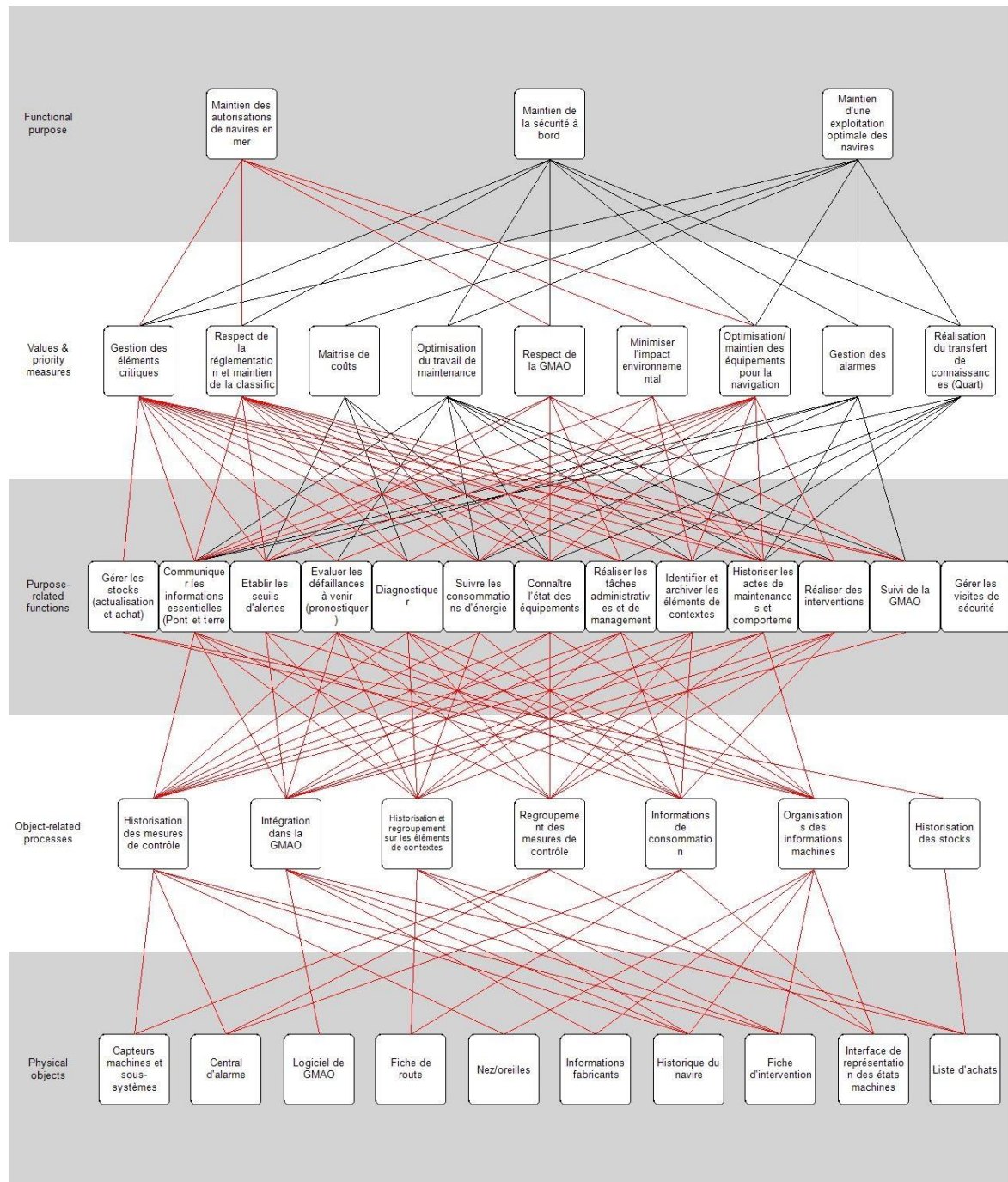
Pour chacune des 15 comparaisons suivantes, choisissez dans le menu déroulant la dimension qui a été la plus importante pour votre charge mentale lors des tâches précédentes.
Effort / Performance
Exigence Temporelle / Frustration
Exigence Temporelle / Effort

Exigence Physique / Frustration
Performance / Frustration
Exigence Physique / Exigence Temporelle
Exigence Physique / Performance
Exigence Temporelle / Exigence Mentale
Frustration / Effort
Performance / Exigence Mentale
Performance / Exigence Temporelle
Exigence Mentale / Effort
Exigence Mentale / Exigence Physique
Effort / Exigence Physique
Frustration / Exigence Mentale

Annexe G : Questions socio-démographique Seanatic 1

Quel âge avez-vous ?	[0 ;∞]
Vous êtes :	[homme, femme, non-genré, ne souhaite pas répondre, autre]
Quelle est votre filière et année d'étude ?	[GC 3, GI3, GI4,GI5,Altergi3,Altergi4,Altergi5,Meca3, Meca4, Meca5, Altermeca3, PeIA-A1, PeIA-A2]
Aviez-vous des connaissances préalables sur les éléments mécaniques d'un bateau ?	[Oui (expert), oui (basique), non]
Avez-vous déjà utilisé un logiciel de gestion de maintenance assistée par ordinateur (GMAO) ?	[Oui, non]

Hiérarchie d'abstraction de supervision et de planification



Annexe I : ConTA et SOCA

Annexe I.1 : CAT de supervision et de planification

Situations / Fonctions	Supervision	Planification
Prendre en compte et diffuser les retours d'expériences		
Ajuster les seuils d'alarmes		
Diagnostiquer		
Suivre les consommations d'énergie		
Connaître l'état des équipements		
Réaliser les tâches administratives et de management		
Prendre en compte les éléments de contextes		
Réaliser et historiser les actes de maintenance et comportements anormaux		
Planifier des interventions		
Suivre la GMAO		
Réajuster le (dérégation au) planning de GMAO		
Gérer les stocks (achat)		
Visualisation des mesures de contrôle		
Visualisation des Informations de consommation énergétique		
Détection des comportements anormaux		
Détection et affichage des alarmes		
Visualisation des informations machines et historique du navire		
Historisation et visualisation des Informations sur les éléments de contextes		
Affichage des données GMAO et intégration d'informations complémentaire		
Actualisation des stocks		
Communication des informations		

Annexe I.2 : CAT de planification (gauche) et de supervision (droite) en fonction des contextes de navigation

Situations Fonctions	Contextes de navigation					Situations Fonctions	Contextes de navigation				
	Mauvais temps	Arrêt technique	Manoeuvre	Eau calme	Quart de nuit		Mauvais temps	Arrêt technique	Manoeuvre	Eau Calme	Quart de nuit
Prendre en compte et diffuser des retours d'expériences						Prendre en compte et diffuser des retours d'expériences					
Ajuster les seuils d'alarmes						Ajuster les seuils d'alarmes					
Evaluer les défaillances à venir (Pronostiquer)						Evaluer les défaillances à venir (Pronostiquer)					
Diagnostiquer						Diagnostiquer					
Suivre les consommations d'énergie						Suivre les consommations d'énergie					
Connaître l'état des équipements						Connaître l'état des équipements					
Réaliser les tâches administratives et de management						Réaliser les tâches administratives et de management					
Prendre en compte les éléments de contextes						Prendre en compte les éléments de contextes					
Réaliser et historiser les actes de maintenance et comportements anormaux						Réaliser et historiser les actes de maintenance et comportements anormaux					
Planifier des interventions						Planifier des interventions					
Suivre la GMAO						Suivre la GMAO					
Réajuster le (dérégation au) planning de GMAO						Réajuster le (dérégation au) planning de GMAO					
Gérer les stocks (achat)						Gérer les stocks (achat)					
Visualisation des mesures de contrôle						Visualisation des mesures de contrôle					
Visualisation des informations de consommation énergétique						Visualisation des informations de consommation énergétique					
Détection des comportements anormaux						Détection des comportements anormaux					
Détection et affichage des alarmes						Détection et affichage des alarmes					
Visualisation des informations machines et historique du navire						Visualisation des informations machines et historique du navire					
Historisation des mesures de contrôle						Historisation des mesures de contrôle					
Historisation et visualisation des informations sur les éléments de contextes						Historisation et visualisation des informations sur les éléments de contextes					
Attichage des données GMAO et intégration d'informations complémentaire						Attichage des données GMAO et intégration d'informations complémentaire					
Actualisation des stocks						Actualisation des stocks					
Communication des informations						Communication des informations					

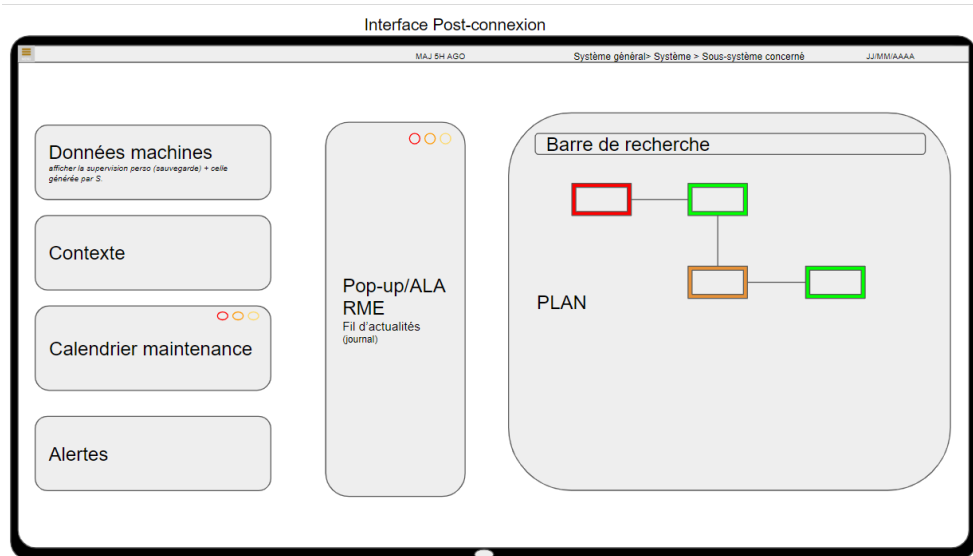
Annexe I.3 : SOCA-CAT des fonctions de supervision et de planification

Situations / Fonctions	Supervision	Planification
Prendre en compte et diffuser des retours d'expériences		
Ajuster les seuils d'alarmes		
Evaluer les défaillances à venir (Pronostiquer)		
Diagnostiquer		
Suivre les consommations d'énergie		
Connaître l'état des équipements		
Réaliser les tâches administratives et de management		
Prendre en compte les éléments de contextes		
Réaliser et historiser les actes de maintenance et comportements anormaux		
Planifier des interventions		
Suivre la GMAO		
Réajuster le (dérogation au) planning de GMAO		
Gérer les stocks (achat)		
Visualisation des mesures de contrôle		
Visualisation des informations de consommation énergétique		
Détection des comportements anormaux		
Détection et affichage des alarmes		
Visualisation des informations machines et historique du navire		
Historisation des mesures de contrôle		
Historisation et visualisation des Informations sur les éléments de contextes		
Attichage des données GMAO et intégration d'informations complémentaire		
Actualisation des stocks		
Communication des informations		

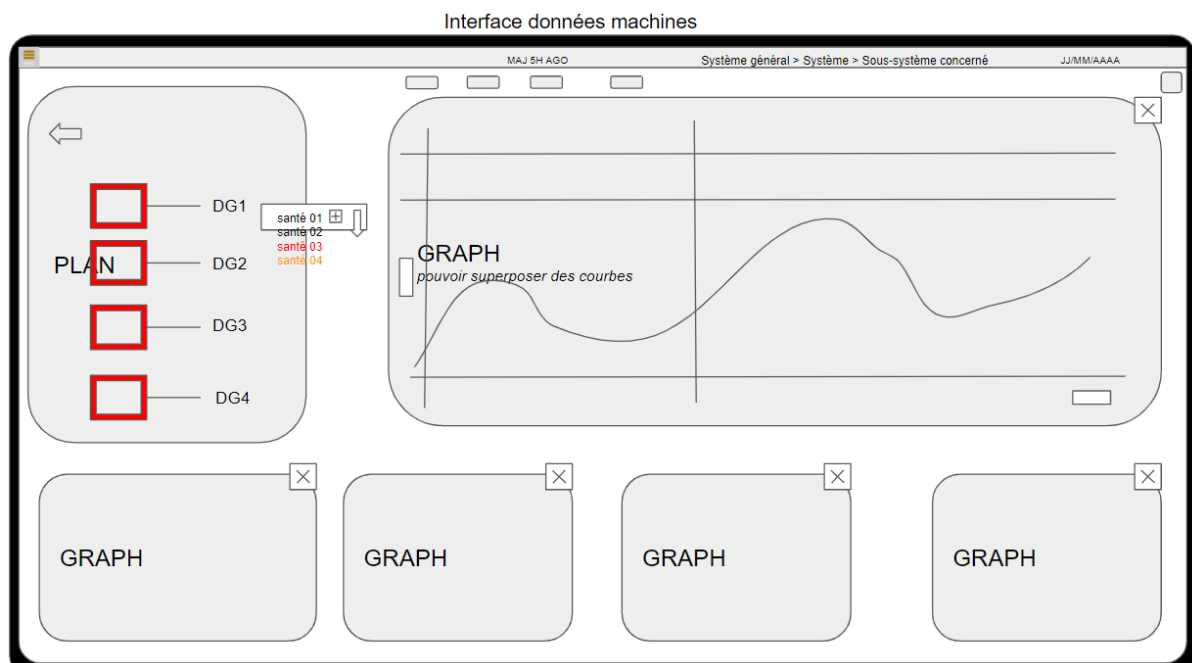
- Chef mécanicien
- Armateur
- Passerelle
- Officier mécanicien
- Technicien de maintenance
- Second mécanicien
- Centrale d'alarme

Annexe J : Atelier co-conception

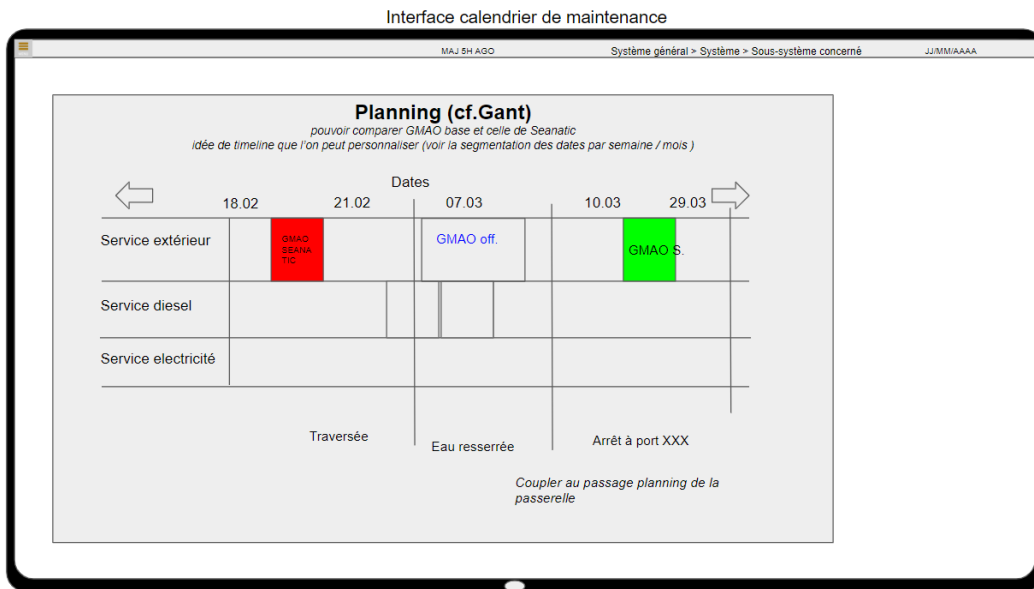
Annexe J.1 : Wireframe de l'interface Post-Connexion



Annexe J.2 : Wireframe de l'interface « données machines »



Annexe J.3 : Wireframe de l'interface « calendrier de maintenance »



Annexe J.4 : Post-its des fonctionnalités du calendrier Senatic

Fonction GMAO +

Un planning - calendrier Ordre chronologique Pondérer en fonction du degré de l'intervention (sur 5 étoiles) Filtrer sur le type d'appareil +0	Pouvoir filtrer Liste des maintenances ayant une date qui varie par rapport à la GMAO +0	Mettre en avant (alerte) les maintenances qui arrivent plus tôt Signaler les éléments critiques (par rapport à la réglementation) +0	Pouvoir consulter les données associées à cette "prédiction" Données des machines Lien avec données +0	Pouvoir entrer du texte / commentaire +0	Pouvoir contredire une prédiction Communiquer un degré d'acceptation à l'IA +0
Préconiser une maintenance plus tôt ou plus tard en fonction du plan de route (port avec des pièces en stock). + Intégration des étapes de maintenance Afficher date GMAO + date Seanatic +0	Ce qui fonctionne aujourd'hui : - Le fait d'avoir accès aux notices - de pouvoir incorporer des pièces jointes(photos, rapports, ...) - de pouvoir avoir accès aux données constructeurs en matière de maintenances qui sont prévues +0	Pouvoir mettre à jour Seanatic en fonction des données provenant de la GMAO +0	Si décision, pouvoir intégrer plusieurs personnes - comptes Capacité de s'identifier, savoir qui a pris la décision +0	Consulter sur l'historique des maintenances et des prédictions Pouvoir revenir/modifier sur une action (en gardant l'historique) +0	Pouvoir qualifier manuellement une maintenance (à qualifier lors de l'import de la GMAO) : - difficulté - nombre de personnes nécessaires - degré d'urgence - + degré d'urgence - temps d'intervention +0

Annexe J.5 : Post-its des fonctionnalités des données machines

Fonction Pré-Alarme +

Il faut avoir accès aux valeurs des seuils critiques (A tels seuils, il faut faire de la maintenance) Montrer les seuils (haut - moyen -bas) +0	Raccourci vers les données machines +0	Créer différents types d'écarts Créer différents types de seuils (valeurs fixes, coefficient de pentes, discontinuité, etc.) pour régler les pré-alarmes +0	Les regrouper par appareil +0	Pouvoir avoir une indication s'il s'agit d'une usure ou d'un défaut (évolution continue ou rupture) Avoir accès aux causes probables et aux tests simples pour pouvoir cibler plus précisément le défaut +0	Pouvoir activer / désactiver une pré-alarme Pouvoir acquitter une alarme +0
Historique des alarmes nombre de fois où l'écart est apparu Pouvoir consulter les derniers déclenchements d'une pré- +0	Définition d'un groupe de pré-alarmes pour déterminer une maintenance ? +0	apparitions synchrones ou pas de différents écarts +0	Permettre à ces pré-alarmes d'être associées à une déconnexion automatique (lors d'une phase de démarrage par exemple) +0	Pré-alarme avec l'objet de la maintenance et une date d'intervention? Positionner pré-alarme par rapport au planning opérationnel (date d'intervention prévue par +0	Popup pour relancer pré-alarme si pas d'action de la part de l'utilisateur +0
Pouvoir prioriser les indices alarmants Classer par ordre d'importance les "niveaux" d'écarts des alarmes Se +0	Au fil de l'eau ou à l'ouverture du logiciel Seanatic (instances en cours?) +0				

Annexe J.6 : Post-its des fonctionnalités de dérives

Fonction Données Machine +

indicateur général de la santé des satellites de l'équipement voir l'état rapidement par une couleur (vert/rouge) + 0	Données températures - Degré Celsius - Courbe si pas affichage instantané + 0	Données NoX PPM, uG/M3, G/KWH Courbes + 0	Données Spectrométriques Mesures d'opacité, %age de lumière absorbé Courbes + 0	Données Débits Courbes + 0	Données Pressions Courbes + 0
Données de Niveaux Courbes + 0	Données Watts Courbes + 0	Mettre en parallèle les courbes réelles, normales et prédites + 0	Schématiser les appareil, les capteurs, leur localisation les uns par rapport aux autres Synoptique général de la machine + 0	Vitesse Courbes + 0	Equipement global vers sous équipement (Diesel Engine => Filtre) Pouvoir choisir ce que l'on souhaite afficher et pouvoir aller dans le détail (micro - macro) + 0
Choisir son mode d'affichage (thèmes sombre ou clair) + 0	Raccourci vers le listing des alarmes Raccourci des seuils d'alertes pour un appareil / capteur + 0	Choisir les échelles d'affichage (temps, valeurs min et max) + 0	Pouvoir afficher dans données de plusieurs appareils / capteurs en même temps. (division de l'écran) Mettre en parallèle plusieurs équipements + 0	Pouvoir choisir parmi plusieurs affichages graphiques (courbe, baton, ...) + 0	Avoir un affichage montrant le lien entre les pages (exemple : à l'extrémité du tuyau du circuit d'air de lancement on peut accéder au compr + 0
Valeurs moyennes + 0	Données Instant T en numérique, et défilement avec variations de valeurs + 0	Déviations entre deux points + 0			

Annexe J.7 : Post-its des fonctionnalités de contexte

Fonction Contexte +

Contexte de transition/changement de régime (démarrage, accélération, décélération, arrêt, etc.) + 0 * Désactiver la collecte des données pour certains contextes	Etat de la mer T°C eau de mer météo en général + 2	Changement d'équipe Contexte de changement/roulement des équipes de maintenance + 0 Disponibilité équipages	Comment passer d'un contexte à un autre ? + 0	tirant d'eau + stab Chargement du navire ----- mouvement du navire + 0	Timeline des différents contextes rencontrés Définir les contextes dans le temps (passés et à venir) + 0
escale appropriées à l'approvisionnement durée d'escale + 0 Feuille de route	Disponibilité d'un stock de pièces disponibilité pièce / délai approvisionnement + 1 outillage disponible				

Annexe K : Consentement libre et éclairé Seanatic 2

Formulaire d'information et de consentement libre et éclairé

Vous êtes invité(e) à participer à un projet de recherche. Le présent document vous renseigne sur les modalités de ce projet de recherche. S'il y a des mots ou des paragraphes que vous ne comprenez pas, n'hésitez pas à poser des questions. Pour participer à ce projet de recherche, vous devrez signer le consentement à la fin de ce document et nous vous en remettrons une copie signée et datée.

Titre du projet : SEANATIC

Sous la direction de : Rauffet Philippe et Guérin Clément

Promoteur : Université Bretagne Sud, Laboratoire Lab-STICC, Rue de Saint-Maudé, 56100 Lorient

Investigateur principal : Loïck Simon

Objectif du projet : Comprendre les mécanismes de relation lors d'une coopération humain-machine dans le cadre de la replanification d'une opération de maintenance.

Raison et nature du projet : Votre participation sera requise pour utiliser un pro-giciel sur ordinateur et répondre aux propositions d'une intelligence artificielle. Avant et après cette passation vous aurez à répondre à des questionnaires. Il n'y a pas de bonne ou de mauvaises réponses dans ces questionnaires. La passation durera au maximum 1h15. Cette passation se déroulera au sein de l'ENSM, 10 Quai Frissard, 76600 Le Havre, France

Avantages pouvant découler de la participation : Votre participation à ce projet de recherche contribuera à l'avancement des connaissances portant sur la coopération humain-machine.

Inconvénients et risques pouvant découler de la participation : Votre participation à la recherche ne devrait pas comporter d'inconvénients significatifs, si ce n'est le fait de donner de votre temps.

Droit de retrait sans préjudice de la participation : Il est entendu que votre participation à ce projet de recherche est tout à fait volontaire et que vous restez libre, à tout moment, de mettre fin à votre participation sans avoir à motiver votre décision ni à subir de préjudice de quelque nature que ce soit.

Confidentialité, partage, surveillance et publications : Durant votre participation à ce projet de recherche, le chercheur responsable ainsi que son personnel recueilleront et consigneront dans un dossier de recherche les renseignements vous concernant. Seuls les renseignements nécessaires à la bonne conduite du projet de recherche seront recueillis. Ils peuvent comprendre les informations suivantes : nom, sexe, date de naissance, origine ethnique, photographies, enregistrements vidéo ou audio, habitudes de vie, résultats de tous les tests, examens et procédures que vous aurez à passer lors de ce projet, etc.

Tous les renseignements recueillis au cours du projet de recherche demeureront strictement confidentiels dans les limites prévues par la loi. Afin de préserver votre identité et la confidentialité de ces renseignements, vous ne serez identifié(e) que par un numéro de code. La clé du code reliant votre nom à votre dossier de recherche sera conservée par le chercheur responsable du projet de recherche.

Le chercheur principal de l'étude utilisera les données à des fins de recherche dans le but de répondre aux objectifs scientifiques du projet de recherche décrits dans ce formulaire d'information et de consentement.

Les données du projet de recherche pourront être publiées dans des revues scientifiques ou partagées avec d'autres personnes lors de discussions scientifiques. Aucune publication ou communication scientifique ne renfermera d'information permettant de vous identifier. Dans le cas contraire, votre permission vous sera demandée au préalable.

Les données recueillies seront conservées, pour une période n'excédant pas 5 ans. Après cette période, les données seront détruites.

Enregistrement vidéo : Durant cette passation un enregistrement vidéo de l'écran sera réalisé. Vous n'apparetez pas dans les enregistrements vidéo.

Résultats de la recherche et publication : Si vous le souhaitez, vous serez informé des résultats de la recherche et des publications qui en découleront, le cas échéant. À tout moment, vous pouvez solliciter le chercheur à ce sujet, il mettra en œuvre les moyens nécessaires pour vous répondre. Nous préserverons l'anonymat des personnes ayant participé à l'étude.

Consentement libre et éclairé :

Je, _____ (nom en caractères d'imprimerie), déclare avoir lu et/ou compris le présent formulaire. Je comprends la nature et le motif de ma participation au projet. J'ai eu l'occasion de poser des questions auxquelles on a répondu, à ma satisfaction.

Par la présente, j'accepte librement de participer au projet.

Signature du ou de la participant.e : _____

Fait à _____, le _____ 2022

Déclaration de responsabilité des chercheurs de l'étude

Je, Loïck Simon, chercheur principal de l'étude, déclare que les chercheurs collaborateurs ainsi que mon équipe de recherche sommes responsables du déroulement du présent projet de recherche. Nous nous engageons à respecter les obligations énoncées dans ce document et également à vous informer de tout élément qui serait susceptible de modifier la nature de votre consentement.

Signature : Loïck Simon

Je, Loïck Simon, certifie avoir expliqué à la participante ou au participant intéressé(e) les termes du présent formulaire, avoir répondu aux questions qu'il ou qu'elle m'a posées à cet égard et lui avoir clairement indiqué qu'il ou qu'elle reste, à tout moment, libre de mettre un terme à sa participation au projet de recherche décrit ci-dessus. Je m'engage à garantir le respect des objectifs de l'étude et à respecter la confidentialité.

Signature : Loïck Simon

Fait au Havre, le _____ 2022.

Annexe L : Questions socio-démographique Seanatic 2

Quel âge avez-vous ?	[0 ;∞]
Quelle est votre identité de genre ?	[homme, femme, non-genré, ne souhaite pas répondre, autre]
En quelle année d'étude êtes-vous ?	[Master 1, Master 2]
Combien de mois avez-vous navigué ?	[0 ;∞]
Avez-vous déjà utilisé une GMAO ?	[Oui, Non]
Quel est votre niveau de connaissances sur les équipements de type « Générateur Diesel »	[Niveau élevé de connaissance, Niveau moyen de connaissance, Niveau faible de connaissance]
Quelle est votre appétence envers le métier de chef.fe mécanicien ?	[Je souhaite devenir chef.fe mécanicien, Je ne souhaite pas devenir chef.fe mécanicien, Je ne sais pas encore]