



HAL
open science

Agentivité dans les systèmes fortement automatisés

Kévin Le Goff

► **To cite this version:**

Kévin Le Goff. Agentivité dans les systèmes fortement automatisés. Modélisation et simulation. AIX-MARSEILLE UNIVERSITE, 2016. Français. NNT: . tel-01455102

HAL Id: tel-01455102

<https://hal.science/tel-01455102v1>

Submitted on 3 Feb 2017

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.

AIX-MARSEILLE UNIVERSITE

ED 62 - Sciences de la Vie et de la Santé

Laboratoire de Psychologie Cognitive - UMR 7290

ONERA - Commande des Systèmes et Dynamique du Vol

Thèse présentée pour obtenir le grade universitaire de docteur

Discipline : Neurosciences

Kevin Le Goff

Agentivité dans les systèmes fortement automatisés

Soutenue le 13.12.2016 devant le jury :

Dr. Elisabeth Pacherie	ENS - Institut Jean-Nicod & CNRS	Rapporteur
Pr. Patrice Terrier	Université de Toulouse - Jean Jaurès	Rapporteur
Dr. Franck Mars	Ecole Centrale de Nantes & CNRS	Examineur
Dr. Valérian Chambon	ENS - Institut Jean-Nicod & CNRS	Examineur
Dr. Bruno Berberian	ONERA	Co-Directeur de thèse
Dr. Arnaud Rey	Aix-Marseille Université & CNRS	Co-Directeur de thèse



Cette oeuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 3.0 France.

Remerciements

Mes premiers remerciements s'adressent à mes deux directeurs de thèse, Arnaud Rey et Bruno Berberian, pour leurs conseils et pour leur soutien. Merci à Arnaud pour m'avoir accompagné depuis la Licence et permis de me lancer sur de nombreux projets ainsi que de m'avoir poussé à partir m'aguerrir à l'étranger. Merci Bruno pour m'avoir permis de passer si facilement du M2 à la Thèse et de la Thèse au premier job! Merci également pour cet encadrement à toute épreuve et ta disponibilité 24/24 et 7/7!

Merci aux membres du jury, Dr. Elisabeth Pacherie, Pr. Patrice Terrier, Dr. Franck Mars et Dr. Valérian Chambon, d'avoir accepté d'évaluer ce travail.

Je remercie vivement toute l'équipe de l'ONERA : les collègues doctorants, les collègues de l'administration et les collègues chercheurs pour m'avoir toujours permis de réaliser ces travaux dans les meilleures conditions possibles. Le seul problème, c'est que vous avez fait passer cette thèse beaucoup trop rapidement!

Je remercie également les collègues de Marseille pour m'avoir formé de la Licence à la Thèse et pour m'avoir donné le goût de la recherche! Mention spéciale pour Olivier avec qui nous avons pu vérifier l'existence du phénomène d'Empreinte de Konrad Lorenz. Merci à tous les amis qui ont été présents au quotidien et qui ont rendu agréable cette vie étudiante! Merci à Fred et Sylvain pour m'avoir inspiré et encouragé depuis le début de ma formation!

Merci à Elodie et Christophe pour ce soutien mutuel qui dure, pour tous les souvenirs accumulés avec notre trio infernal et tous ces repas de la Trinité!

Merci infiniment à mes parents, ma petite sœur et toute ma famille pour m'avoir toujours soutenu dans les bons et les moins bons moments depuis le début de cette aventure marseillaise!

Et pour finir, ces derniers mots sont pour toi Alaïs, pour m'avoir encouragé, soutenu, aidé et supporté au quotidien pendant ces trois ans de Thèse!

Table des matières

Liste des Figures	vi
Liste des Tableaux	vi
Liste des abréviations	viii
Introduction	3
I Automatisation	3
I.1 Contexte actuel	3
I.2 Aspects positifs de l'automatisation	6
I.3 Le phénomène de sortie de boucle	9
I.4 Comprendre l'impact de l'automatisation	12
I.5 Compenser les effets de la sortie de boucle	20
II Agentivité	26
II.1 Définition	26
II.2 Composante rétrospective de l'agentivité	28
II.3 Composante prospective de l'agentivité	32
II.4 Intégration multiple	34
II.5 Problème HCI-Agentivité	36
III Problématique	39

IV Matériel et Méthodes	42
IV.1 Dispositif expérimental	42
IV.2 Mesures	43
Contributions Expérimentales	53
Première Etude	53
Seconde Etude	84
Troisième Etude	103
Discussion Générale	119
Contexte et Discussion des résultats	119
Limites de notre approche	131
Perspectives	135
Conclusion	140
Bibliographie	162
Annexes	164

Table des figures

II.1	Dispositif expérimental de l'étude de Wegner et Wheatley (1999).	30
II.2	Dispositif expérimental de l'étude de Wegner, Sparrow et Winerman (2004).	31
IV.1	Exemple d'essai sur l'interface du LIPS	43
V.1	Dispositif expérimental avec l'affichage de la navigation et l'interface de pilotage	61
V.2	Séquence d'événements pour chaque essai dans chacune des quatre conditions	63
V.3	Evaluation de l'estimation temporelle pour chaque condition	66
V.4	Evaluation du sentiment de contrôle dans la reprise en main du système	67
V.5	Description du feedback visuel lorsque le système prend une mauvaise décision	72
V.6	Représentation de la cible et du curseur à ajuster	73
V.7	Gauche : Temps de réaction pour la détection d'erreur ; Centre : Nombre moyen de fausses alarmes ; Droite : Evaluation du d'	75
V.8	Gauche : Temps passé sur la tâche de tracking ; Droite : Nombre d'allers-retours entre les deux tâches	76
VI.1	Représentation du dispositif expérimental	91
VI.2	Affichage de l'écran de navigation	92
VI.3	Représentation de la cible et du curseur à ajuster	95
VI.4	Gauche : Temps de réaction pour la détection d'erreur ; Droite : Temps passé sur la tâche de tracking	97

VII.1	Affichage de l'écran de navigation	110
VII.2	Description du feedback visuel lorsque le système prend une mauvaise décision	111
VII.3	Representation of the tasks for the Multi-Attribute Task Battery II.	112
VII.4	Temps de réaction pour la détection d'erreur en fonction des quatre délais	114
VII.5	Valeurs moyennes du sentiment de contrôle et du l'acceptabilité	115

Liste des tableaux

I.1	Echelle d'automatisation d'après Endsley (1999)	4
I.2	Echelle d'automatisation d'après Parasuraman, Sheridan et Wickens (2000)	5

Liste des abréviations

ANOVA	Analyse of Variance
ASRS	Aviation Safety Reporting System database
BEA	Bureau d'Enquêtes et d'Analyses
CIAIAC	Comisión de Investigación de Accidentes e Incidentes de Aviación Civil
DSB	Dutch Safety Board
FAA	Federal Aviation Administration
FMS	Flight Management System
GPWS	Ground Proximity Warning System
HCI	Human-Computer Interaction
IB	Intentional Binding
ICAO	International Civil Aviation Organization
LIPS	Laboratoire d'études des Interactions Pilote-Systèmes
MATB	Multi-Attribute Task Battery
NASA	National Aeronautics and Space Administration
NCE	Negative Compatibility Effect
NTSB	National Transportation Safety Board
ONERA	Office national d'Etudes et de Recherches Aérospatiales
OOTL	Out Of The Loop performance problem
PCE	Positive Compatibility Effect
TCAS	Traffic alert and Collision Avoidance System
TMS	Transcranial Magnetic Stimulation

"On sent son avion, comme un cavalier sent son cheval! On fait corps avec lui et c'est avec ses jambes qu'on doit piloter! J'ai horreur de tous ces instruments de contrôle!

Mon SPAD, moi, je le sens avec mes fesses!"

Georges Guynemer

Introduction

Chapitre I

Automatisation

I.1 Contexte actuel

Au cours des 50 dernières années, les mutations technologiques à l'œuvre ont profondément changé l'interaction de l'humain avec son environnement. L'influence de l'automatisation est aujourd'hui perçue dans tous les aspects de la vie quotidienne et non plus seulement dans le monde de l'industrie. Que ce soit au travail ou à domicile, nous sommes habitués à interagir avec des systèmes informatiques sophistiqués conçus pour nous assister dans nos activités. Cette automatisation a permis de rendre certains aspects de la vie plus *faciles* : en permettant aux personnes en situation de handicap d'être en mesure de se déplacer et de communiquer ; plus *rapides* : avec la généralisation de dispositifs informatisés et l'augmentation de la productivité ; et plus *sûrs* : le taux d'accidents dans l'aviation et l'industrie a baissé grâce à la mise en œuvre de ces systèmes (Norman, 1990).

Comme pour toute notion complexe, il existe autant de définitions de l'automatisation que de livres sur le sujet. Dans son sens le plus large, le terme d'automatisation est défini comme (1) la mécanisation et l'intégration de mesures de variables environnementales, (2) le traitement de données et la prise de décision, (3) l'action mécanique (par moteurs ou

appareils agissant sur l'environnement) et (4) la communication d'informations traitées à l'opérateur (Sheridan & Parasuraman, 2005). Une définition plus succincte établit que l'automatisation correspond à la prise en charge totale ou partielle par la technologie, d'une fonction gérée précédemment par l'humain (Parasuraman & Riley, 1997).

Automatiser une tâche n'est pas un concept binaire (tout ou rien) mais plutôt un continuum avec différents niveaux d'implication de l'automatisme pour chaque fonction nécessaire à la tâche (Billings, 1991). De très nombreuses taxonomies ont été proposées afin de rendre compte de ces différents niveaux (Endsley, 1999 ; Parasuraman, Sheridan, & Wickens, 2000 ; Sheridan & Verplank, 1978). Les tableaux 1 et 2 présentent deux exemples de taxonomies développées respectivement par Endsley (1999) et Parasuraman et al. (2000). Ces taxonomies ont pour point commun leur niveau de base qui correspond à une action complètement réalisée par l'opérateur et leur niveau maximal d'automatisation pour lequel l'ensemble des opérations est géré par la machine (exemple : le guichetier à la banque / le distributeur automatique de billets).

Endsley (1999)
Automatisation complète
Supervision
Décision automatisée
Système rigide
Décision partagée
Aide à la décision
Contrôle partagé
Implémentation automatisée
Aide à l'action
Contrôle manuel

TABLE I.1 – Echelle d'automatisation d'après Endsley (1999)

Parasuraman, Sheridan et Wickens (2000)
Le système décide de tout et agit de façon autonome, ignorant l'opérateur
Le système informe l'opérateur seulement s'il décide de le faire
Le système informe l'opérateur seulement s'il le demande
Le système agit de façon autonome et informe l'opérateur
Le système laisse à l'opérateur un veto avant l'implémentation de l'action
Le système implémente sa décision si l'opérateur l'approuve
Le système propose une action possible
Le système propose quelques alternatives
Le système offre un set complet d'actions et de décisions à l'opérateur
L'opérateur doit prendre toutes les décisions sans l'aide du système

TABLE I.2 – Echelle d'automatisation d'après Parasuraman, Sheridan et Wickens (2000)

L'automatisation est largement présente dans l'industrie (chimique, pétrolière, nucléaire, etc.), les transports (avions, métros, voitures, etc.) et dans nos maisons (ordinateurs, domotique, etc.). Malgré son caractère ubiquitaire, nous nous concentrerons à travers cette thèse sur sa présence dans le monde aéronautique et nous utiliserons également des études issues du monde de l'automobile.

La question de l'automatisation s'est posée très tôt dans l'histoire de l'aviation. Les pilotes ont rapidement reconnu l'instabilité de leur appareil et leur difficulté à gérer l'ensemble des paramètres de vol (Billings, 1991). Le développement d'aides au pilotage était donc une nécessité. Dès 1913, un système de stabilisation automatique a permis à Orville Wright, l'un des pionniers de l'aviation, de piloter quelques instants sans toucher aux commandes de vol (Billings, 1991). Les premiers pilotes automatiques quant à eux, sont apparus à la fin de la Grande Guerre. Grâce à une combinaison de gyroscopes, ils étaient capables de maintenir l'assiette longitudinale et l'inclinaison latérale de l'appareil en envoyant aux gouvernes des ordres par le biais des câbles de commandes (Carpentier, 2011). Ensuite, tout s'est rapidement enchaîné, chaque amélioration technique donnant naissance à un système automatisé. Par exemple, l'implémentation du train d'atterrissage rétractable a nécessité la mise en place d'une alarme pour avertir le pilote en cas d'oubli

de sortie du train.

A la suite de la Seconde Guerre Mondiale, les progrès technologiques ont entraîné une diminution des coûts, de taille, de poids et de consommation d'énergie des équipements informatisés, ce qui a largement facilité leur propagation à tous les niveaux du processus de pilotage (commandes de vol électriques, commandes de gestion de vol, systèmes de localisation et de communication, aides à l'atterrissage, etc.). C'est à cette période que l'optimisation du pilote automatique a permis d'accroître son activité, jusqu'à aboutir à la possibilité d'une prise en charge de la totalité d'un vol. Ainsi en 1947, l'armée américaine a réussi un vol transatlantique (décollage et atterrissage compris) à l'aide du pilote automatique (Carr, 2014). L'implémentation d'un tel pilote automatique dans l'aviation civile a pris davantage de temps mais de nos jours, l'ensemble des avions de lignes long-courriers ou gros porteurs modernes sont équipés de pilotes automatiques et de systèmes automatisés permettant de gérer l'ensemble du vol.

I.2 Aspects positifs de l'automatisation

I.2.1 Sécurité

Wiener et Curry (1980) rappellent que les raisons motivant l'automatisation des systèmes à grande échelle étaient de fournir une alternative plus sûre et moins coûteuse que l'humain. L'Organisation International de l'Aviation Civile (ICAO) reporte régulièrement dans ses rapports annuels une augmentation du trafic aérien d'environ 5% chaque année (Ky & Miaillier, 2006). Cette augmentation constante nécessite une coordination parfaite de tous les acteurs impliqués (pilotes, contrôleurs aériens, techniciens et ingénieurs, etc.) pour maximiser la sécurité des passagers. Du côté de l'appareil, un avion obtient sa certification lorsque la probabilité d'occurrence d'une catastrophe est estimée à 10^{-9} par heure de vol (Rushby, 2011). Le challenge pour le secteur aéronautique a donc depuis toujours été les erreurs humaines, souvent citées comme la cause principale des

incidents/accidents. Jusque dans les années 80, plus de la moitié des accidents aériens était due à ce type d'erreurs (Amalberti, 2013 ; Helmreich & Foushee, 1993 ; Wiener & Curry, 1980). Cette prédominance de l'erreur humaine ne se retrouve pas seulement dans le monde de l'aéronautique. Des études sur les risques liés à des opérations chirurgicales reportent également un taux d'erreurs de 75% (Chopra, Bovill, Spierdijk, & Koornneef, 1992 ; Cooper, Newbower, & Kitz, 1984 ; Wright, Mackenzie, Buchan, Cairns, & Price, 1991).

Nous verrons dans le chapitre suivant que nous n'approuvons pas forcément ce terme d'erreur "humaine" qui donne l'impression que l'opérateur ne sait tout simplement pas se servir correctement de systèmes automatisés parfaitement fiables (Woods, Dekker, Cook, Johannesen, & Sarter, 2012). Toutefois, établir que les opérateurs humains n'étaient pas aussi fiables que l'automatisation a motivé l'industrie aéronautique à automatiser toujours plus leurs appareils (Billings, 1996 ; Wiener, 1988). Il est maintenant clair que l'introduction de systèmes automatisés a permis de diminuer drastiquement le taux d'accidents malgré une augmentation constante du trafic aérien (Norman, 1990).

L'implémentation du système d'avertissement de proximité du sol (*Ground Proximity Warning System*, GPWS) illustre parfaitement l'apport de l'automatisation en matière de sécurité. Avant son implémentation, le nombre de collisions au sol alors que l'avion était sous contrôle des pilotes restait relativement élevé (Wiener & Curry, 1980). La surcharge de travail des pilotes les distrait du monitoring de l'altitude de leur appareil. Le phénomène de désorientation spatiale, correspondant à la perception erronée du pilote de la position, du mouvement ou de l'attitude de son avion, est également pointé du doigt comme cause de ces collisions (Mars, 2001). Depuis 1975, un système avertit les pilotes si l'avion s'approche anormalement d'un relief ou du sol (hors phase d'atterrissage) réduisant ainsi considérablement le nombre de ces accidents (Billings, 1991).

Un autre exemple est fourni par le système d'alerte de trafic et d'évitement de collision (*Traffic alert and Collision Avoidance System*, TCAS), une autre avancée majeure pour

la sécurité aérienne. Cet instrument de bord a pour but d'éviter les risques de collisions en plein air qui augmentaient avec l'accroissement du trafic aérien. Dans un premier temps, le TCAS va avertir les pilotes de la présence d'un autre avion dans le même secteur de vol. Si le risque de conflit augmente, le TCAS va alors conseiller les pilotes sur une trajectoire d'évitement à prendre (Bliss & Acton, 2003 ; Mohleji, Lamiano, & Massimini, 2009).

En plus des systèmes de détection d'obstacles ou de reliefs, le système de gestion de vol (*Flight Management System*, FMS) a permis d'optimiser l'ensemble des trajectoires du décollage à l'atterrissage, de détecter des erreurs ou pannes des systèmes et de faciliter le vol de l'avion. Cette optimisation a joué un rôle dans la sécurité et le confort des passagers et de l'équipage en permettant de mieux anticiper les zones de turbulences ou de mauvaises conditions climatiques (Amalberti, 1993). S'il est établi que l'automatisation représente la solution idéale pour gérer des fonctions de façon plus sûre que les opérateurs humains, il faut également la considérer comme un outil permettant une réduction des coûts pour les deux plus importants postes de dépenses d'une compagnie aérienne : la masse salariale et le carburant.

I.2.2 Economie

Les progrès techniques ont rendu possible le développement d'avions de plus en plus gros et de plus en plus rapides. Le nombre d'instruments de communication, de localisation et de capteurs n'a cessé d'augmenter. Dans les années 50, le cockpit d'un avion long-courrier abritait cinq personnes. Petit à petit, la prise en charge de certaines fonctions par des systèmes automatisés et notamment par le pilote automatique a permis de diminuer la charge de travail de l'équipage. Malgré la réticence des syndicats de pilotes, cette baisse s'est accompagnée d'une diminution du nombre de personnes nécessaires pour le vol d'un long courrier en passant à deux pilotes dans un cockpit tout en maintenant un haut niveau de sécurité (Galan, 2012). Les deux pilotes, assistés par l'automatisation, ont

pu prendre en charge à la fois les tâches de communication puis les tâches de navigation et la gestion des moteurs et du carburant, rendant obsolète la présence, respectivement, de l'opérateur radio, du navigateur et du mécanicien navigant (Billings, 1991 ; Norman, 1990). Le passage de cinq à deux personnes grâce à l'automatisation a ainsi permis une forte réduction de la masse salariale pour les compagnies aériennes.

Ensuite, le contexte économique des années 70 avec le premier choc pétrolier a rapidement motivé les constructeurs et compagnies aériennes à trouver des solutions pour réduire la consommation de carburant. Les outils que nous avons cités précédemment ont également permis d'atteindre cet objectif. Le système de gestion de vol a permis de réduire le temps de vol, de voler par mauvais temps avec une visibilité nulle. Le TCAS a permis de réduire la distance minimale entre deux avions tout en maintenant un niveau de sécurité élevé au cœur d'un trafic aérien toujours plus dense. Cette réduction de la distance a permis d'optimiser les trajectoires de descente et de montée à proximité des aéroports, ce qui a aussi permis d'économiser du carburant et d'augmenter les rotations des vols. Ainsi, malgré le coût non négligeable nécessaire à l'introduction des systèmes automatisés (achat, implémentation, entraînement, maintenance, etc.), l'automatisation a eu un impact économique positif pour les compagnies aériennes (Wiener & Curry, 1980).

I.3 Le phénomène de sortie de boucle

Nous avons vu précédemment que les progrès technologiques ont permis de donner aux systèmes automatisés la capacité de prendre des décisions et de les implémenter. A travers cette thèse, nous avons choisi de nous concentrer sur la supervision (niveau 9 d'automatisation d'après la nomenclature d'Endsley du Tableau 1). A ce niveau d'automatisation, l'ensemble des actions est pris en charge par l'automatisation sous la surveillance de l'opérateur. Les tâches du superviseur correspondent donc à : la planification des ac-

tions; l'implémentation de ces actions dans le système automatisé; le monitoring des actions réalisées par le système, la reprise en main si nécessaire et l'analyse des résultats (Sheridan & Parasuraman, 2005). Un tel niveau d'automatisation a permis d'allonger la distance des vols tout en les rendant plus sûrs et moins coûteux. Toutefois, ces bénéfices tendent à masquer un problème fondamental : l'interposition de systèmes automatisés entre les opérateurs et les différentes tâches à effectuer a profondément modifié l'activité de l'opérateur humain et ses modalités d'interaction avec l'environnement, souvent d'une façon inattendue et non anticipée par les concepteurs (Dekker & Woods, 2002). En effet, de nombreux travaux soulignent les conséquences négatives en matière de sécurité et de performance de l'approche traditionnelle de l'automatisation donnant naissance au phénomène de sortie de boucle de contrôle (*Out Of the Loop performance problem*, OOTL Endsley & Kiris, 1995 ; Kaber & Endsley, 1997).

Un opérateur hors de la boucle de contrôle a des difficultés pour détecter les erreurs ou pannes du système (*detection failures*, Ephrath & Young, 1981 ; Kessel & Wickens, 1982), pour comprendre son état courant (*automation surprises*, Christoffersen & Woods, 2002 ; Sarter & Woods, 1995 ; Sarter, Woods, & Billings, 1997) et pour déterminer les actions appropriées pour la suite de la tâche (Endsley, 1999). Les trois exemples suivants issus de l'aviation, de l'industrie nucléaire et du secteur boursier illustrent le phénomène de sortie de boucle et son omniprésence dans tous les secteurs fortement automatisés.

I.3.1 Exemple 1 : Aviation

Le premier exemple concerne le vol 447 d'Air France. Le 31 mai 2009, l'Airbus 330 décolle de Rio de Janeiro pour Paris. Quatre heures après le décollage, les sondes Pitot gèlent à cause des mauvaises conditions climatiques. Les indications de vitesse de l'appareil étaient donc erronées entraînant la déconnexion du pilote automatique et la reprise en main par le pilote. Pendant les quatre minutes qui ont suivi, les trois pilotes présents dans le cockpit ont été incapables de diagnostiquer la situation et d'appliquer la

procédure appropriée. L'alarme incompréhensible pour les pilotes, couplée à la situation hautement stressante et à son caractère inattendu ont certainement empêché les pilotes d'agir correctement et de reprendre en main l'appareil (BEA, 2012).

I.3.2 Exemple 2 : Industrie nucléaire

Le deuxième exemple concerne l'incident de la centrale nucléaire de Three Miles Island (Pennsylvanie, USA) en 1979. Une valve servant à réguler l'arrivée d'eau au cœur du réacteur était bloquée en position ouverte, malgré le signal lumineux sur l'interface de contrôle indiquant que la valve était en position fermée. En fait, ce signal lumineux n'indiquait pas la position réelle de la valve mais seulement que l'ordre de fermeture avait été envoyé. A cause de ces informations ambiguës, les opérateurs ont été incapables de diagnostiquer le problème correctement (Norman, 2013). Durant cette période, une succession de pannes et d'actions inappropriées a conduit à une fonte partielle du cœur du réacteur. Heureusement, grâce à la redondance des barrières de protection, les rejets radioactifs n'ont pas été assez importants pour causer des dommages sanitaires et environnementaux. Une catastrophe nucléaire majeure a été évitée.

I.3.3 Exemple 3 : Bourse

Dans un domaine complètement différent cette fois, nous pouvons citer l'un des bugs informatiques les plus coûteux de l'histoire. Le courtier américain Knight Capital est une firme spécialisée dans le trading haute fréquence (technique basée sur des algorithmes puissants pouvant envoyer de manière automatisée des ordres d'achat ou de vente en une fraction de seconde). Le 1er août 2012, la firme teste une nouvelle version de son algorithme de trading. Cependant à cause d'un bug, le système automatisé a commencé à acheter et vendre des actions de façon totalement erratique entraînant un mini-krach boursier. Les superviseurs n'étaient pas au courant du comportement de leur système, cela a donc pris presque une heure pour comprendre que le problème venait de leur

logiciel. Ce bug d'une heure a coûté à Knight Capital environ 400 millions de dollars (Jones, 2013).

Bien que ces trois exemples proviennent de différents domaines, ils montrent que lorsque les systèmes automatisés tombent en panne, les superviseurs peuvent se retrouver impuissants pour diagnostiquer la situation et déterminer les actions à réaliser car ils ne sont pas au courant de l'état du système avant la panne : ils sont hors de la boucle de contrôle (Billings, 1991). Nous allons voir dans la partie suivante que le phénomène de sortie de boucle provient à la fois de changements quantitatifs et qualitatifs dans le travail des opérateurs.

I.4 Comprendre l'impact de l'automatisation

I.4.1 Approche quantitative : un opérateur moins investi

D'un point de vue chronologique, l'impact négatif de l'automatisation a été dans un premier temps expliqué en des termes quantitatifs. A partir des années 70, la présence de l'automatisation dans le secteur aéronautique a explosé. A cette époque, la question de savoir quelle fonction automatiser avait une réponse très simple : il fallait automatiser toutes les fonctions possibles (Billings, 1996). La tâche des concepteurs était donc de créer des systèmes automatisés efficaces répondant à un cahier des charges précis. Leurs seules limites étaient la faisabilité technologique et le coût (Dekker & Woods, 2002 ; Parasuraman & Riley, 1997). Avec l'automatisation massive des systèmes aéronautiques, l'opérateur s'est vu décharger d'une multitude d'activités dont il avait initialement la responsabilité. Si cette automatisation a permis de diminuer la charge de travail des opérateurs, elle s'est également accompagnée d'une dégradation tant des mécanismes attentionnels que des habiletés motrices qui peuvent expliquer, du moins en partie, les difficultés inhérentes au phénomène de sortie de boucle (Endsley & Kiris, 1995 ; Kaber & Endsley, 1997).

Baisse de Vigilance

Lorsque l'avion atteint son altitude de croisière, le pilote enclenche le pilote automatique qui se charge de maintenir la trajectoire de l'avion, sa vitesse et son altitude. Le rôle du pilote n'est donc plus de faire voler son appareil mais plutôt de superviser le pilote automatique et de gérer les communications avec les contrôleurs aériens. Ce faible investissement dans la tâche en cours s'est notamment traduit par une dégradation des mécanismes de vigilance. Mosier, Skitka, Heers, et Burdick (1998) ont analysé une base de données de la NASA regroupant des incidents volontairement et anonymement reportés par les pilotes (*Aviation Safety Reporting System database, ASRS*). Ils ont trouvé que 77% des incidents liés à des facteurs humains étaient dus à ces problèmes de vigilance.

Comment expliquer une telle dégradation des mécanismes de vigilance ? L'humain peut facilement être attentif à des stimuli prédictibles, fréquents et pendant de courtes périodes. A l'inverse, de nombreuses études sur la vigilance soulignent l'impact du temps sur la capacité à discriminer correctement un signal imprévisible et rare dans du bruit (Mackworth, 1948 ; Parasuraman, 1979 ; Teichner, 1974). Avec l'automatisation des cockpits, les pilotes se retrouvent dans une configuration similaire : aucune interaction avec le système pendant de longue période et une faible probabilité d'occurrence d'une erreur ou d'une panne (Parasuraman & Davies, 1977 ; See, Howe, Warm, & Dember, 1995 ; Wiener & Curry, 1980). Cette baisse de vigilance se traduit notamment par une diminution dans la fréquence de monitoring et de vérifications des systèmes par l'opérateur (Billings, 1991 ; Kaber & Endsley, 1997). En parallèle (le lien de causalité restant à déterminer), l'absence d'implication dans la tâche se traduit par l'avènement de pensées non reliées à la tâche, phénomène plus connue sous le terme de divagation attentionnelle (ou *mind-wandering*, Casner & Schooler, 2015), exacerbant ainsi la perte de vigilance vis-à-vis de la tâche en cours.

Le crash du vol 3407 de la Colgan Air, le 12 février 2009, illustre parfaitement comment des problèmes de vigilance peuvent faire sortir les pilotes de la boucle de contrôle.

L'avion était en phase d'approche pour atterrir à Buffalo (New York, USA) mais les pilotes étaient en retard sur la vérification de la checklist d'atterrissage. Lors de la vérification, aucun des deux pilotes ne monitorait la vitesse de l'appareil qui ralentissait à cause du givre se formant sur les ailes. Lorsque l'alarme de décrochage a retenti, les deux pilotes ont été surpris et n'ont pas su réagir correctement. L'enquête révéla que les deux pilotes étaient très fatigués pendant le vol et qu'ils avaient pendant la phase d'approche une discussion personnelle non appropriée pendant cette phase alors que seules les conversations ayant un lien avec la procédure d'atterrissage sont normalement autorisées (NTSB, 2010).

Perte des capacités manuelles

En plus des problèmes liés à la vigilance, le faible investissement de l'opérateur dans la tâche en cours se traduit par une dégradation des capacités manuelles (Bainbridge, 1983 ; Endsley & Kiris, 1995 ; Norman, 1990), dégradation souvent présentée comme un élément central dans le phénomène de sortie de boucle notamment au regard des difficultés rencontrées par les opérateurs humains à reprendre les systèmes en main en cas d'urgence. En effet, les actions manuelles ont progressivement été remplacées par des tâches de monitoring. Ainsi par manque de pratique, les pilotes tendent à perdre leur capacité à piloter manuellement leur avion. Cette dégradation des capacités manuelles est régulièrement citée comme une des causes qui a conduit au crash du vol AF447 Rio-Paris en 2009 (Arnold, 2015 ; BEA, 2012). Pourtant, les pilotes eux-mêmes sont conscients de cette dégradation et certains désengagent volontairement le pilote automatique de temps en temps pour maintenir leurs capacités à un niveau satisfaisant (Wiener & Curry, 1980).

Gillen (2010) a réalisé une étude avec 30 pilotes expérimentés. En simulateur, les pilotes devaient réaliser cinq manœuvres dites "de base" (décollage, voler avec un moteur en panne, se maintenir en croisière, atterrissage guidé par instruments au sol, approche manquée avec remise de gaz) uniquement avec des indicateurs classiques et sans

système automatisé. Les résultats ont montré que le score moyen pour chacune des manœuvres était en dessous du minimum requis par la *Federal Aviation Administration* (FAA), l'agence américaine responsable des réglementations aériennes. Pour cause, l'investissement de l'opérateur dans ces différentes manœuvres s'est vu clairement réduit au cours des dernières décennies. Il est d'ailleurs intéressant de constater que le score le plus faible concernait le maintien de cap qui se fait quasiment tout le temps avec le pilote automatique à l'heure actuelle, alors que des manœuvres telles que le décollage et l'approche manquée restent moins impactés (même s'ils étaient sous les standards de la FAA) dû à un niveau d'automatisation généralement plus faible.

Le manque de contrôle manuel a également un impact sur la capacité du pilote à détecter le comportement déviant de son appareil. Piloter en manuel permet au pilote de "(res)sentir" le comportement de son appareil à travers la dynamique de ses mouvements (vibrations, roulis, etc.). Cette capacité à ressentir le comportement de son appareil tend à s'estomper par manque de pratique rendant un monitoring déjà complexe, encore plus difficile (Geiselman, Johnson, & Buck, 2013).

I.4.2 Approche qualitative : un opérateur investi différemment

L'automatisation des systèmes a été classiquement (est encore ?) considérée comme une simple substitution d'une activité humaine par la machine, raccourci nommé le mythe de la substitution (*substitution myth*, Christoffersen & Woods, 2002 ; Dekker & Woods, 2002). L'idée véhiculée par un tel raccourci est celle d'un statu quo qualitatif vis-à-vis de l'activité de l'opérateur humain. Le seul changement au regard de l'activité humaine serait d'ordre quantitatif, l'automatisation déléstant l'opérateur d'une partie du travail à effectuer sans fondamentalement modifier la nature de son activité (Christoffersen & Woods, 2002). Nous avons vu dans la partie précédente que les changements quantitatifs pouvaient effectivement favoriser le phénomène de sortie de boucle de contrôle. Toutefois, de nombreux travaux montrent que la sortie de boucle est également causée par un

changement dans la nature du travail de l'opérateur (Parasuraman, Molloy, & Hilburn, 1993 ; Parasuraman & Wickens, 2008 ; Woods & Dekker, 2000). En effet, à de nombreux égards, l'automatisation peut être perçue comme l'introduction d'un nouvel agent dans la boucle de contrôle avec lequel le pilote devra interagir. Autrement dit, l'automatisation des systèmes peut être assimilée à l'ajout d'un nouveau "membre d'équipage", impliquant de nouvelles demandes en termes de coordination pour l'opérateur humain (Sarter et al., 1997). Ce changement qualitatif, largement mésestimé par les concepteurs, s'est traduit par de nombreuses difficultés liées à la compréhension de ces nouveaux partenaires, à la confiance qui leur est accordée ainsi qu'à leur acceptabilité. Chacune de ces difficultés participent à leur manière à la dégradation des performances qui accompagne le phénomène de sortie de boucle.

Opacité des systèmes

Pour interagir efficacement avec les automatismes, l'agent humain se doit de comprendre et d'anticiper les actions de ce nouveau compagnon. Pour ce faire, il s'appuie non seulement sur la connaissance acquise par rapport au système en question mais également par les feedbacks renvoyés par celui-ci. Or, il est clair que la complexité de l'automatisation est liée à la large palette de fonctions disponibles et à l'absence de feedbacks appropriés les rendant difficilement compréhensibles par les opérateurs (Norman, 1990). Cette opacité des systèmes est aujourd'hui reconnue comme un facteur clé dans les problèmes de supervision (Christoffersen & Woods, 2002 ; Dekker & Woods, 2002 ; Klein, Woods, Bradshaw, Hoffman, & Feltovich, 2004 ; Norman, 1990).

Alors que la baisse de vigilance entraîne des temps de détections d'erreurs plus longs, l'opacité du système va jouer sur le temps nécessaire à l'opérateur pour se mettre à jour sur l'état courant du système, comprendre la cause du problème, stabiliser le système et en reprendre le contrôle. De nombreux exemples d'incidents/accidents illustrent les difficultés pour les pilotes à comprendre le comportement du système automatique (Billings,

1996 ; Funk et al., 1999 ; Sarter & Woods, 1995). C'est cette difficulté qui est à l'origine des *automation surprises*, situations au cours desquelles les opérateurs sont surpris du comportement de l'automatisation qui n'agit pas de manière attendue (Christoffersen & Woods, 2002 ; Sarter & Woods, 1995 ; Sarter et al., 1997). Cette incompréhension pousse les pilotes à se demander ce que le système est en train de faire, pourquoi il le fait et ce qu'il va faire ensuite (Wiener, 1988).

Le crash du vol Turkish Airlines 1951 Istanbul-Amsterdam en 2009 (DSB, 2010) est un exemple de l'opacité du pilote automatique et de la difficulté pour les pilotes à suivre son comportement. Lors de la phase d'approche, un des radioaltimètres est tombé en panne et indiquait que l'avion était au sol. Les pilotes ont pris connaissance de cette panne mais ils ignoraient que cette panne allait faire passer le pilote automatique en mode "arrondi" c'est-à-dire le mode enclenché normalement lorsque l'avion touche la piste. Le changement de mode est pourtant indiqué mais n'est pas suffisamment saillant pour être détecté lors d'une phase pendant laquelle les pilotes vérifient leur checklist. Le pilote automatique a fait perdre de la vitesse à l'avion jusqu'à l'envoyer au sol. Trois pilotes étaient présents dans le cockpit ce jour-là et aucun n'a vu ni compris le changement de mode de l'automatisation. Malgré leur implication dans les procédures d'atterrissage, les trois pilotes étaient hors de la boucle de contrôle effective de leur appareil.

Confiance

Il est depuis longtemps établi que la confiance est un facteur déterminant pour l'utilisation des systèmes automatisés (Verberne, Ham, & Midden, 2015). Ainsi, pour des interactions homme-machine optimales, le niveau de confiance des opérateurs vis-à-vis de l'automatisation se doit d'être adapté. Toutefois, la fiabilité très élevée de l'automatisation favorise la confiance excessive des opérateurs (Sheridan & Parasuraman, 2005). Cette confiance excessive se traduit par une absence de remise en question par l'opérateur des actions ou décisions de l'automatisation et une attention réduite sur les tâches de mo-

monitoring qui se surajoute aux problèmes de vigilance (Bahner, Hüper, & Manzey, 2008 ; Metzger & Parasuraman, 2005 ; Wickens, Clegg, Vieane, & Sebok, 2015). Les exemples d'accidents liés à une confiance excessive des opérateurs sont nombreux (Endsley, 1996 ; Mosier et al., 1998).

Bien que les problèmes liés à la confiance des opérateurs vis-à-vis de leurs systèmes automatisés sont difficiles à évaluer car ils dépendent souvent de facteurs individuels (Lee & See, 2004), Parasuraman, Molloy, et Singh (1993) ont montré que cette confiance excessive pouvait être diminuée lorsque le niveau de fiabilité de l'automatisation était variable au cours du temps par rapport à un niveau de fiabilité constant. Le nombre d'erreurs du système était le même dans les deux conditions mais les opérateurs détectaient davantage les erreurs dans le cas d'un niveau de fiabilité variable. Dans la même étude, les auteurs ont montré que cet effet de confiance excessive n'était pas présent lorsque les opérateurs n'avaient qu'une seule tâche à monitorer, ce qui valide l'hypothèse selon laquelle l'allocation des ressources attentionnelles joue également un rôle important dans le phénomène de confiance excessive.

Il peut exister également des cas pour lesquels le problème n'est pas une confiance excessive vis-à-vis du système mais plutôt une confiance insuffisante. Nous avons parlé précédemment de l'introduction du système d'avertissement de proximité du sol (GWPS) au milieu des années 70. Malgré son impact positif sur la diminution du nombre d'accidents par collision avec le sol, le trop grand nombre de fausses alarmes qu'il générerait au début a diminué la confiance des pilotes envers ce système. Ainsi, les pilotes préféraient parfois éteindre le système, couper l'alarme ou tout simplement ne pas en tenir compte (Wiener & Curry, 1980).

Problème d'acceptabilité

Le phénomène de sortie de boucle de contrôle n'a pas qu'un rôle sur la performance globale de pilotage. Un autre problème qui n'a pas été envisagé par les concepteurs des

systèmes autonomes est celui de l'acceptabilité. Même si les performances et l'acceptabilité sont souvent corrélées positivement (Nielsen & Levy, 1994), une amélioration de la performance n'entraîne pas toujours une acceptabilité élevée. En effet, l'évolution du rôle du pilote vers celui de superviseur s'est accompagnée parfois de certaines réticences vis-à-vis de l'automatisation. Certains pilotes reportent se sentir relégués au second plan et se posent la question de l'identité de celui qui est réellement en contrôle de l'appareil : le système ou le pilote (Baron, 1988) ? Le problème d'acceptabilité a été accentué par l'opacité des systèmes et leur manque de communication, donnant aux pilotes l'impression que le système "agissait dans leur dos" (Sarter & Woods, 1995).

Préserver le sentiment de contrôle des utilisateurs fait partie des principes à respecter pour la conception d'interfaces, comme l'ont souligné Shneiderman et Plaisant (2004) avec leur septième principe de conception : *Maintenir un lieu de contrôle interne : les opérateurs expérimentés désirent fortement se sentir en contrôle de l'interface et que l'interface réponde à leurs actions*. Les actions inattendues de l'interface, la difficulté ou l'impossibilité d'obtenir les informations nécessaires et l'incapacité à produire l'action désirée sont tous des facteurs qui entraînent l'anxiété et la non satisfaction de l'opérateur. Gaines (1981) avait déjà fait part de cette idée avec ses encouragements à faire de l'utilisateur l'initiateur des actions plutôt que d'en être le spectateur.

La question de l'acceptabilité ne doit pas être reléguée au second plan car un système qui n'est pas acceptable par les pilotes ne sera pas utilisé malgré la plus-value au niveau de la sécurité ou de la performance (voir le cas de GPWS). Inagaki et Nagai (2007) ont fait le même constat dans le domaine de l'automobile en montrant que les conducteurs préféraient être assistés par une alarme indiquant la présence d'obstacles plutôt que de laisser le système automatisé, qui est mal compris par les conducteurs, gérer la conduite malgré une performance améliorée.

L'accident du vol 006 de la China Airlines, en 1985, démontre parfaitement les problèmes liés à une sortie de boucle de contrôle. A la suite de la perte d'un moteur, le pilote

automatique a essayé de compenser au maximum le déséquilibre lié à la perte de puissance. Or compenser un tel déséquilibre nécessite le gouvernail de direction mais le pilote automatique ne pouvait contrôler que les ailerons, ce que le pilote ignorait (connaissance insuffisante du système). Les pilotes étaient davantage focalisés sur le redémarrage du moteur en panne que sur le monitoring du système (confiance excessive envers le système). Lorsque le déséquilibre était trop important et que le pilote automatique ne pouvait plus le compenser, il s'est désengagé (comme il a été programmé pour le faire) pour laisser le contrôle au pilote. Le problème était que le pilote a dû reprendre en main un avion quasiment à la verticale ce qu'il n'avait absolument pas anticipé (*automation surprise*). Heureusement, après une chute de plusieurs milliers de mètres, le pilote a réussi à rétablir l'assiette de l'avion et à reprendre l'appareil en main (Norman, 1990 ; NTSB, 1986).

I.5 Compenser les effets de la sortie de boucle

Afin d'éviter le catastrophisme, il est important de rappeler que le taux d'erreurs ou de pannes des systèmes fortement automatisés est (exceptionnellement) faible. De plus, la plupart des accidents sont dus à des causes multiples. Il est rare que la simple défaillance d'un système ou de l'humain soit à l'origine d'une catastrophe. C'est plutôt un concours de circonstances qui entraîne un accident (voir la théorie du *Swiss Cheese*, Reason, 1990, 2000). Les rapports d'accidents du vol Air France 447 Rio-Paris en 2009 (BEA, 2012), du vol Turkish Airlines 1951 Istanbul-Amsterdam en 2009 (DSB, 2010) et du vol SpanAir 5022 Madrid-Las Palmas en 2008 (CIAIAC, 2011) suffisent à nous convaincre du caractère multifactoriel des crashes et incidents aériens. Il est donc souvent difficile de déterminer la part de responsabilité des automatismes ou de l'humain (pilotes, contrôleurs aériens, opérateurs de maintenance, etc). C'est cette part variable dans la responsabilité des accidents qui pousse la communauté des facteurs humains à préférer le terme d'erreurs d'interactions plutôt que celui d'erreurs humaines.

Nous avons décrit le phénomène de sortie de boucle de contrôle à travers les changements quantitatifs et qualitatifs induits par l'automatisation. Nous allons à présent décrire les différentes solutions proposées afin de compenser ce phénomène, en suivant un découpage identique.

I.5.1 Approche quantitative : investir davantage l'opérateur

Pour compenser le changement lié à la diminution de l'investissement des opérateurs, la nécessité d'améliorer les entraînements est régulièrement citée lors des rapports du NTSB ou du BEA notamment pour améliorer les connaissances du fonctionnement des automatismes et le maintien des capacités à piloter manuellement. Toutefois, créer un entraînement qui permette de limiter le phénomène de sortie de boucle n'est pas une chose aisée. Parmi les différentes méthodes, nous pouvons citer l'entraînement aux aléas, qui consiste à exposer l'opérateur humain à des situations nécessitant une reprise en main ce qui permet notamment de diminuer le niveau de complaisance de l'opérateur vis-à-vis de l'automatisme (Bahner et al., 2008 ; Plat & Amalberti, 2000). Mais l'entraînement en simulateur présente également des limites. En effet, il est impossible de prévoir toutes les pannes ou situations possibles et de les implémenter lors des phases d'entraînement des opérateurs. Il est donc important d'entraîner les opérateurs à développer des stratégies générales et non des réponses spécifiques (Bainbridge, 1983). En plus des entraînements, un des moyens envisagés pour limiter le phénomène de sortie de boucle de contrôle a consisté à déterminer un meilleur partage des tâches pour maintenir le niveau d'engagement des opérateurs.

Afin de réinvestir les opérateurs dans leurs tâches, différentes méthodes d'allocation de fonctions ont été proposées. Les concepteurs n'avaient plus pour consigne d'automatiser l'ensemble des fonctions mais de trouver le juste équilibre pour maintenir l'opérateur dans la boucle de contrôle. Déterminer quelle fonction automatiser ou non est une difficulté récurrente pour les concepteurs et les ergonomes. L'approche MABA-MABA (*Men-*

Are-Better-At/Machines-Are-Better-At) a constitué une des méthodes les plus utilisées (Dekker & Woods, 2002). Cette méthode consiste à définir l'allocation des différentes fonctions à réaliser entre les opérateurs et les systèmes automatisés en fonction des forces et faiblesses de chacun (Fitts, 1951). Elle permet ainsi de manipuler le niveau d'automatisation des systèmes et de partager les tâches à effectuer entre l'automatisme et l'opérateur (Parasuraman, Molloy, & Hilburn, 1993 ; Parasuraman et al., 2000). L'opérateur reste tout de même responsable de l'ensemble des actions, y compris celles réalisées par le système (Lee & Seppelt, 2009).

L'allocation statique de fonctions a progressivement laissé place à l'allocation dynamique. Le partage des tâches n'est plus figé mais évolue en fonction des différentes phases. Par exemple, l'automatisation peut fournir une aide à l'opérateur uniquement lorsque sa charge de travail est élevée (Kaber & Endsley, 2004). Pour éviter la confiance excessive, l'opérateur peut reprendre temporairement la main pour le maintenir dans la boucle de contrôle et maintenir ses capacités manuelles (Parasuraman, Molloy, & Hilburn, 1993).

Qu'elle soit statique ou dynamique, cette approche repose sur l'idée que le partage des tâches aura pour résultat de garder l'opérateur impliqué dans les actions du système, permettant ainsi d'améliorer les performances et la conscience de situation de cet opérateur par rapport à une simple supervision d'un système fortement automatisé (Endsley, 1987 ; Endsley & Kiris, 1995). Depuis, des données empiriques sont venues confirmer cette hypothèse (Endsley, 1999 ; Kaber & Endsley, 2004 ; Kaber, Onal, & Endsley, 2000).

Ces méthodes permettent sans conteste de limiter l'impact du phénomène de sortie de boucle. Toutefois, de nombreux chercheurs soulignent la nécessité de dépasser cette simple approche quantitative du problème posée par l'automatisation des systèmes (Dekker & Woods, 2002). En effet, un des inconvénients d'une telle approche est le découpage arbitraire de chacune des actions à effectuer, sans prise en compte de leur interdépendance. Il arrive donc que les opérateurs doivent gérer une somme d'actions incohérentes entre elles (Bainbridge, 1983).

I.5.2 Approche qualitative : améliorer les relations hommes-systèmes

Nous avons vu que l'automatisation changeait la nature du travail de l'opérateur (passage du contrôle manuel à la supervision). Pour compenser les problèmes d'opacité, de confiance et d'acceptabilité, il est nécessaire d'identifier comment l'automatisation et les opérateurs peuvent se compléter et se coordonner pour réaliser l'ensemble des fonctions de façon optimale.

Dès 1990, Norman souligna que le problème de sortie de boucle de contrôle n'était pas dû à l'automatisation en elle-même mais plutôt au manque de feedbacks appropriés. Or, les feedbacks sont au cœur du contrôle (Lee & Seppelt, 2009). Avant l'ère de l'automatisation, les opérateurs étaient activement engagés dans l'ensemble des tâches à effectuer. Ils devaient tout monitorer et tout contrôler entraînant ainsi des surcharges de travail. Toutefois, cela permettait à l'opérateur d'être continuellement informé de l'état courant du système. Depuis, l'opacité des systèmes tend à isoler les opérateurs des actions à effectuer, ce qui est un des facteurs majeurs de la sortie de boucle de contrôle (Christoffersen & Woods, 2002 ; Dekker & Woods, 2002 ; Klein et al., 2004 ; Norman, 1990). Pour contrer cette opacité, Norman (1990) explique la nécessité de fournir à l'opérateur des feedbacks et des interactions en continu de la même façon que lorsque deux individus agissent ensemble.

Le point crucial pour les concepteurs n'est plus de se concentrer uniquement sur les performances propres de l'automatisation (Lee & Seppelt, 2009) mais surtout d'arriver à réaliser des interfaces permettant une coordination efficace entre l'opérateur et l'automatisme, pas seulement lors de situations standards et facilement prédictibles mais également lors de circonstances imprévues. Comme un bon équipier, les systèmes automatisés doivent faire en sorte que leurs actions soient facilement compréhensibles par le reste de l'équipe à travers des feedbacks et le partage d'intentions. Ainsi, de nombreuses études se sont penchées sur la question de comment transformer l'automatisme en agent collaboratif (Christoffersen & Woods, 2002 ; Dragan, Lee, & Srinivasa, 2013 ; Hoc, 2000,

2001 ; Klein et al., 2004 ; Zimmermann, Bauer, Lutteken, Rothkirch, & Bengler, 2014). Nous allons voir trois exemples d'études issues du monde de l'automobile et proposant différentes approches pour concevoir des systèmes plus coopératifs.

La première étude s'est concentrée sur la nécessité de partager avec l'automatisation des buts communs. Verberne, Ham, et Midden (2012) ont réalisé une expérience lors de laquelle ils ont demandé à 57 participants de classer les quatre critères suivants : confort, rapidité, sécurité et consommation d'énergie, selon l'importance qu'ils leur accordaient. Les participants devaient ensuite évaluer différents systèmes d'aide à la conduite selon une échelle d'acceptabilité et de confiance. Conformément à leur hypothèse, les systèmes partageant le même ordre de priorité étaient évalués plus acceptables et plus fiables par les participants. Ainsi pour être acceptable, les systèmes automatisés doivent partager leurs buts avec les opérateurs, ce qui confirme que les personnes interagissent avec la technologie de la même façon qu'ils interagissent avec d'autres personnes (Lee & See, 2004 ; Nass & Moon, 2000).

La seconde étude s'est intéressée à la nécessité d'anthropomorphiser les systèmes automatisés. Cette étude a également été réalisée par Verberne et al. (2015). Elle est basée sur le fait que les individus tendent à faire davantage confiance aux personnes ayant un visage partageant des traits similaires (DeBruine, 2002). L'expérience se déroule dans un simulateur de conduite automatisée. Le système automatisé était représenté par un agent virtuel qui partageait ou non des traits faciaux avec les participants. Comme pour le partage des buts, les résultats ont montré que les participants faisaient davantage confiance à l'agent virtuel leur ressemblant.

La dernière étude a montré l'intérêt d'utiliser des messages amorces permettant de diminuer l'opacité du système. (Koo et al., 2015) ont démontré que les bénéfices des messages amorces dépendaient de la nature et de la quantité des informations affichées. Dans un simulateur de conduite, 64 participants ont réalisé une tâche de conduite semi-automatisée. La voiture était équipée d'une fonction de freinage automatisé permettant

de freiner en cas d'obstacles (par exemple, un piéton, un débris) sur la trajectoire de la voiture sans intervention du conducteur. Lorsqu'un freinage d'urgence avait lieu, un message vocal était joué fournissant une explication sur l'action future de la voiture. Les résultats ont montré que le message expliquant l'action de la voiture (*how : the car is braking*) et la raison de cette action (*why : obstacle ahead*) en même temps, augmente les performances de conduite mais diminue l'acceptabilité du système. Avec le développement des voitures autonomes, de nombreuses études se sont penchées sur l'utilisation de ces amorces pour optimiser les interactions entre un système à haut niveau d'automatisation et un opérateur humain (Gold, Damböck, Lorenz, & Bengler, 2013 ; Navarro, Mars, Forzy, El-Jaafari, & Hoc, 2010).

Ces exemples d'études empiriques nous montrent qu'il existe un certain nombre de propositions qui ont été faites pour compenser l'effet de l'automatisation sur l'opérateur (voir également Nowak et Biocca (2003), sur la téléprésence ; Eriksson et Stanton (2015), sur les règles de communication). Toutes ces études ont en commun la nécessité de concevoir des systèmes automatisés ressemblant de plus en plus à l'humain sur le plan physique et/ou cognitif.

A travers cette thèse, nous allons poursuivre les efforts déjà entrepris pour améliorer la coordination entre un opérateur et un système artificiel. Pour cela nous souhaitons nous concentrer sur la compréhension des transformations induites par l'automatisation à travers les avancées récentes relatives aux mécanismes sous-tendant le contrôle et la compréhension de nos propres actions. En particulier, nous défendons l'idée que l'opérateur interprète les intentions et le résultat des actions produits par un système à travers son propre "outillage cognitif" de la même manière qu'il le fait pour les actions d'autrui. Dès lors, la compréhension de cet "outillage" pourrait permettre de proposer des principes de conception permettant de concevoir des systèmes plus coopératifs, et donc plus contrôlables.

Chapitre II

Agentivité

II.1 Définition

Lorsque nous agissons, nous avons le sentiment que nous contrôlons nos propres actions et que nous pouvons ainsi agir sur l'environnement extérieur. Ce sentiment est décrit comme le "sentiment d'agentivité" (*sense of agency*, Gallagher, 2000, 2012 ; Haggard & Tsakiris, 2009). Ce sentiment nous permet de discriminer naturellement les actions dont nous sommes les auteurs de celles dont nous ne le sommes pas. De nombreuses études en psychologie (Aarts, Custers, & Wegner, 2005 ; Moore, Wegner, & Haggard, 2009), psychopathologie (Franck et al., 2001) et neurosciences (Farrer et al., 2003) se sont penchées sur les mécanismes permettant de s'attribuer ou non une pensée, des intentions ou des actions.

Bien que les mécanismes ne soient pas encore parfaitement compris, les différentes approches proposent que les mécanismes cérébraux de l'agentivité agissent comme un système central de traitements qui intègre une large variété d'indices internes (signaux moteurs, feedbacks sensoriels, etc.) et externes (amorces, contexte, etc.), qui ne sont pas mutuellement exclusifs mais qui sont combinées pour créer la représentation la plus robuste du sentiment d'agentivité (Moore & Fletcher, 2012 ; Synofzik, Vosgerau, & Newen,

2008 ; Synofzik, Vosgerau, & Voss, 2013). Ces différents indices servent à confirmer ou non l'intégrité de la chaîne causale "intention-action-effet", clé de voûte de l'agentivité. En effet, bien que nous exprimons rarement un sentiment intense d'être l'agent d'une action (expérience phénoménologique), nous reconnaissons facilement cette rupture d'agentivité lorsque nos actions ne se déroulent pas comme prévu ou ne produisent pas les effets attendus (Chambon, Sidarus, & Haggard, 2015). L'agentivité serait donc un "mode par défaut" perturbé par des ruptures dans la chaîne causale "intention-action-effet" (Blakemore, Wolpert, & Frith, 2002 ; Sidarus & Haggard, 2016).

Lorsque nous interagissons par le biais de systèmes automatisés, l'ensemble des variables intermédiaires entre l'initiation d'une action et la perception d'un effet peut largement impacter notre sentiment d'agentivité (Limerick, Coyle, & Moore, 2014). En effet, l'automatisation peut perturber, dégrader ou supprimer les différents indices nécessaires à la vérification de l'intégrité de la chaîne "intention-action-effet". L'intérêt de la communauté des facteurs humains pour ce cadre théorique est très récent et peu d'études se sont penchées, à ce jour, sur l'impact que pouvait avoir l'utilisation de systèmes automatisés sur le sentiment d'agentivité des opérateurs et utilisateurs. Pourtant, Berberian, Sarrazin, Le Blaye, et Haggard (2012) ont récemment confirmé de façon empirique l'impact négatif de l'automatisation sur l'agentivité. Cette dégradation de l'agentivité a également été montrée par Coyle, Moore, Kristensson, Fletcher, et Blackwell (2012) dans une tâche de pointage assistée par ordinateur. Les résultats ont montré que, bien que les participants étaient au courant de l'assistance, ils se sentaient quand même agents pour un niveau faible d'assistance. A l'inverse pour des niveaux plus élevés, il y avait une perte du sentiment d'agentivité (voir également Kumar & Srinivasan, 2013).

Afin de mieux comprendre l'impact de l'automatisation sur le sentiment d'agentivité, il nous apparaît nécessaire de décrire la nature des indices qui sont intégrés lors de l'établissement par les différents mécanismes de l'agentivité. Ces indices sont généralement regroupés en composantes prospective et rétrospective selon qu'ils interviennent avant la

réalisation de l'action ou après.

II.2 Composante rétrospective de l'agentivité

Initialement, les études empiriques sur l'agentivité ont mis l'accent sur des indices moteurs directement liés aux commandes motrices et à l'exécution d'un mouvement. Ces études ont suggéré que les modèles computationnels du contrôle moteur pouvaient également être responsables de la naissance du sentiment d'agentivité (Blakemore & Frith, 2003 ; Blakemore, Wolpert, & Frith, 1998). Le comparateur est un mécanisme qui va comparer à la fois les conséquences sensorielles prédites avant l'action par l'envoi d'une copie d'efférence, avec les conséquences sensorielles réelles (Blakemore, Wolpert, & Frith, 2000 ; Blakemore et al., 2002). Si les deux types d'informations correspondent, alors nous percevons les informations sensorielles comme étant des conséquences de notre action (i.e., sentiment d'agentivité). A l'inverse, lorsque les deux types d'informations ne correspondent pas, alors le sentiment d'agentivité est atténué, voire absent en cas de trop grande différence (Moore & Fletcher, 2012).

D'après le cadre théorique du comparateur, le sentiment d'agentivité se produit logiquement après que l'action ait été effectuée et que les conséquences sensorielles (visuelles, motrices, ou proprioceptives) de l'action soient disponibles, d'où l'appartenance à la composante rétrospective. Les résultats empiriques ont validé cette hypothèse et ont montré que des divergences spatiales (Farrer et al., 2003) et temporelles (Farrer, Bouchereau, Jeannerod, & Franck, 2008) entre l'action effectuée et les feedbacks visuels de l'action réduisait le sentiment d'agentivité envers cette action proportionnellement à la divergence introduite.

Une autre approche a été proposée par Wegner et Wheatley (1999) avec la théorie de l'*apparent mental causation* reliant l'agentivité à une intégration rétrospective d'informations. Wegner et Wheatley ont proposé trois principes permettant de "faire naître

un sentiment d'agentivité". Les principes de priorité, de cohérence et d'exclusivité traduisent le fait que lorsque nous avons une pensée (ou intention) qui émerge avant une action, que cette pensée est cohérente avec l'action et que l'action n'a pas d'autres causes alternatives possibles, alors nous ressentons le sentiment d'être l'auteur de cette action. Le sentiment d'agentivité serait donc, en partie, une reconstruction a posteriori de l'action, après vérification du respect des trois principes. Plus les principes seront respectés et plus la personne se sentira agent.

Pour illustrer leur proposition, Wegner et Weatherley ont mis en place une expérience pour démontrer que nous pouvions nous sentir en contrôle d'événements que nous n'avions en réalité pas causés. Un participant et un complice utilisaient simultanément une souris pour contrôler la position du curseur sur l'écran (voir Figure II.1). Sur cet écran était affichée une photographie représentant 50 objets. Les participants entendaient à travers un casque audio le nom d'un objet présent à l'écran. Le complice avait pour consigne (qu'il recevait à travers son casque) de s'arrêter sur cet objet après que le participant ait entendu le nom de l'objet. Les résultats ont montré que plus l'intervalle temporel entre le mot présenté dans le casque et l'arrêt sur cet objet était court et plus le participant reportait une forte volonté d'action pour l'arrêt réalisé, alors même que le participant suivait passivement le mouvement du complice. Ainsi lorsque la situation est ambiguë (principe d'exclusivité caduque), le respect des principes de priorité et de cohérence permet la naissance d'un sentiment de contrôle sur une action produite par autrui.

A la suite des travaux de Wegner et Wheatley (1999) sur *l'apparent mental causation*, Wegner, Sparrow, et Winerman (2004) ont montré que l'effet d'amorçage pouvait être une bonne manière d'influencer ou de simuler les pensées a priori et par conséquent de créer un sentiment d'agentivité même en l'absence de mouvement. Dans leur étude, les participants se tenaient debout, les bras le long du corps en face d'un miroir pendant qu'une autre personne était placée derrière eux, cachée par un rideau, avec les bras tendus de chaque côté du participant et réalisant une série de mouvements (voir Figure II.2).



FIGURE II.1 – Dispositif expérimental de l'étude de Wegner et Wheatley (1999).

Lorsque les participants pouvaient entendre la description du mouvement qui allait être réalisé par l'autre participant, ils reportaient un plus grand sentiment d'agentivité pour ces mouvements. Il est à noter que l'amorçage n'augmentait pas le sentiment de posséder les mains effectuant les mouvements (*sense of ownership*) mais seulement le sentiment de contrôle sur les actions produites, cela montre que le principe d'exclusivité a été compensé par le respect des principes de priorité et de cohérence.

Dans une tâche de production de sons, Sato (2009) a montré avec un dispositif similaire à Wegner et al. (2004) que l'amorce du mouvement effectué par l'expérimentateur permettait de se sentir plus en contrôle que dans la condition sans amorce. Toutefois, pour avoir une pleine expérience de contrôle, il fallait produire le mouvement soi-même. Son étude justifie donc la nécessité d'une combinaison de plusieurs indices afin de maximiser le sentiment d'agentivité.

D'autres indices interviennent également dans la construction de l'agentivité. Par exemple, la valence émotionnelle de l'effet produit est un des facteurs de la construc-



FIGURE II.2 – Dispositif expérimental de l'étude de Wegner, Sparrow et Winerman (2004).

tion de l'agentivité. Dans une tâche de pointage, Wilke, Synofzik, et Lindner (2012) ont montré que la direction de pointage perçue par les participants ne dépendait pas que d'informations liées à l'action en elle-même mais également de la valence affective de l'effet : les participants percevaient leurs mouvements attirés vers les effets à valence positive et éloignés des effets à valence négative (voir également Yoshie & Haggard, 2013). Wen, Yamashita, et Asama (2015) ont également montré que la réussite dans la tâche à réaliser était également un facteur influençant le développement du sentiment d'agentivité.

II.3 Composante prospective de l'agentivité

Les différents travaux présentés précédemment illustrent la manière dont des indices pouvant être à la fois internes (i.e., les signaux moteurs) ou externes (i.e., les amorces) peuvent agir rétrospectivement grâce notamment à un mécanisme de comparaison des retours sensoriels issus de nos actions, pour moduler voire faire naître un sentiment d'agentivité. Des recherches plus récentes ont montré également l'existence d'un mécanisme prospectif intervenant avant l'action. Ce mécanisme se concentre, dans la chaîne "intention-action-effet", sur le lien entre intention et action, c'est-à-dire aux processus se déroulant avant l'action elle-même et avant de recevoir les feedbacks sur les effets (Wenke, Fleming, & Haggard, 2010). Selon cette composante, la sélection entre plusieurs actions possibles pourrait également faciliter la création d'un sentiment d'agentivité. Cela suggère que l'agentivité dépend également de signaux prospectifs naissant des circuits internes liés à la préparation de l'action en plus des signaux rétrospectifs comparant les états prédits et réels de l'environnement.

Contrairement aux études précédentes qui se sont intéressées à l'amorçage de l'effet à produire, Wenke et al. (2010) ont été les premiers à montrer que la facilité à sélectionner parmi des actions alternatives pouvait affecter l'expérience subjective de l'agentivité. Dans cette étude, les participants appuyaient sur une touche à gauche ou à droite en réponse à une flèche-cible. L'appui enclenchait l'apparition d'un cercle coloré. La présentation par amorçage de la flèche avant la flèche-cible entraînait une facilitation de la sélection de l'action pour des amorces et des cibles congruentes dans leur direction. Les résultats montrent que les participants ressentent plus de contrôle sur l'apparition des couleurs suivant les actions compatibles avec les amorces.

Cependant ces résultats sont compatibles avec une autre interprétation. Etant donné que les amorces accélèrent les temps de réaction quand elles sont compatibles (Dehaene et al., 1998), les participants pourraient par conséquent se sentir plus en contrôle dans les

essais compatibles car ils y répondent plus rapidement. Avec cette interprétation alternative, l'agentivité dépendrait d'un monitoring rétrospectif (Corallo, Sackur, Dehaene, & Sigman, 2008 ; Marti, Sackur, Sigman, & Dehaene, 2010) de la performance de l'action et non d'un monitoring prospectif de la fluidité des signaux pré-moteurs. Chambon et Haggard (2012) ont testé ces deux hypothèses alternatives. Le but de leur expérience était donc de dissocier l'apport de la facilité à sélectionner la réponse de l'apport de la performance sur le sentiment de contrôle en utilisant l'effet négatif de compatibilité (*Negative Compatibility Effect*, NCE). De précédentes études ont montré qu'augmenter l'intervalle entre l'amorce masquée et le stimulus cible renverse la polarité de l'effet d'amorçage sur les temps de réaction (Eimer & Schlaghecken, 1998, 2003 ; Lingnau & Vorberg, 2005). Pour des délais très courts (< 70 ms), nous observons le classique "effet positif de compatibilité" (*Positive Compatibility Effect*, PCE) d'amorçage avec des réductions des temps de réaction et des erreurs pour les essais congruents. Cependant, pour des délais plus longs, un "effet négatif de compatibilité" (NCE) apparaît : la congruence entre l'amorce et la cible entraîne maintenant une augmentation du temps de réaction et des erreurs. Leurs résultats ont montré que les jugements de contrôle étaient plus élevés pour les essais congruents que pour les essais incongruents, même lors de l'utilisation d'intervalles correspondant au NCE qui diminuent les performances motrices des essais congruents par rapport aux essais incongruents. Cela suggère que l'effet d'amorçage est effectivement indépendant de l'exécution motrice et naît des étapes pré-motrices.

En plus de la fluidité de l'action, le nombre de choix possibles a également un lien avec le sentiment d'agentivité (Barlas & Obhi, 2013). Dans une tâche classique d'appui-bouton, les participants avaient soit un, trois ou sept boutons de disponibles pour déclencher l'effet (un son). Les auteurs ont utilisé le liage intentionnel, une mesure implicite de l'agentivité (voir section "Matériel et méthodes"), pour montrer que le nombre d'alternatives possibles était positivement corrélé au sentiment d'agentivité. Nous pouvons toutefois faire l'hypothèse que ce résultat doit être dépendant de la tâche à effectuer et

du nombre d'alternatives possibles. En effet, nous pouvons imaginer qu'un trop grand nombre d'options vienne freiner cette fluidité dans la sélection de l'action à effectuer et ainsi diminuer le sentiment d'agentivité de l'individu.

Pour clôturer cette partie sur la composante prospective de l'agentivité, nous pouvons noter que contrairement à Chambon, Wenke, Fleming, Prinz, et Haggard (2013), Synofzik et al. (2013) définissent les mécanismes liés au comparateur comme étant prospectifs. Pour ces auteurs, l'information critique du comparateur est générée par le mécanisme de prédiction des conséquences sensorielles alors que les feedbacks réels ne sont là que pour la comparaison. Toutefois, comme nous l'avons expliqué précédemment, l'enjeu n'est plus de déterminer si l'agentivité dépend de mécanismes prospectifs ou rétrospectifs puisque la complémentarité des deux composantes est maintenant admise.

II.4 Intégration multiple

Les études précédentes confirment la théorie de l'intégration multiple en montrant que le sentiment d'agentivité dépend en partie de la comparaison rétrospective entre les attentes et les effets et en partie des processus prospectifs se produisant avant l'action notamment lors de la sélection de l'action. Synofzik et al. (2013) listent également les connaissances préalables, les croyances, le contexte et les récompenses associées aux actions comme étant tous des facteurs pouvant impacter le sentiment d'agentivité. Les auteurs proposent que l'agentivité serait donc issue d'une intégration optimale d'indices (Synofzik et al., 2013). Cette combinaison d'informations permet de limiter les erreurs qui seraient liées à un signal bruité ou ambigu (Synofzik et al., 2008). Le cerveau intègre constamment les différents indices que nous avons décrit précédemment et les pondère en fonction de leur fiabilité selon le contexte (Synofzik, Vosgerau, & Lindner, 2009 ; Synofzik & Voss, 2010). Généralement, les signaux moteurs liés à la préparation de l'action et les feedbacks proprioceptifs et visuels fournissent l'information la plus rapide et la moins

bruitée concernant l'action d'une personne. Ces informations sont qualifiées de sensori-motrices (Synofzik et al., 2013). Si ces signaux sont ambigus, les informations qualifiées de cognitives (croyance, attente, performance, etc.) auront un poids plus importants pour compenser le bruit des signaux sensorimoteurs. Ainsi, pour l'expérience de Wegner et al. (2004) citée plus haut, les participants n'effectuant pas de mouvements, les signaux moteurs et proprioceptifs sont absents ce qui doit entraîner une augmentation du poids de l'amorce auditive et du feedback visuel pouvant expliquer l'augmentation du sentiment de contrôle ressenti par le sujet pour des mouvements qu'il ne produit pas.

Un autre exemple illustrant cette pondération en fonction du contexte est donné par l'expérience récente de Sidarus, Chambon, et Haggard (2013). Ils ont fait varier factuellement l'amorçage subliminal de l'action à sélectionner et la probabilité d'apparition des effets. Les résultats ont montré que lorsque l'effet était attendu, les participants évaluaient leur sentiment de contrôle comme étant plus élevé qu'avec un effet inattendu. De façon intéressante, il existe un effet d'interaction entre la fluidité de la sélection de l'action et la probabilité d'obtention d'un effet. Lorsque l'action était suivie d'un effet attendu, il n'y avait pas de différence dans l'évaluation du contrôle que l'amorce soit congruente ou non. A l'inverse, lorsque l'action produisait un effet inattendu, l'impact de la fluidité de la sélection de l'action était plus important, c'est-à-dire que les amorces congruentes conduisaient à un sentiment de contrôle plus élevé que les amorces incongruentes.

Un autre exemple de pondération des indices en fonction du contexte nous est fourni par l'étude de Wen et al. (2015). Les auteurs ont montré comment le niveau global de performance pouvait également compenser la dégradation des signaux sensori-moteurs. Les participants devaient utiliser les touches droite et gauche du clavier pour déplacer un point vers une cible. Dans la condition "soi", l'ensemble des actions des participants était pris en compte; dans la condition "assistance", seules les actions rapprochant le point de la cible étaient prises en comptes, les actions éloignant le point de la cible étant ignorées. Ainsi, la condition "assistance" ne permettait pas une parfaite associa-

tion "intention-action-effet" car certaines actions n'étaient pas réalisées. Toutefois, ce système de corrections permettait aux participants de réaliser la tâche plus rapidement et maintenait leur sentiment d'agentivité. Les auteurs ont conclu que lorsque l'association action-effet était bruitée, les indices externes comme la performance avaient un poids plus important dans le maintien du sentiment d'agentivité.

Nous allons voir à présent comment l'utilisation de l'automatisation va dégrader la qualité des différentes informations nécessaires au développement du sentiment d'agentivité, voire les rendre absentes et comment compenser cette dégradation.

II.5 Problème HCI-Agentivité

Peu d'études se sont intéressées à l'impact de l'automatisation sur les processus sous-tendant le sentiment de contrôle des opérateurs et des utilisateurs. Pourtant, l'automatisation engendre des variables intermédiaires entre l'initiation d'une action et la perception d'un effet pouvant altérer notre sentiment de contrôle (Berberian et al., 2012 ; Limerick et al., 2014). En effet, comme nous l'avons dit précédemment, l'augmentation de l'autonomie des systèmes a entraîné un désengagement progressif de l'opérateur dans les actions à effectuer. De plus, l'opérateur se retrouve distancé de l'effet produit par l'action du système. Cette distance est à la fois physique avec l'élimination du bruit et des vibrations, ainsi que le contrôle déporté et la latence avant apparition de l'effet (e.g., pour les drones). Cette distance peut aussi être cognitive, avec une action qui est de plus en plus médiée et assistée par la technologie (e.g., les commandes de vol électriques remplaçant les commandes de vol classiques) ou encore à cause de l'opacité des systèmes qui peut perturber le sentiment de contrôle des opérateurs en leur empêchant d'accéder aux intentions du système.

Cet impact négatif de l'automatisation sur l'agentivité a été confirmé empiriquement par Berberian et al. (2012). La tâche des participants était de superviser la progression

d'un avion sur un plan de vol préétabli et d'intervenir si nécessaire. Lorsqu'un autre avion apparaissait sur ce plan de vol, les participants devaient effectuer une série d'actions pour changer le cap de leur avion par l'intermédiaire d'une interface de pilotage. Le niveau d'automatisation a été manipulé afin de prendre en charge une partie de plus en plus importante des actions à effectuer pour le changement de cap. Les résultats ont montré une diminution de l'agentivité (grâce à des mesures implicites et explicites) concomitante avec une augmentation de l'automatisation.

Un autre exemple concerne les tâches de téléopération lors desquelles la latence entre l'action de l'opérateur sur le système et la perception des effets associés va dégrader le sentiment d'agentivité de l'opérateur en l'empêchant d'utiliser des feedbacks visuels et/ou moteurs pour confirmer qu'il est bien l'auteur de l'action produite. Dans une tâche de pointage de cible, Berberian, Le Blaye, Schulte, Kinani, et Sim (2013) ont montré que la latence, c'est-à-dire le temps entre le mouvement du joystick et le mouvement du curseur, diminuait le sentiment d'agentivité du participant. Ils ont également montré que ce problème de latence pouvait être compensé par une aide visuelle affichant en temps réel la position estimée du curseur pendant le mouvement du participant, lui fournissant ainsi un feedback visuel précis de son action. Cette correction permet aux mouvements de l'opérateur d'être plus fluides et lui permet de lier facilement intention-action-effet répondant ainsi aux principes de l'agentivité.

Les modalités d'interaction ont également un impact sur l'agentivité. Coyle et al. (2012) ont montré qu'une action déclenchée par un appui sur son propre poignet entraîne plus d'agentivité (mesurée implicitement) qu'une action déclenchée par un appui sur le clavier. Limerick, Moore, et Coyle (2015) ont montré, avec un dispositif similaire, que l'agentivité des participants était diminuée lorsqu'ils utilisaient une commande vocale pour agir plutôt qu'un appui sur un clavier. Les auteurs de ces études justifient leurs résultats selon la théorie de l'intégration multiple en expliquant que l'appui sur la peau met en jeu davantage d'informations sensori-motrices que l'appui sur le clavier, lui-même

offrant plus d'informations que la commande vocale.

Chapitre III

Problématique

Dans cette thèse, nous souhaitons proposer des solutions permettant de remédier aux problèmes induits par l'automatisation, notamment au regard des mécanismes d'agentivité. Nous avons vu que l'agentivité était sous-tendue par l'intégration de différents indices provenant de l'action en elle-même, de l'environnement et du contexte. L'interposition de systèmes automatisés entre les opérateurs et les différentes tâches à effectuer a dégradé la qualité de ces indices et a donc entraîné une diminution de sentiment d'agentivité (Berberian et al., 2012 ; Limerick et al., 2014). Nous avons également vu que l'utilisation d'indices externes (aides visuelles, performance globale) permettait de limiter cette diminution (Berberian et al., 2013 ; Wen et al., 2015). Toutefois, ces deux résultats ont été obtenus lors de tâches avec un contrôle manuel du participant. Bien que les informations sensori-motrices étaient dégradées, elles étaient toujours présentes. En revanche dans le cadre de la supervision de systèmes hautement automatisés, ces informations sont absentes car l'action n'est pas produite par l'opérateur. Lorsqu'en plus le système automatisé est opaque, l'opérateur n'a pas accès aux intentions du système, ce qui l'isole encore plus des actions produites (sortie de boucle de contrôle).

Le modèle d'inférence proposé par Wegner et Wheatley (1999) nous semble être un cadre théorique pertinent pour traiter la question du sentiment d'agentivité en situation

de supervision. Les expériences de Wegner et Wheatley (1999) et Wegner et al. (2004) nous ont montré que l'utilisation d'amorces permet de recréer un sentiment d'agentivité même dans le cas où la personne ne produit pas l'action (i.e., *vicarious agency*). En particulier, les amorces permettent de faciliter la visualisation de l'effet à venir, recréant ainsi le lien entre intention et effet, lien nécessaire à l'établissement du sentiment de contrôle. L'utilisation d'amorces pourrait donc être un bon moyen pour rendre visibles les intentions d'un système et ainsi maintenir l'agentivité de l'opérateur qui le supervise et les principes de priorité et de cohérence de Wegner et Wheatley (1999) sont autant de guides pour mettre en forme ces amorces.

Les amorces ont déjà été utilisées dans le cadre des interactions homme-machine (Koo et al., 2015 ; Navarro et al., 2010). Nous avons vu précédemment que les résultats de Koo et al. (2015) montrent que le message expliquant l'action de la voiture et la raison de cette action augmentait les performances de conduite mais diminuait l'acceptabilité du système. Toujours dans le domaine de l'automobile, Navarro et al. (2010) ont également observé une dissociation entre performance et acceptabilité. Dans une tâche de simulation de conduite, ils ont trouvé que le système d'assistance amorçant au niveau du volant la trajectoire à prendre grâce à des vibrations directionnelles (*motor priming assistance*) était plus efficace que l'avertissement sonore latéralisé. Toutefois, les participants évaluaient le signal sonore plus acceptable que le système d'amorce motrice.

Ces résultats rejoignent ceux de Damen, van Baaren, et Dijksterhuis (2014). Dans leur étude, les participants devaient appuyer sur un bouton gauche ou droit pour générer des sons et il devait évaluer leur degré de contrôle sur l'apparition du son à la place de l'ordinateur. Avant d'appuyer, les participants étaient amorcés sub- ou supra-liminalement avec les mots "gauche" ou "droite". Les participants étaient libres d'appuyer sur le bouton qu'ils voulaient et ainsi d'effectuer une action congruente ou incongruente avec l'amorce. Les résultats ont montré que les amorces congruentes subliminales augmentaient le jugement d'agentivité comparé aux amorces subliminales incongruentes. A l'inverse, les

amorces congruentes supraliminales diminuaient le jugement d'agentivité comparé aux amorces incongruentes. Les auteurs ont conclu que les participants étaient conscients que leur choix était biaisé par l'amorce et donc un input extérieur ce qui réduit le sentiment de libre arbitre et donc d'agentivité.

Les études de Koo et al. (2015), Navarro et al. (2010), Damen et al. (2014) semblent indiquer que l'utilisation d'amorces consciemment perçues par l'opérateur peut être considérée comme une forme d'intrusion dans le processus de décision de l'opérateur, diminuant ainsi son sentiment de contrôle et son acceptabilité vis-à-vis du système. Toutefois dans le cas de la supervision, nous faisons l'hypothèse que ces amorces ne seront pas perçues comme intrusives car l'opérateur n'est pas celui qui produit l'action, comme pour les études de Sato (2009) et Wegner et al. (2004). Nous pensons au contraire que ces informations seront perçues comme une forme de communication de la part du système.

L'objectif général de cette thèse est donc d'utiliser le cadre théorique de l'agentivité pour aller plus loin dans la compréhension du rôle des amorces afin d'améliorer la supervision de systèmes fortement automatisés tant au niveau des performances que de l'acceptabilité du système par les opérateurs. Afin d'y parvenir, nous allons tester dans un premier temps dans quelle mesure la présentation des intentions du système, grâce à la technique d'amorçage, permet de diminuer son opacité, de maintenir l'agentivité de l'opérateur et son contrôle effectif (Expériences 1 et 2). Dans un second temps, nous souhaitons montrer que le maintien de l'agentivité s'accompagne d'une augmentation de l'acceptabilité du système par les opérateurs (Expérience 3). Pour finir, nous faisons l'hypothèse que les études empiriques sur l'agentivité peuvent nous aider à optimiser cet effet d'amorçage en déterminant la meilleure temporalité à laquelle présenter l'amorce (Expérience 4).

Chapitre IV

Matériel et Méthodes

IV.1 Dispositif expérimental

Nous avons utilisé pour les quatre expériences de cette thèse un simulateur de page radar développé par l'ONERA : le LIPS (Laboratoire d'études des Interactions Pilote-Systèmes, voir Figure IV.1). Ce simulateur nous a permis de réaliser des tâches de détection d'obstacles, tâches plus adaptées au cadre des interactions homme-machine que celles traditionnellement employées dans les études sur l'agentivité. Chacune des tâches est décrite dans la partie "Matériel et Méthodes" des articles. Le principe général est toujours le même : un avion sous contrôle du pilote automatique poursuit sa route selon un plan de vol préétabli. Des obstacles peuvent apparaître sur le plan de vol de l'avion. Lorsque l'obstacle est détecté par le système automatisé, le participant en est averti et une nouvelle direction va être implémentée pour l'éviter. En fonction des expériences et des différentes conditions au sein des expériences, un message amorce peut apparaître afin d'indiquer aux participants quelle sera la future direction de l'avion avant son implémentation. Si la direction ne permet pas la résolution du conflit, le participant devra détecter l'erreur le plus rapidement possible.

Lors des Expériences 2, 3 et 4, les participants doivent effectuer une autre tâche en

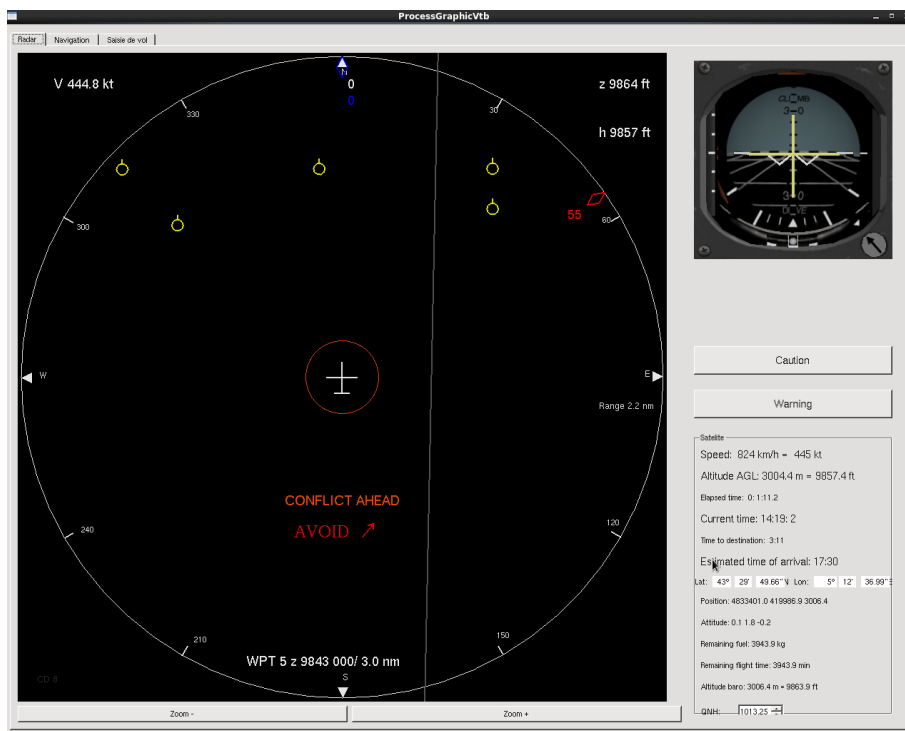


FIGURE IV.1 – Exemple d’essai sur l’interface du LIPS

parallèle en plus de la tâche de détection de conflit. Ces tâches secondaires permettent d’alourdir la charge de travail des participants et ainsi de complexifier la tâche principale. Ces tâches secondaires servent également à rendre l’expérience encore plus écologique. En effet dans les Expériences 2 et 3, la tâche secondaire est une tâche de poursuite de cible et pour l’Expérience 4, la tâche est la *Multi-Attribute Tasks Battery* (Comstock & Arnegard, 1992 ; Santiago-Espada, Myer, Latorella, & Comstock, 2011) développée par la NASA et très utilisée dans le monde de l’aéronautique.

IV.2 Mesures

Dans cette partie, nous allons présenter les différentes mesures utilisées à travers cette thèse. Dans un premier temps, nous présenterons les mesures explicites et implicites de

l'agentivité que nous avons utilisées pour la première étude (Expériences 1 et 2). Ensuite, nous présenterons l'échelle d'acceptabilité ainsi que les mesures d'attention obtenues grâce à l'oculomètre que nous avons utilisées dans les études 2 et 3. Pour finir, nous dirons quelques mots sur la mesure de performance que nous avons utilisée sur l'ensemble des études.

IV.2.1 Mesures de l'agentivité

Deux types de mesures sont couramment utilisés dans les études sur l'agentivité : les mesures explicites et implicites. Ces deux types de mesures sont à la fois complémentaires mais non redondantes (Dewey & Knoblich, 2014) car elles reflèteraient deux niveaux distincts du sentiment d'agentivité. En effet, Synofzik et al. (2008) ont proposé un modèle du sentiment d'agentivité à deux niveaux (*a multifactorial two-step account of agency*). Le premier niveau, la sensation d'agentivité (*feeling of agency*), correspond à l'agentivité pré-réflexive et automatique. Ce niveau ne nécessite pas un accès conscient et se base sur des informations de "bas niveau cognitif" (signaux moteurs, etc.). A ce niveau, l'action est simplement catégorisée comme étant réalisée par soi-même ou non, il n'y a donc pas d'attribution causale externe possible. Le second niveau, le jugement d'agentivité (*judgment of agency*), correspond à la capacité à attribuer explicitement la source qui a causé l'action (soi ou autrui). Ce niveau se base sur des informations de plus "haut niveau cognitif" (contexte, croyances, etc.). Ces deux niveaux sont sous-tendus par les mécanismes intégrant de manière optimale les différents indices internes et externes. Ainsi, les mesures explicites permettraient de rendre compte du jugement d'agentivité alors que les mesures implicites seraient des indicateurs de la sensation d'agentivité (Moore, Middleton, Haggard, & Fletcher, 2012).

Mesures explicites

Les mesures explicites représentent le type de mesures le plus utilisé dans les études sur l'agentivité grâce à sa facilité de mise en place. Elles consistent pour la plupart à demander aux participants d'évaluer sur une échelle de Likert ou de pourcentage, le niveau de contrôle ou d'agentivité qu'ils ressentent sur l'apparition d'un effet. Par exemple, Sato et Yasuda (2005) demandaient aux participants d'évaluer sur une échelle de 0 (absolument pas d'accord) à 100 (totalement d'accord), la proposition : *"je suis celui qui a produit l'effet"*. Dans l'expérience de Chambon et Haggard (2012), les participants devaient évaluer le contrôle qu'ils ressentaient sur l'effet produit en utilisant une échelle de 1 (pas de contrôle) à 8 (contrôle total).

Nous avons vu précédemment que l'utilisation d'amorces permettait d'influencer ces mesures explicites (Aarts et al., 2005 ; Wegner et al., 2004 ; Wegner & Wheatley, 1999). D'autres études ont également montré que ces mesures explicites pouvaient être sensibles à d'autres paramètres, comme la probabilité d'apparition d'un effet. Par exemple, Sato et Yasuda (2005) ont montré que les participants pouvaient ressentir une diminution du sentiment d'agentivité pour les effets inattendus de leurs propres actions.

Dans l'Expérience 1 de la première étude, les participants devaient évaluer leur capacité à reprendre en main le système automatisé en cas d'erreur, en utilisant une échelle de 0 (pas capable) à 7 (hautement capable). Cette question est très différente des questions traditionnellement posées sur l'agentivité. Toutefois dans cette expérience, les participants réalisaient des essais en contrôle direct ou en supervision. Nous avons fait l'hypothèse qu'une question classique du type : *"je suis en contrôle de l'action"* donnerait des résultats binaires car elle pouvait être comprise par les participants comme *"qui contrôle l'appareil, vous ou le système automatisé ?"* Contrairement à l'étude de Wegner et Wheatley (1999), la source de l'action n'est pas ambiguë, c'est clairement le système automatisé qui produit le changement de cap de l'appareil. Nous ne souhaitons pas que le participant s'attribue les actions de ce système, mais plutôt qu'ils se sentent en contrôle

de ces actions de la même façon que pour l'expérience de Wegner et al. (2004). Une façon d'évaluer le contrôle qu'à un opérateur sur un système qu'il supervise est d'évaluer la qualité de la reprise en main lorsqu'il y a une panne ou erreur du système, ce qui explique pourquoi nous avons choisi cette question pour évaluer explicitement l'agentivité des participants.

A l'inverse, pour l'Expérience 4, les participants n'ont pas eu à réaliser la tâche en contrôle manuel mais seulement en supervision, ce qui diminuait la probabilité d'obtenir des réponses stéréotypées. De plus, nous préférons utiliser une question plus classique ("*je me sens en contrôle lorsque j'utilise ce système*") afin d'établir plus facilement le parallèle avec l'expérience de Wegner et Wheatley (1999).

Mesures implicites

Les mesures implicites se basent sur les différences perceptuelles qui existent entre les effets produits par mes actions et les effets produits par l'action d'autrui (Dewey & Knoblich, 2014). Il existe deux mesures implicites du sentiment d'agentivité : le liage intentionnel et l'atténuation sensorielle (Moore, 2016). En effet, il a été montré que lorsqu'une personne se sentait agent d'une action, il percevait une contraction temporelle entre l'action et l'effet et tendait à atténuer les conséquences sensorielles de cette action.

Le phénomène d'atténuation sensorielle a été étudié dans le domaine somato-sensoriel (Blakemore et al., 1998), auditif (Sato, 2009) et visuel (Cardoso-Leite, Mamassian, Schütz-Bosbach, & Waszak, 2010 ; Roussel, 2013), ce qui en fait une mesure pertinente du sentiment d'agentivité pour les interactions homme-machine étant donné la diversification des modalités d'interaction (contrôle vocal, tactile, etc.). Elle semble toutefois plus compliquée à mettre en place, au moins pour des recherches exploratoires, que son homologue implicite, le liage intentionnel, car comme toutes mesures en psychophysique, elle nécessite un très grand nombre d'essais et une discrimination dans les variables utilisées non compatibles avec cette thèse. Nous allons donc surtout nous concentrer sur le liage

intentionnel.

Le liage intentionnel

Le liage intentionnel (*Intentional Binding*, IB) fournit également une autre preuve du rôle important de la contiguïté temporelle entre action et effet dans la genèse de l'agentivité. L'IB a été reporté pour la première fois par Haggard, Clark, et Kalogeras (2002). Ils ont trouvé que pour des participants qui réalisaient une action volontaire (i.e., un appui sur un bouton) les actions étaient perçues comme se décalant vers les effets (et donc arrivant plus tardivement) et les effets étaient perçus comme décalés vers l'action (et donc arrivant plus tôt). A l'inverse, lorsque le mouvement d'appui sur le bouton était réalisé involontairement par l'utilisation de la stimulation magnétique transcrânienne (TMS), les participants n'étaient plus sujets à cette compression de l'intervalle temporel entre l'action et les conséquences sensorielles.

La mesure du liage intentionnel a été largement utilisée et répliquée avec des protocoles différents (pour une revue, voir Moore & Obhi, 2012). Différentes méthodes d'opérationnalisation de la mesure ont été proposées : la méthode initialement utilisée par Haggard, Clark et Kalogeras (2002) s'appuyait sur la méthode de l'horloge de Libet (Libet, Gleason, Wright, & Pearl, 1983). Les participants devaient estimer l'occurrence de l'action ou de l'effet lors d'essais séparés en estimant la position de l'aiguille sur l'horloge (Engbert, Wohlschläger, & Haggard, 2008 ; Haggard et al., 2002 ; Moore et al., 2009 ; Wohlschläger, Haggard, Gesierich, & Prinz, 2003). Plus récemment, une méthode de mesure alternative a été proposée, l'estimation directe de l'intervalle, et non plus l'estimation temporelle de l'action et de l'effet dans des essais séparés (Buehner & Humphreys, 2009 ; Coyle et al., 2012 ; Moore et al., 2009). La reproduction de l'intervalle perçu a également été utilisée (Faro, 2010 ; Humphreys & Buehner, 2010) mais ne constitue pas la méthode la plus fréquente.

Initialement, ce phénomène a été démontré pour des échelles de temps très courtes (de

l'ordre de la milliseconde). Haggard et al. (2002) ont montré que le phénomène de liage intentionnel ne fonctionnait plus pour des délais supérieurs à 650 ms. Toutefois, cette estimation a été réalisée sur une tâche simple d'appui-bouton. Berberian et al. (2012) ont montré que ce phénomène était toujours présent pour des délais supérieurs à 2 s lors d'une tâche de supervision. Faro, McGill, et Hastie (2013) ont montré que cet effet était encore présent pour des échelles de temps de l'ordre de la minute pour une expérience de consommation et de perception d'effets associés. La fenêtre temporelle du phénomène de liage intentionnel serait donc dépendante de la tâche effectuée.

Parmi les différentes méthodes d'évaluation du liage intentionnel, nous avons choisi d'utiliser l'estimation directe de l'intervalle temporelle pour sa facilité de mise en place et parce qu'elle correspondait à la méthode déjà utilisée par l'équipe pour de précédentes recherches sur le sujet (Berberian et al., 2013, 2012). Les participants devaient évaluer sur une échelle de 0 à 3 s, l'intervalle de temps perçu entre l'action, ici matérialisée par l'envoi de la commande (feedback sonore) et l'effet, ici matérialisé par le changement de cap de l'avion (feedback sonore et visuel).

IV.2.2 Mesures d'acceptabilité

Nous avons vu précédemment que l'automatisation pouvait avoir un impact négatif sur l'acceptabilité des systèmes. Etant donné que le contrôle perçu par l'opérateur apparaît comme déterminant pour l'acceptabilité des systèmes (Shneiderman & Plaisant, 2004), nous avons émis l'hypothèse que l'utilisation de principes issus des recherches sur le sentiment d'agentivité pourrait compenser cette baisse d'acceptabilité ressentie par les opérateurs. Il existe une multitude de questionnaires évaluant l'acceptabilité des utilisateurs. En effet, en fonction du domaine (aviation, automobile, etc.), du niveau d'expertise (pilote confirmé, élève pilote, quidam, etc.) ou encore de la nature du système automatisé (aide à la conduite, voiture autonome), les questionnaires sont adaptés à chaque type de situations. Au regard du temps de chaque expérience (pratiquement 2 heures), de la

population visée (des étudiants sans connaissance avancée en aéronautique) et de la nature de la tâche (supervision d'un pilote automatique), nous avons choisi d'élaborer une version simplifiée et raccourcie des questionnaires de Verberne et al. (2012) et Midden, Verberne, et Ham (2012) dont les expériences ont été citées précédemment. Ainsi dans les Expériences 3 et 4, nous avons mesuré le niveau d'acceptabilité grâce à un questionnaire contenant sept propositions que les participants devaient évaluer en utilisant une échelle de Likert en sept points (1 = totalement pas d'accord avec la proposition, 7 = totalement d'accord) :

1. Utiliser ce système est stressant
2. Je considère ce système comme fiable
3. J'ai confiance en ce système
4. Il est difficile pour moi de comprendre ce que va faire le système
5. Utiliser ce système ne demande pas un niveau d'attention élevé
6. Ce système augmente mes performances de supervision
7. Ce système augmente le niveau de sécurité

Les réponses à chaque proposition étaient moyennées pour former une mesure plus robuste de l'acceptabilité (alpha de Cronbach = .87). Un score élevé reflète une meilleure acceptabilité vis-à-vis de l'automatisme.

IV.2.3 Performance

L'automatisation n'étant pas encore assez puissante pour pouvoir gérer l'ensemble des événements se produisant dans l'environnement, la question de la reprise en main par l'opérateur est centrale pour améliorer les interactions homme-système. Les mesures présentées précédemment constituent des mesures subjectives de l'acceptabilité et de l'agentivité; même si le liage intentionnel est une mesure implicite de l'agentivité, elle

n'en reste pas moins subjective. En plus de vérifier l'impact des messages amorces sur le sentiment de contrôle, il nous fallait mesurer le contrôle objectif des participants. Afin de simplifier le dispositif expérimental, nous nous sommes concentrés sur le temps que le participant mettait à détecter une erreur du système et à la reprendre en main. Le participant n'avait pas le choix entre plusieurs actions possibles pour reprendre en main le système, donc la qualité de sa reprise en main n'a pas pu être évaluée contrairement à l'étude de Gold et al. (2013). C'est donc sur le temps de réaction que s'est fait l'ensemble des analyses de performance à travers les quatre expériences de cette thèse.

IV.2.4 Oculométrie

Afin de vérifier si la supervision de systèmes automatisés répondant aux principes de l'agentivité entraînait une gestion différente des ressources attentionnelles, nous avons utilisé un système d'oculométrie portatif (SensoMotoric Instruments© Eye Tracking Glasses 2GW ; Sampling rate : 120Hz) pour les Expériences 2 et 3 et un système d'oculométrie fixe (Smart Eye 3D ; Sampling rate : 120Hz) pour l'Expérience 4. Le but de ces outils était d'évaluer le nombre d'allers-retours effectués entre les deux écrans et le temps passé sur chaque tâche. Adler et Benbunan-Fich (2012) ont montré qu'alterner trop fréquemment entre plusieurs tâches tendait à dégrader la performance globale. De plus, le temps passé sur une zone d'intérêt est traditionnellement considéré comme étant une indication de la difficulté à extraire l'information pertinente de cette zone (Fitts, Jones, & Milton, 1950 ; Jacob & Karn, 2003). Ces mesures objectives de la gestion des ressources attentionnelles nous permettent de compléter les mesures de performances.

Nous ne nous sommes pas intéressés au diamètre pupillaire qui est également considéré comme un marqueur physiologique de la charge de travail (Jepma & Nieuwenhuis, 2011) et de l'attention (Verney, Granholm, & Marshall, 2004). En effet, les participants réalisaient des tâches en parallèle nécessitant un grand nombre d'allers-retours entre les deux écrans ce qui aurait largement bruité la mesure étant donné la sensibilité de la

pupille aux différences de luminosité.

Contributions Expérimentales

Première Etude

Automation technology and sense of agency

Kevin Le Goff^{1,2}, Arnaud Rey², Patrick Haggard³, Bruno Berberian¹

¹ Systems Control and Flight Dynamics Department, ONERA, Salon de Provence, France

² Laboratoire de Psychologie Cognitive & CNRS, Aix-Marseille University, Marseille, France

³ Institute of Cognitive Neuroscience, University College London, United Kingdom

Cette étude est actuellement soumise à International Journal of Human-Computer Interaction

Cette première étude avait pour but de vérifier qu'un système automatisé fournissant des informations sur ses futures actions permettrait de maintenir le sentiment d'agentivité de l'opérateur tout en améliorant son contrôle effectif par rapport à la supervision d'un système opaque. Afin de tester nos hypothèses, nous avons réalisé deux expériences dans lesquelles les participants ont réalisé une tâche de gestion du trafic aérien. Dans la première expérience, nous nous sommes concentrés sur le sentiment d'agentivité et l'utilisation de mesures implicite (le liage intentionnel) et explicite (report verbal) en comparant plusieurs niveaux d'interactions entre les participants et le système automatisé. Ces différents niveaux allaient du contrôle manuel jusqu'à la présentation de messages amorces indiquant à la fois la temporalité et la direction de la prochaine action du système. Dans une seconde expérience, nous avons vérifié l'impact de ces messages amorces sur la rapidité de la reprise en main des opérateurs et sur la gestion de leurs ressources attentionnelles. Dans cette expérience, les participants ont réalisé deux tâches en parallèle. La première tâche correspondait à la tâche de supervision et la seconde était une tâche de poursuite de cible.

Abstract

Operators seem often helpless to takeover automation in case of failure. This "out-of-the-loop" problem occurs when operators are unable to predict the outcome of actions of the system, causing a decrease of control. In this context, we propose that making the system's intentions more visible and accessible to the operator could both increase human operator performances in takeover situations, and increase the sense of agency of human operators. In two experiments, we tested our hypothesis in an aircraft supervision task by using supraliminal primes providing information about the action the automation is about to perform. We showed that priming the system's intentions increases both performance in takeover situation and in the feeling of control, together with a change in the way participants allocated their attentional resources. Similar as to human interactions, the present results indicate that having information about the system's intentions improves human-system interactions by increasing the system's acceptability and by enabling a much more efficient supervision.

Introduction

Automation technology is a ubiquitous phenomenon that is profoundly changing our everyday life. While the public is generally enthusiastic about the possibilities that it offers, recent tragedies remind us of the difficulties human operators have in supervising highly automated systems. Understanding and overcoming these difficulties remain a crucial challenge in the Human-Computer Interaction (HCI) domain.

Automation technology and takeover difficulties

Automation systems are classically considered as a simple substitution of a human activity by a machine activity (*the substitution myth*, Woods & Tinapple, 1999). However, studies on the relationship between humans and technology strongly question this traditional view. Empirical data indeed suggest that automation is frequently accompanied by a decrease in operator's performance, decrements in vigilance such as reduced sensitivity to important signals (Billings, 1991 ; Wiener, 1988), complacent or excessive trust in system ability (Parasuraman, Molloy, & Singh, 1993), and loss of operator situation awareness (Carmody & Gluckman, 1993 ; Endsley, 1996 ; Endsley & Kiris, 1995). These performance decrements are examples of the so-called out-of-the-loop (OOL) performance problem, in which human operators become disconnected from the automation system (Endsley & Kiris, 1995 ; Kaber & Endsley, 1997). The OOL performance problem is particularly problematic in cases of system failure. In these critical situations, human supervisors seem effectively helpless at diagnosing the situation, determining the appropriate solution and retaking control over the automated system (Billings, 1991).

These difficulties point to a paradoxical situation where the human operator is replaced by automation technology in the loop of control and the automated technology is designed in a way that complicates the transfer of control from the machine to the human in takeover situations. Nevertheless, system designers and safety authorities consi-

der the human operator as the last barrier in case of system failure. In that sense, until automation becomes powerful enough to handle all abnormalities, enabling more efficient transfer of control from system to operators in case of system failure appears as a critical issue. How to make this transfer of control easier and more efficient is the main purpose of this paper.

Together with this safety concern, the interposition of automated systems between human operators and processes creates new difficulties related to the sense of control. Improving acceptance of new technologies by human operators is an important area of concern to equipment suppliers (Horberry, Stevens, & Regan, 2014). To be acceptable, new technologies must be reliable, efficient and useful. However, such qualities do not guarantee acceptance from human operators. As pointed by Shneiderman and Plaisant (2004), users strongly desire the sense that they are in charge of the system and that the system responds to their actions. Increasing automation has the potential to seriously threaten this sense of control. This echoes Baron (1988), who claimed : "Perhaps the major human factors concern of pilots in regard to introduction of automation is that, in some circumstances, operations with such aids may leave the critical question, who is in control now, the human or the machine?" The second purpose of this paper is precisely to determine how to compensate for this decrease in the sense of control in presence of highly automated technologies.

OOL performance problem and system opacity

The lack of system predictability is certainly a central point in understanding OOL phenomena and associated difficulties of takeover (Christoffersen & Woods, 2002 ; Dekker & Woods, 2002 ; Klein et al., 2004 ; Norman, 1990). With the progress of technology, current man-made complex systems tend to develop cascades and runaway chains of automatic reactions that decrease, or even eliminate predictability and cause outsized and unpredicted events (Taleb, 2012). This is what we may call "system opacity" : the

difficulty for a human operator to have a clear idea of the system's intentions and to predict the sequence of events that will occur. In that sense, the main problem with automation is not the automation per se, but rather its inappropriate design within the human-computer interaction (Norman, 1990). Although such difficulties have been identified for a long time, how to design more predictable systems remains a critical challenge (Bainbridge, 1983 ; Baxter, Rooksby, Wang, & Khajeh-Hosseini, 2012).

As proposed by (Norman, 1990), what is needed is continual feedback about the state of the system, in a normal, natural way, much in the same way that human participants in a joint problem-solving activity will discuss the issues among themselves. A possible approach to reach this goal focusses on how humans understand and control their own actions. Indeed, we can assume that operators interpret the intentions and the outcomes' actions of a system with their own "cognitive toolkit". Thus, understanding how this "cognitive toolkit" works could be relevant in order to propose design principles for improving the system's controllability.

A new theoretical framework : the sense of agency

The mechanisms underlying the experience of intentional causation and the sense of control of our own actions are the main concerns of the science of agency (Pacherie, 2008). Gallagher (2000) defined agency as "the sense that I am the one who is causing or generating an action". Synofzik, Vosgerau, and Newen (2008) proposed a two-step account of agency. According to this model, the basic non-conceptual feeling of agency is produced by a gradual and highly plastic sub-personal weighting process of different action-related perceptual and motor cues (feeling of agency). This pre-conceptual core is further processed by conceptual capacities and attitudes (e.g., beliefs, desires) to form an attribution of agency (judgement of agency). The extent to which the feeling and judgement of agency, respectively, contribute to the overall sense of agency depends on the context and task requirements.

Recently, the concept of agency has been applied to the HCI domain (Berberian et al., 2013, 2012 ; Limerick et al., 2014 ; McEneaney, 2009, 2013 ; Obhi & Hall, 2011). During the interactions with technology, it has been shown that the simple process of producing an action to cause an intended outcome depends upon several variables that can alter the agentic experience dramatically (Berberian et al., 2012 ; Limerick et al., 2014). Here we ask what information and at what time is necessary for promoting the sense of agency and facilitating the takeover.

Although the mental processes contributing to the sense of agency are not yet fully understood, current approaches suggest that we derive a sense of being the agent of our own actions from a cognitive mechanism that computes the discrepancies between the predicted consequences of our own actions and the actual consequences of these actions, as in action control models (Blakemore et al., 2000, 1998, 2002). Recently, Pacherie (2011) suggested that the various mechanisms underlying the sense of agency for individual actions are of the same kind as those underlying the sense of agency that one experiences while being engaged in joint action. That is, the sense of agency while being engaged in joint action is based on the same principle of congruence between predicted and actual outcome. In particular, some representation of the other's intended or predicted outcome is required. On this view, joint action with an artificial agent, such as an automation system, should be possible, as long as the system's intentions are clear. If the system's opacity makes its intentions difficult to detect and control, sense of agency during joint action with an artificial agent will be low (Berberian et al., 2012 ; Obhi & Hall, 2011 ; Wohlschläger et al., 2003).

In this context, we propose that making the system's intentions more visible and accessible to the operator should both improve the joint action between the human and the automation system (and particularly, in takeover situations) and increase the sense of agency of the human operator. Wegner, Sparrow, and Winerman (2004) have notably shown that priming (i.e., presenting predictive information of what will happen next) can

create and communicate intentions, and can therefore facilitate sense of agency (Aarts et al., 2005 ; Gentsch, Kathmann, & Schütz-Bosbach, 2012 ; Gentsch & Schütz-Bosbach, 2011 ; Linser & Goschke, 2007 ; Wegner & Wheatley, 1999). Importantly, priming can produce a sense of agency even in the absence of any movement (but see also Moore et al., 2009 ; Sato, 2009). In this study, participants watched themselves in a mirror while another person behind them, hidden from their view, moved her/his hands forward on each side where participants' hands would normally appear and performed a series of movements. When participants could hear instructions priming a particular movement, and then subsequently saw that movement, they reported a high degree of agency for the movements, even if they did not themselves act at all.

In this paper, we investigate whether priming the system's intentions, by giving information about which action the system is about to perform, would influence performances in takeover situations, and would increase the sense of agency of the human operator. If primes allow participants to access the system's intentions, and thereby to increase its predictability, we should observe an increase in the sense of agency through both verbal reports and implicit measures of agency (i.e., intentional binding, IB, Experiment 1) together with an increase in performance in takeover situations (Experiment 2).

Experiment 1

Method

Participants

Twenty-five undergraduate students from the Aix-Marseille University with normal or corrected to normal vision participated in this experiment (mean age = 26.6 ; SD = 3.6 ; 15 females). All participants were naive to the purposes of the study, and had no particular expertise in aircraft control.

Materials

This experiment was conducted using ONERA's airplane simulator (LabSim, ONERA-Salon de Provence). The simulator was composed of two operating windows (see Fig. V.1). On the left, there was a navigation display representing the aircraft in the horizontal plane together with the surrounding traffic. On the right, there was an autopilot interface that allowed the participant to change the horizontal trajectory of the aircraft in case of conflict, defined as being too close to another aircraft.

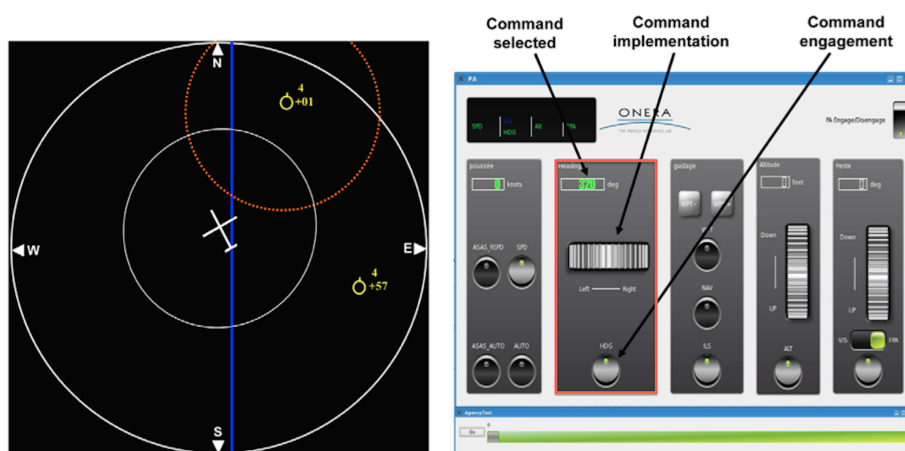


FIGURE V.1 – Experimental set up with the navigation display on the left and the autopilot interface on the right. Left : the large white circle represents the angular indication. The smaller white circle that surrounds the aircraft represents the safety envelope. A conflict appears when an obstacle (yellow circle) penetrates the safety envelope leading to the appearance of a red circle surrounding the obstacle. Right : only the commands surrounded by a red rectangle are used in this experiment. Participants had to implement the adapted direction by scrolling the horizontal wheel and to validate it by pressing the engagement command. A digital dial ("command selected") indicates the value of the direction selected by the participant.

Experimental procedure

Participants had to supervise the progress of an aircraft on a predefined flight path, through four conditions : one "manual condition" and three "automated conditions". At

the beginning of each trial, participants were informed about the condition. Figure V.2 provides a descriptive summary of the four experimental conditions. After a delay of 3 seconds, an obstacle (another aircraft) appeared on their path and could penetrate the safety envelope of the plane if there was no change in its direction. In the manual condition, participants had to implement the adapted direction (by scrolling a horizontal wheel) and to engage it (by pushing the button situated below the wheel) in order to avoid the conflict (see Fig. V.1 for a visual description of both commands). In the three automated conditions, the autopilot system automatically detected and selected a new direction to avoid the potential conflict without any intervention of the participant. Then, in contrast to the manual condition, no movement was performed during these automated conditions. The three automated conditions were defined by the degree of information displayed by the autopilot to participants when a conflict was detected. When an obstacle was detected by the system in the "no prime" condition, the autopilot selected a new direction to avoid a conflict, without displaying any information. In the "temporal prime condition", when the autopilot selected a new direction, the message "AVOID" appeared on the screen indicating that the plane was about to avoid the obstacle. In the "directional prime condition", the message "AVOID" appeared together with an arrow indicating the direction selected by the system before implementing this command. The "temporal prime condition" was used to evaluate the effect of providing temporal information about the command to be executed. By adding an arrow in the "directional prime condition", it was possible to test the effect of providing information about the subsequent action. Note that primes were always predicting the action. Regardless of the condition, a sound was sent when the new direction was engaged. After a controlled temporal delay (see below "Action/Effect delay"), feedback concerning the success of the performed action was given to the participant. It consisted of a green message "Resolving" plus a sound. This feedback indicated the end of a trial.

According to the two-step model of agency, we performed both implicit and expli-

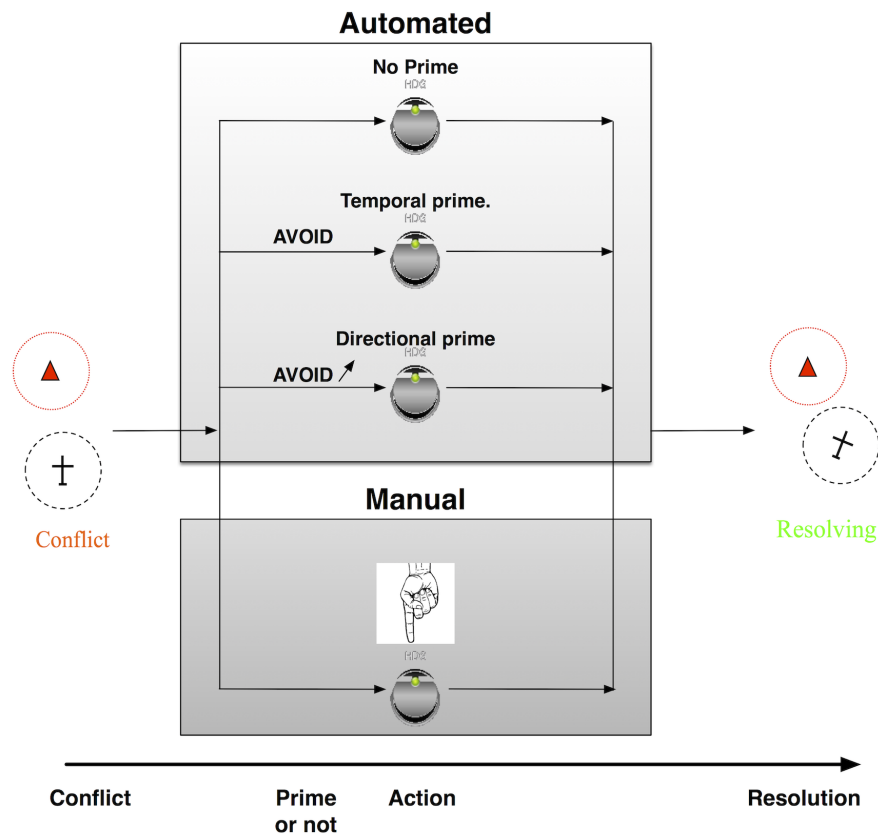


FIGURE V.2 – Sequence of events on each trial in the four experimental conditions : manual condition ; no prime condition ; temporal prime condition ; directional prime condition.

cit measures to respectively evaluate the feeling and the judgement of agency, the two component of the overall sense of agency. At the end of each trial, participants had to estimate the perceived temporal delay between the engagement of the new direction and the appearance of the visual feedback ("Resolving") and the sound on a scale ranging from 1 ms to 3000 ms. In fact, only three Action/Effect delays (1000 ms, 1750 ms, and 2500 ms) were presented in a random order. This delay estimation task was performed to evaluate intentional binding (IB) (Haggard et al., 2002 ; Moore & Obhi, 2012). IB corresponds to the phenomenon in which the interval between a voluntary action and an outcome is perceived as shorter than the interval between a physically similar involun-

tary movement and the same outcome event. IB is traditionally considered as an implicit measure of agency (Moore et al., 2012) and refers to the feeling of agency.

In the three automated conditions, after the delay estimation task, participants made explicit judgements of agency, by verbally reporting how much they felt able to take over the system in case of failure, using a scale from 0 (not able) to 7 (highly able). This explicit question allowed us to estimate the participant's judgement of control during this supervision task.

The experimental session was organized in the following way. Participants began with a training session during which they had to perform three "manual" trials to learn how to implement a new direction in the manual condition, and one trial for each "automated condition" to learn how the autopilot worked depending on the degree of information displayed by the autopilot. Following this training session, there was a "takeover" block composed of three "automated" trials wherein a takeover was necessary. A takeover was considered as necessary if the direction selected by the autopilot led to a collision with an obstacle. If it was the case, participants had to select a new direction to avoid the conflict in the same way as in the manual condition.

After the "takeover" block, participants had to perform an "experimental" block composed of nine experimental safe trials. The participants were informed that these trials did not necessitate any takeover. After each trial, they had to evaluate the temporal Action/Effect delay and to answer the question about agency. Participants then had to perform another "takeover" block followed by another "experimental" block. This sequence was repeated four times. At the beginning of each block, the experimenter informed the participant about the type of block she/he was about to perform. Each participant performed 36 experimental trials : 3 trials * 3 Action/Effect delays * 4 conditions in a random order. "Takeover" blocks were included in order to justify the takeover question after each experimental trial, but were not considered as experimental trials. These "takeover" blocks were explicitly separated from "experimental" blocks so that

participants knew that the system was flawless during experimental trials. They could therefore focus on the supervision, temporal estimation, and questions about agency.

Results

Each participant made nine temporal judgements in each condition (mixing together the three Action/Effect delays). We performed a repeated measures ANOVA on these temporal judgements and we found a significant effect of automation level ($M_{\text{manual}} = 1130$ ms, $M_{\text{no prime}} = 1321$ ms, $M_{\text{temporal}} = 1322$ ms, $M_{\text{directional}} = 1409$ ms; $F(3, 72) = 6.54$; p -value < 0.01 ; see also Figure V.3). Post-hoc analyses revealed that interval estimates depended on the level of automation. We performed comparisons between all four groups. We applied the Benjamini-Hochberg method to correct each p -value (Benjamini & Hochberg, 1995, 2000). First, when we contrasted the values in the manual condition with each of the other automated conditions, the three paired t -tests (manual vs. no prime; manual vs. temporal prime; manual vs. directional prime) revealed a significant difference (all $p_s < 0.05$ after corrections for multiple comparisons). However, no significant differences were found between the three automated conditions (all $p_s > 0.05$). Therefore, although participants tended to underestimate the Action/Effect interval regardless of the condition, when they directly controlled the system (manual condition), they perceived the Action/Effect delay as shorter than when the system was controlled by the autopilot.

On trials in which participants had to report their ability to take over the system in case of failure, we performed a one-way ANOVA with repeated measures on the three automated conditions. Results revealed that the level of automation impacted the feeling of control ($F(2, 48) = 6.927$; p -value < 0.01 ; see also Fig. V.4). To study the benefit of providing temporal information, we first performed a post-hoc comparison between the "no prime" and "temporal prime" conditions, and we found no difference between these

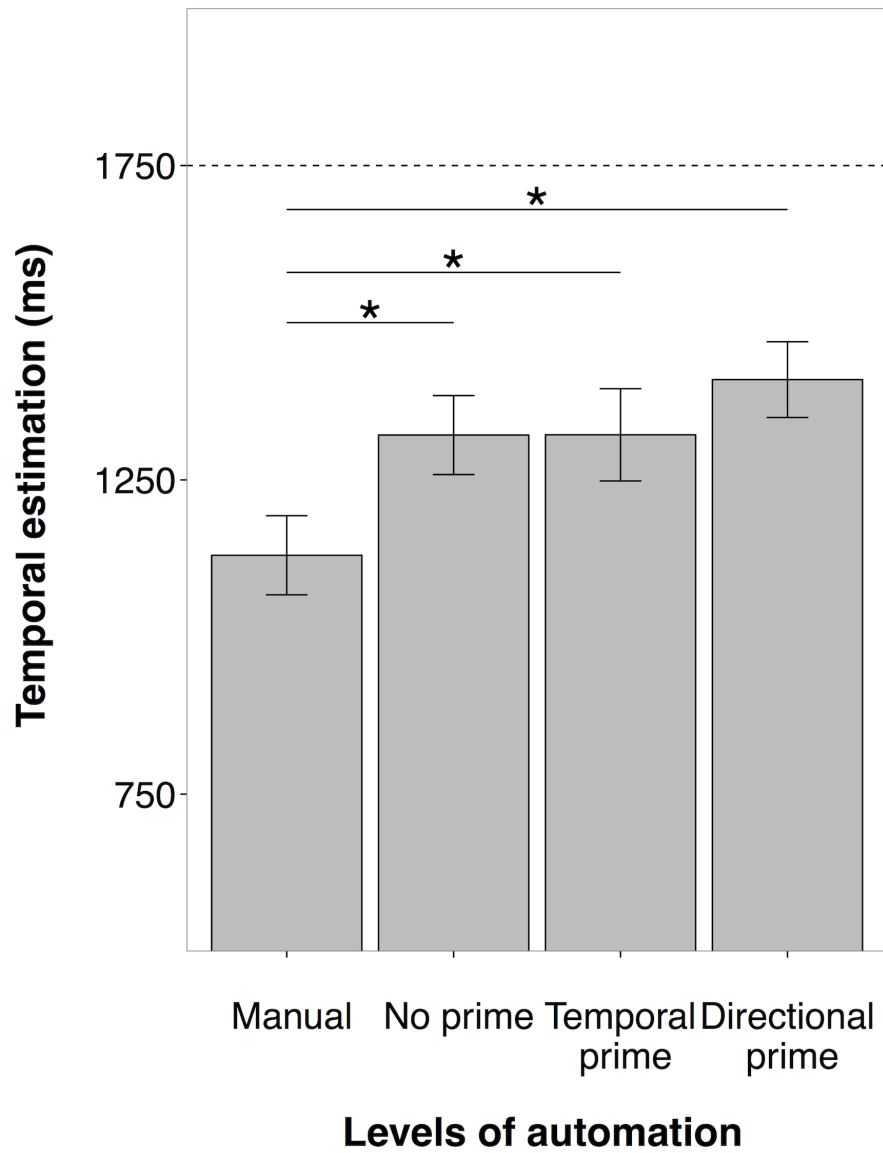


FIGURE V.3 – Mean values of temporal estimation per condition. Error bars represent standard error of the mean. The dashed line at 1750ms indicates the mean of the actual intervals presented.

conditions ($t(24)=-0.51$; p -value > 0.05). Second, to test the benefit of providing the system's intentions, we performed two post-hoc comparisons with Benjamini-Hochberg corrections between "no prime vs. directional prime" ($t(24)= -2.96$; p -value < 0.05 , with

correction) and "temporal prime vs. directional prime" ($t(24) = -3.49$; p -value <0.05 , with correction) and found that when the arrow indicating the direction of the command to be executed was provided, participants reported a greater confidence in their ability to take over the system.

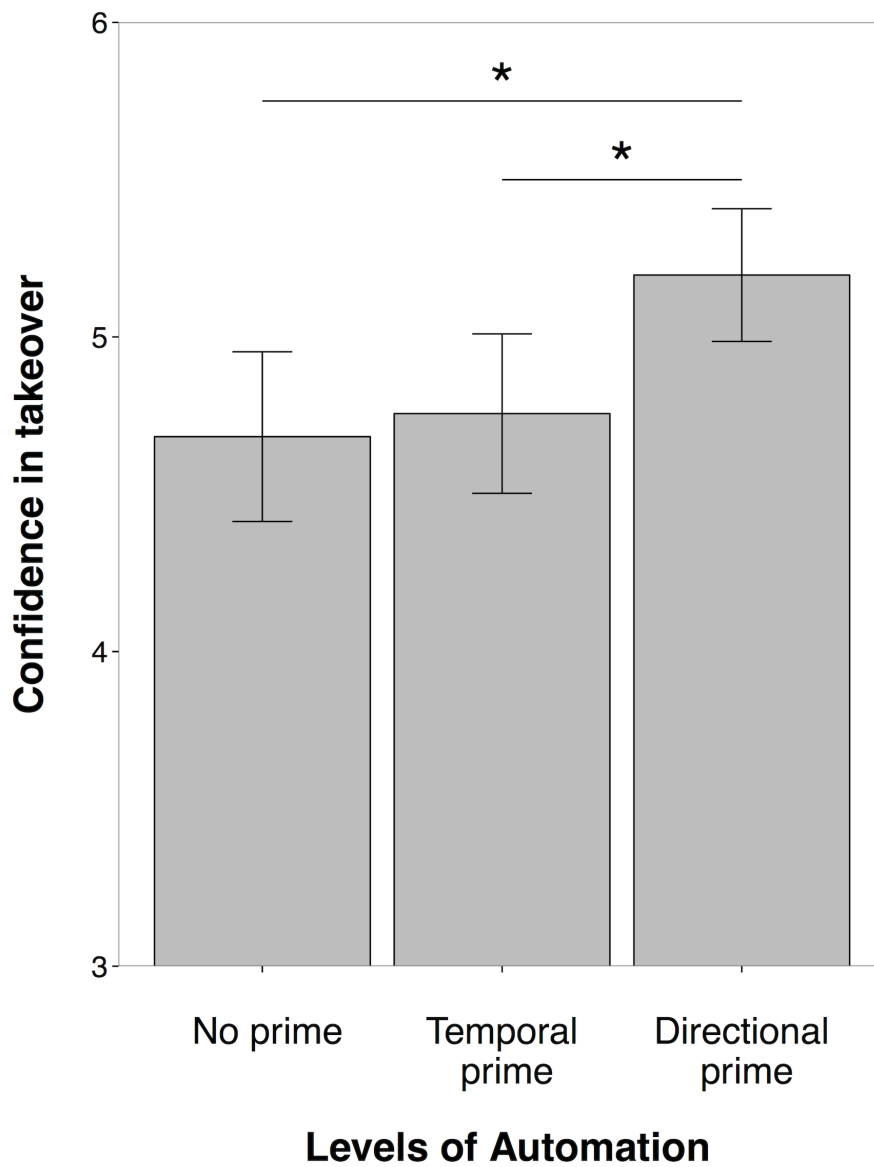


FIGURE V.4 – Mean values for the feeling of being able to take over the system in the three automated conditions. Error bars represent standard errors of the mean.

Discussion

The present results suggest that participants display the strongest temporal estimation bias in the fully manual condition. Assuming that this temporal bias reflects a form of IB, it indicates that the level of automation has an impact on the feeling of agency and also on the IB effect. This result is consistent with previous studies in which IB decreased when participants had to interact with an automation system or a computer (Berberian et al., 2012 ; Obhi & Hall, 2011). However, contrary to our hypothesis, there was no significant difference between the three automated conditions (all $ps > 0.05$). Although this null effect could be due to a weakness in statistical power (with only 36 trials per participant), it could also suggest that participants may have experienced the same feeling of agency in these various automated conditions. Indeed, as proposed by Obhi and Hall (2011), sense of agency while interacting with others is based on our understanding that others have intentions to act that are similar to one's own intentions, which may be more difficult with automation. In their study, they found that IB was present when participants performed a joint action with another human, but not when the other agent was replaced by a computer. Similarly, in our case, assessing the system's intentions through primes does not seem sufficient enough to generate a genuine feeling of agency and the related IB effect.

Providing information about the autopilot's intentions did have an effect on explicit judgements of control. When both the temporal and directional information were provided, participants felt that they would be better at taking over the control of the aircraft. These results suggest that it is certainly appropriate to dissociate the feeling of agency and the judgement of control. Increasing the amount of information about the system's intentions produced stronger explicit judgements of control, but did not influence an implicit measure of sense of agency based on time perception (intentional binding). The results suggest that implicit measures of agency, such as intentional binding, may be

relatively unaffected by predictive information, while explicit judgements of control are strongly influenced by predictability.

Experiment 2

In Experiment 1, we found that participants reported a better confidence in their ability to take over the system when it indicated clearly which action it was about to perform. However, these results were based on subjective ratings. The aim of Experiment 2 was to assess whether such priming also influenced performance. Instead of asking participants to rate their sense of control over the system, they were confronted with real "takeover situations" and we measured their performance in these situations that are critical in aeronautic contexts.

Method

Participants

Twenty-eight undergraduate students from Aix-Marseille University participated in this experiment (mean age = 23.7; SD = 2.45; 17 females). All were naive to the purposes of the manipulation, with normal or corrected-to-normal vision and no expertise in controlling an aircraft. Two participants were excluded from the final analyses because of technical problems and a third participant was also excluded because of misunderstandings in running the experiment.

Material

In this experiment, participants were expected to perform two tasks in parallel : the conflict detection task used in Experiment 1 and a tracking task. The same simulator as in experiment 1 was used for the conflict detection task. Adding a second task increased

working memory load of participants, and therefore made the detection task harder. This situation is also more ecological since, in standard aeronautic situations, human operators usually have more than one screen or one action to supervise and often need to switch quickly and efficiently between various tasks.

Conflict detection task

The participant's task was the same as in Experiment 1 except that only two types of priming conditions were used : the "temporal prime" and the "directional prime" conditions. To reduce the complexity of the experimental design and maximize statistical power, we decided not to use a "no prime" condition, since results in the first experiment indicated an absence of difference with the "temporal prime" condition. To ensure a high confidence in the automation system, participants were told that the system was flawless and that it did not require their assistance or any monitoring unless it generated a conflict detection signal. This signal was presented 13s before a potential conflict. When such a conflict was detected by the autopilot, the system was programmed to make a decision about the new direction to select in order to avoid the conflict. In the same way as in Experiment 1, in the "temporal prime condition", the message "AVOID" appeared on the screen indicating that the plane was about to avoid the obstacle. In the "directional prime condition", the message "AVOID" appeared together with an arrow indicating in advance the direction selected by the system before implementing this command. If the decision chosen by the autopilot was perceived as optimal, participants were instructed to let the autopilot control the aircraft and execute the command. Otherwise, if the decision implemented by the system was perceived as not optimal, participants had to press the "e" key (for "error") as quickly as possible. These situations were relatively rare during the experiment (8 % of the trials) and they simulated a typical case in which a takeover had to be performed. These takeover situations always had the same structure : the autopilot selected a direction that prevented a collision with a first obstacle but this

direction would lead to a collision with another obstacle.

Contrary to Experiment 1, the experiment was organized in continuous blocks during which the aircraft was supposed to follow a predefined trajectory. Each block lasted about seven minutes. In each block, 10 potential conflicts with the surrounding traffic were implemented. For each potential conflict, the autopilot system indicated that it detected an incoming obstacle and it automatically selected a new direction to avoid the conflict. Participants had to judge the autopilot decision and to press the "e" key when they considered that the decision was inappropriate or not optimal. The nonoptimal autopilot decisions would lead to collision with a second obstacle. Failure to detect a nonoptimal decision lead to a red circle was shown surrounding the aircraft. If they correctly press the "e" key, participants were told that their aircraft reduced his altitude to avoid collision and therefore, the red circle did not appear (Fig V.5). There were 5 blocks in each experimental condition (i.e., in the "temporal prime" and the "directional prime" conditions) resulting in a total of 50 conflicts to supervise in each condition. Among these 100 conflicts, 60 were used as fillers (30 per condition). These filler trials were simple situations with few obstacles (1 or 3) leading systematically to a clear and unambiguous decision of the system. These simple conflicts were used to introduce some variability in the task. The 40 remaining conflicts were more complex with three fixed obstacles and two moving aircraft. Among these 40 conflicts (20 per condition), eight required to press the "e" key and to take over the autopilot's decision (four for each prime condition). Therefore, out of 100 conflicts, there were a total of 8 takeover conflicts appearing in a pseudo-random order. The two constraints for the pseudo-randomization were to avoid a takeover at the beginning of a block and to avoid two errors in a row. Finally, the 10 experimental blocks appeared in a random order. At the beginning of each block, the experimenter informed the participant about the type of block she/he was about to perform.

Two types of behavioral measures were recorded in order to evaluate the performance

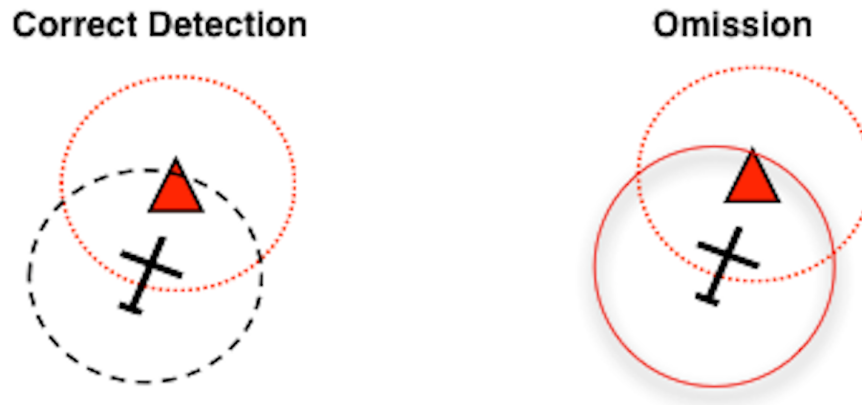


FIGURE V.5 – Description of the visual feedback provided to the participant when the system takes an inappropriate decision (8% of cases). Left : feedback after participant correctly detected the error by pressing the "e" key. Right : feedback displayed if participants did not press the "e" key showing the aircraft dangerously approaching another obstacle. In this case, a red circle surrounded the aircraft.

of participants in these takeover situations. We first recorded the time-to-conflict at the "e" key press, that is, the remaining time before a potential collision with the incoming obstacle. The second measure was the number of times a participant incorrectly pressed the "e" key to signal a system non-optimal decision (i.e., false alarm).

Tracking Task

This task was performed on a second screen, in parallel to the conflict detection task and was similar to the tracking subtask from the Multi-Attribute Task Battery (MATB) (Comstock & Arnegard, 1992 ; Molloy & Parasuraman, 1996). This 2D compensatory tracking task took place above a lake which was represented by a "God's eye view" (Fig. V.6). Participants were instructed to keep a square cursor as close as possible to a moving target (a boat) by using a joystick. The color of the square cursor depended on its position relative to the target in order to provide a visual feedback to the operator : it was green when the cursor was on the target (i.e., when it covered at least 75% of the

target's surface) and red otherwise. The target alternated between stationary position and movements. When the target moved, it was following a predefined straight trajectory. This second task was selected to be coherent with the supervision task and maintain the participant in an aeronautical framework. The two screens were placed side by side, so they can both be monitored without changing posture. Participants were told that if a takeover in the conflict detection task was needed, this task became the priority even if the target moved outside the cursor.

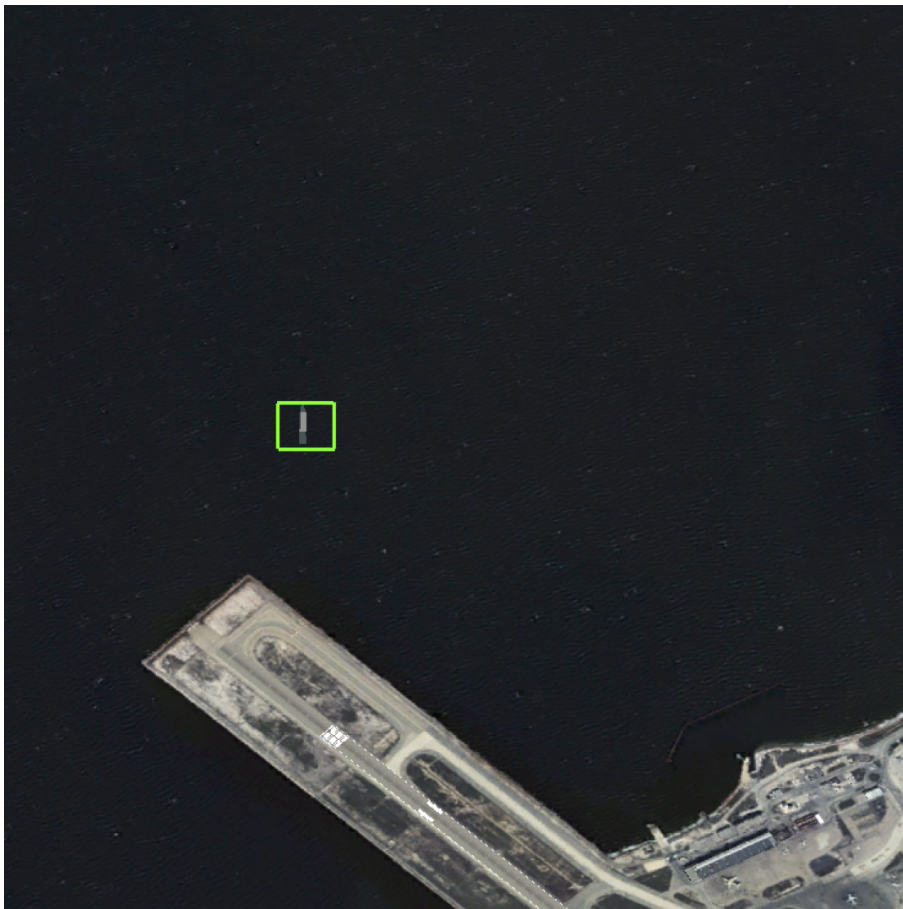


FIGURE V.6 – Representation of the target and the cursor. If the coverage is higher than 75% of the target's surface, the cursor appears in green otherwise it appears in red.

During each block of the conflict detection task, 20 movements of the target were

implemented in the tracking task. Each movement lasted about 8s. The trajectory was random as well as the interval between two movements (i.e., stationary position). The participant's performance during the tracking task was evaluated by measuring the proportion of time on target

Additional measures

At the end of each block participants were asked to evaluate the percentage of time they spent on the conflict detection task by setting a cursor on a scale between 0% and 100%. In addition to this subjective measure, we counted the number of switches of attention between the two computer screens using a head tracker. The aim of these two measures was to evaluate the allocation of attentional resources depending on the experimental condition. We hypothesised that more information about the automation system's intentions would lead to less time spent on the conflict detection task and to better performance in takeover situations.

Results

Conflict detection task

Concerning the delay in pressing the "e" key, there was a significant difference between the two experimental conditions ($t(24) = -2.10$, p -value < 0.05). Participants detected a system error sooner in the directional prime condition than in the temporal prime condition, as revealed by their respective time-to-conflict (3.74 sec and 3.21 sec on average) (higher time-to-conflict meaning a faster response time). Participant made also fewer false alarms ($t(24)=2.2$, p -value < 0.05) in the temporal prime condition (3.6 false alarms) compared to the directional prime condition (2.36 false alarms).

To complement these measures of performance, we computed an index of discriminability (d') for the trials requiring a takeover. However, because many participants did not

produce any miss (i.e., they always pressed the "e" key on critical trials), we artificially generated misses by using the following time threshold : if participants did not press the "e" key after a delay of 2 sec., then the trial was considered as a miss. Based on the rate of hits (correct identification of non-optimal decisions) and the rate of false alarms (incorrect identification of non-optimal decisions), we calculated the sensitivity ($d' = Z(H)-Z(FA)$). A paired t -test revealed a higher sensitivity to conflict when the system's intentions were provided ($d' = 1.6$) than when participants only had temporal information ($d' = 1.2$) ($t(24)=-2.15$, p -value <0.05). Figure V.7 provides a summary of these measures of takeover performance.

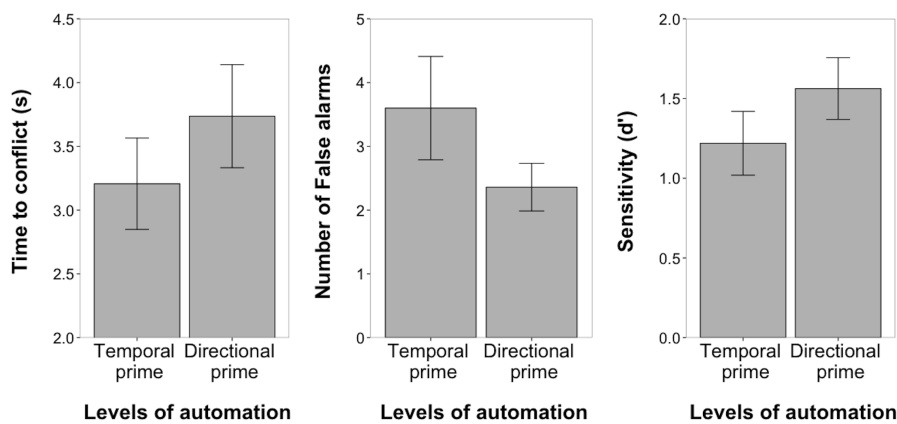


FIGURE V.7 – Left : Mean response times to press the "e" key in the two experimental conditions. Middle : Mean number of false alarms. Right : Mean d' measures of sensitivity. Error bars represent standard errors of the mean.

Tracking task

A paired t -test revealed that there was no significant difference in the time ratio spent on the target during the tracking task between the two experimental conditions ($t(24) = 0.57$, p -value > 0.05). Participants spent 76.94% of the time on the target in the temporal prime condition and 76.57% of the time in the directional prime condition.

Attentional measures

A paired t -test revealed that participants reported spending less time on the supervision task in the directional prime condition (34.9%) compared to the temporal prime condition (49.45%) ($t(24) = 5.25$, p -value < 0.01). Head-tracking measures revealed that when the system's intentions were provided, participants produced less switches between the two tasks ($M_{\text{temporal}} = 213.8$ and $M_{\text{directional}} = 197.6$; $t(24) = 5.78$, p -value < 0.01). Subjective reports were therefore corroborated by an objective measure. Both indicated that providing system's intention leads to a change in allocation of the attentional resources when performing a dual-task.

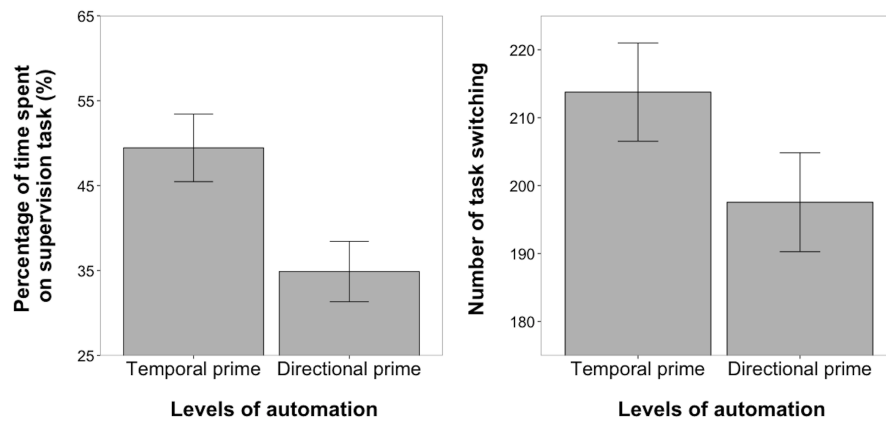


FIGURE V.8 – Left : Percentage of time spent on the tracking task, estimated by participants. Right : Number of task switches per condition. Error bars represent standard error of the mean.

Discussion

The present results indicate that providing information about the system's intentions is helpful for participants to improve their detection of non-optimal decisions from the automation system. We also found that this increase in performance in the conflict detection task is associated with a change in the way participants allocate their attentional

resources in the two tasks. When participants had access to the system's intentions, they both engaged fewer resources in the conflict detection task and had greater performances in this task. Finally, we did not observe any difference in the tracking task, although this might be due to a ceiling effect, the task being too easy to really help discriminating the performances in the two experimental conditions.

General discussion

Four main results were obtained in the present study. Experiment 1 showed that (1) when participants benefit both from temporal and spatial information about the action to be executed by the automated system, they felt more confident in their ability to take it over (i.e., subjective control). However, (2) these conditions did not produce any change in intentional binding relative to the no prime condition. Experiment 2 showed that providing additional directional information about the system's intentions led to both (3) better performance in detecting the system's non-optimal decisions (shorter reaction times and less false alarms; i.e., objective control) and (4) a change in the way participants allocated their attentional resources. The present results indicate that priming human operators with information about the system's intentions improves : 1) the operator's judgement of control ; 2) her/his performance in takeover situations ; and 3) her/his management of attentional resources. Our results in an aeronautic context are in line with the work on (semi-) automated vehicles by Neville Stanton's group that proposed to design automation systems as chatty co-drivers providing continuous relevant feedback to the driver (Eriksson & Stanton, 2015 ; Stanton, Dunoyer, & Leatherland, 2011). This trend of researches follows Norman's (1990) recommendations about the need to provide feedback about the state of the system. It is also consistent with previous studies that have shown a benefit of prime messages (Koo et al., 2015 ; Navarro et al., 2010 ; Navarro, Mars, & Hoc, 2007). Particularly, our data suggest that prime messages

do enable more efficient transfer of control between system and operators in the case of system failure. On this view, OOL phenomena during takeover situations could be largely consequences of system opacity.

How information should be provided by the system to optimize these benefits remains largely unknown. For instance, Koo et al. (2015) demonstrated that the benefits of prime messages depended on the nature and the quantity of information displayed. In a semi-autonomous driving task, they tested different messages that provided an explanation of the car's imminent autonomous action in advance (e.g. auto-braking function). They found that providing the car's future action ("how" : "the car is braking") and the reason for that action ("why" : "obstacle ahead") at the same time increased driving performance but decreased acceptance. Moreover, they found that messages providing only "how" information led to poor driving performance, whereas "why" information was preferred by drivers and led to better driving performance. Clearly, the content (e.g., "how" and "why" information) of the prime message should be explored in future research. In our experiment 2, we noticed a particular benefit of spatial, as opposed to temporal priming. This perhaps indicates a specific requirement to attend to the location of the takeover situation.

In this context, using the theoretical framework of agency might help determine the minimal information necessary for such transfer of control. Our results reveal that displaying the system's intentions prior to an action is a good candidate for maximizing takeover efficiency. Both subjective (takeover confidence) and objective (performance) control increased in the presence of system intention priming. In addition to the increase of control, participants reported spending less time on the supervision task when they received the system's intentions. These results are consistent with Cook's studies, which showed that using interfaces which enabled proactive rather than retroactive monitoring led to earlier problem detection and to an increased operator span of control (Cook, Rieth, & Ngo, 2015 ; Cook, Smallman, & Rieth, 2014). Those studies used a no-

vel trend-icon hybrid indicator to give quantitative information about worsening trends. An experiment contrasting these trend-icons against digital values revealed that these icons supported more frequent proactive detections, and led to a fivefold increase in the number of processing indicators that participants could effectively oversee.

In our data, verbal reports about allocation of attentional resources were corroborated by a head-tracking system. We found a reduced level of switching between the two tasks when spatial and temporal information were provided. Adler and Benbunan-Fich (2012) demonstrated that high frequency of task switching tends to degrade overall performance. Therefore, providing information about system intentions simplifies the supervision task, and leads to less cognitive effort in deciding whether a takeover is required. This in turn allows operators to spend more time on other tasks.

However, contrary to our hypothesis, we did not observe any difference in terms of performance on the tracking task. This could likely be due to a ceiling effect. A more difficult tracking task might generate significant differences between these two conditions. Eye-tracking techniques could also be used in order to provide more sensitive measures and to gain further insight into the allocation of attention in these multi-tasks situations depending on the prime message provided to the operator.

These different results show that models of sense of agency provide a useful framework for thinking about interactions with technology. We assumed that Wegner's (1999) three principles of agency could be made relevant to human-machine interfaces. Wegner claimed that a thought is perceived as willed when the thought precedes the action at a proper interval (called the priority principle), when the thought is compatible with the action (consistency principle) and when the thought is the only apparent cause of the action (exclusivity principle). For example, the consistency principle suggests one possible way that interfaces could compensate for negative effects of latency on action control (Berberian et al., 2013). In the same vein, the priority principle could be relevant to the time at which automated systems transfer information to the operator, particu-

larly in aeronautical situations where there is both a high level of complexity and strong time constraint. Wegner and Wheatley (1999) indeed demonstrated that an increase of the interval between thought and act led to a decrease of sense of agency. They claimed that "thought that occurs too far in advance of an action is not likely to be seen as the cause of it." More recently, Haering and Kiesel (2015) found that sense of agency is not always stronger at shorter action-effect intervals, but rather depends on the matching between the agent's expectations and the actual timing of the events. Therefore, advance warnings have to be displayed neither too soon nor too late, but with precise timing otherwise performance in terms of supervision might not increase (see also Berberian et al., 2012). In this context, we assume that both consistency and priority principles could be helpful for system designers and that the framework of agency could be used to propose guidelines to design more controllable artificial agents.

We also investigated whether priming messages could promote the sense of agency, both at an explicit and implicit level. Priming by the automated system indeed improved participants' explicit judgements of control. Similarly to Wegner et al. (2004), we found that the increase of system predictability through the use of prime messages increases sense of control (see also Sato (2009)).

These results, however, are inconsistent with those reported by Damen, van Baaren, and Dijksterhuis (2014). In that study, participants performed left or right button-presses that generated tones, and were subsequently asked to indicate the degree to which they felt that they, instead of the computer, had caused the tones. Prior to button-presses, participants were subliminally or supraliminally primed with "left" or "right". Participants were free to press one of the two buttons, and thus could perform prime-compatible or prime-incompatible actions. Results showed that compatible subliminal primes increased judgement of agency compared to subliminal incompatible primes. In contrast, supraliminal compatible primes decreased judgements agency compared to incompatible primes. The authors suggested that conscious awareness of the primes might reduce sense of

agency when acting compatibly with an external prime. By analogy, in our task, supraliminal priming with the intention of the system could be perceived as intrusive and lead to a decreased sense of control. In fact, we found the opposite effect (see also Moore & al., 2009). Perhaps, the difference between Damen and colleagues' task and our task may explain this contradiction. In a supervision context, an external agent performs the action in any case. Thus, the use of prime messages might be interpreted as communication rather than a form of intrusion and the use of supraliminal prime messages can create a sense of control.

At the implicit level, a different pattern of results was observed. As in Berberian et al., (2012), we replicated the basic finding that automation decreases binding relative to feeling of agency. However, we observed no effects of prime on the implicit level of agency. Our results revealed a clear difference between action-outcome association (as in the manual condition) and the visual event-outcome association present in the three automated conditions.

The dissociation between explicit and implicit measures could be interpreted via the two-step model of agency described by Synofzik et al. (2008). In our first experiment, participants produced movements in the manual condition but did nothing in the automated conditions. Therefore, our results could be accounted for by the fact that intentional binding is more related to motoric agency, while the explicit component of agency is more related to outcome prediction.

In a recent study, Poonian and Cunnington (2013) found intentional binding for both self-made and observed actions (see also Wohlschläger et al., 2003). In their experiment, participants were required to make a button-press response to indicate the perceived interval between a self-made action and a tone, between a closely matched observed action and tone, and between two tones. For both self-made and observed actions, they found a significant perceived shortening of the interval between the actions and tones as compared to the interval between two tones. Intentional binding was found for both

self-made and observed actions. This result indicates that binding could be obtained even without any movement of the participant. Then, the absence of movement in our experiment does not seem sufficient to explain the absence of intentional binding.

A complementary explanation stems from the field of human-machine interactions. Previous work by Wohlschläger et al. (2003) examined the time-course of the conscious awareness of self, other and machine-generated actions. Their results showed that conscious judgments of self- and other generated actions were similar (same as Poonian & Cunningham, 2013), but both were very different from judgments of actions generated by a machine. They suggested that attribution of intentions to other human agents, but not machines, explained this pattern of action awareness (see also Gentsch & Schütz-Bosbach, 2011). Obhi and Hall (2011) found similar results : an absence of binding when an individual is paired with a computer compared to when the individual is paired with another person. Therefore, we could interpret our result as demonstrating lack of effectively intention attribution with respect to the autopilot.

Wohlschläger et al., (2003) and Obhi and Hall (2011) suggested that the mirror system could be involved when interacting with human partners enabling the formation of a "we-" identity. The mirror system would not form the representation of action plans and intentions when interacting with computer partners. This echoes the growing trend of research promoting the use of more human-like interfaces to improve interactions between human and artificial agents. Previous studies revealed that robots miming the nonverbal behavior of humans (bodily cues, emotional feedback) tend to be more persuasive (Chidambaram, Chiang, & Mutlu, 2012), lead to better learning when used by students (Lester, Towns, & Fitzgerald, 1998) and enabling people to build trust with these artificial agents (DeSteno et al., 2012). Human-like agents can apparently support intention attribution, leading to more engagement during interactions (Gratch, Hill, Morency, Pynadath, & Traum, 2015). We hypothesize that using such human-like interfaces could facilitate the formation of this "we-unit" and lead to an intentional binding effect, even

in the case of automated systems. However, the inability to create a new unit with an artificial system does not prevent other possible increases in the efficiency of supervising the system. It would be interesting to test if a system that supported intention attribution could further improve performance in a supervision task.

Acknowledgements

We thank the "Direction Générale de l'Armement" (DGA) for their financial support to the first author.

Seconde Etude

Agency modulates interactions with automation technologies

Kevin Le Goff^{1,2}, Arnaud Rey², Patrick Haggard³, Olivier Oullier², Bruno Berberian¹

¹ Systems Control and Flight Dynamics Department, ONERA, Salon de Provence, France

² Laboratoire de Psychologie Cognitive & CNRS, Aix-Marseille University, Marseille, France

³ Institute of Cognitive Neuroscience, University College London, United Kingdom

Cette étude est actuellement en révisions pour la revue Ergonomics.

Bien qu'aucun effet des messages amorces n'eut été constaté sur la mesure implicite d'agentivité (i.e., liage intentionnel), notre première étude a permis de confirmer un impact positif des messages indiquant les futures actions du système automatisé que ce soit en termes de jugement de contrôle ou au niveau de la performance des participants en situation de reprise en main. Ainsi, le premier objectif de cette seconde étude était de répliquer les résultats obtenus précédemment sur le contrôle effectif de l'opérateur tout en étudiant l'évolution de la gestion des ressources attentionnelles. Le second objectif était d'étudier l'impact de ces messages amorces sur le niveau d'acceptabilité du système par les opérateurs. Pour y parvenir, les participants devaient toujours réaliser deux tâches en parallèle. La tâche de supervision restait la même mais la tâche annexe correspondait cette fois à la *Multi-Attribute Task Battery* souvent utilisée dans le secteur aéronautique.

Abstract

The increasing presence of automation between operators and automated systems tends to disrupt operators from action outcomes leading them to leave the control loop. The theoretical framework of agency suggests that priming the operator about the system's upcoming behavior could help restoring an appropriate sense of control and increase the user acceptance of what the system is doing. Using an aircraft supervision task, participants (N=24) were to take over the automated system if it offered a non-optimal course of actions. Two automation systems were compared : one providing both temporal and directional information and one providing only temporal information about the action to be done. Our results showed that providing directional information about the system's upcoming behavior led to (1) better performances in detecting the system's non-optimal decisions (i.e., shorter reaction times), (2) to increase the level of user acceptance and (3) to better allocation of the operator's attentional resources.

Introduction

Recently, the concept of agency has been applied to the HCI domain (Berberian et al., 2013, 2012 ; Limerick et al., 2014 ; McEneaney, 2009, 2013 ; Obhi & Hall, 2011). During the interactions with technology, it has been shown that the simple process of producing an action to cause an intended outcome depends upon several variables that can alter the agentic experience dramatically (Limerick et al., 2014). For instance, when tweaking the level of automation in an aircraft supervision task, Berberian et al. (2012) found a decrease in agency (for both implicit and explicit measures) concomitant with the increase in automation. They argued that the interposition of increasing automation between operators and automated systems tends to distract operators from action outcomes, decrease their sense of control and therefore disrupt their overall performance. This echoes Baron (1988), who claimed : "Perhaps the major human factors concern of pilots in regard to introduction of automation is that, in some circumstances, operations with such aids may leave the critical question, who is in control now, the human or the machine?"

This ambiguity about who is in control has also an impact on user acceptance. Although performance and acceptance are often positively correlated (Nielsen & Levy, 1994), high levels of performance do not guarantee user acceptance. Indeed, users tend to reject systems that enhance their performance in favour of systems that are less efficient but more acceptable. For instance, Inagaki, Itoh, and Nagai (2007) showed that drivers preferred collision warnings over automated control, which tends to be misunderstood, even when automated control led to better performance. If the question of what we should automate remains (for a discussion of this point see Hancock (2014)), a major challenge in the HCI community, with the next generations of highly automated systems, is precisely to determine how to compensate for this decrease in the sense of control and the user acceptance.

These negative impacts of automation are well-known by the HCI community and

different solutions have been proposed. Some of them consist in training human operator to produce efficient behavior in case of system failure. However, recent dramatic events indicate that such training does not ensure efficient takeover for trained situations, whereas the apparition of unexpected failure are not considered by such approach. Moreover, a recent study demonstrated that operator errors resulting from automation bias were much higher when automation misdiagnosed a fault than when it missed one (Sauer, Chavaillaz, & Wastell, 2015). This result shows that complacency cannot be entirely overcome with training. Other solutions propose to manipulate the level of system automation, sharing the authority between the automation and the human operator (for example MABA-MABA methods, adaptive function allocation). Such approach rests on the hypothesis that new technologies can be introduced as a simple substitution of machines for people - preserving the basic system while improving it on some output measures. Unfortunately, such assumption corresponds to a vague and bleak reflection of the real impact of automation : automation technology transforms human practice and forces people to adapt their skills and routines (Dekker & Woods, 2002).

A better understanding of the role played by agency may therefore provide a useful framework for thinking about interactions with automated technology. Notably, three principles of agency proposed by Wegner and Wheatley (1999) are relevant to improve human-machine interfaces. These authors claim that an action is perceived as willed 1) when the thought precedes the action at a proper interval (called the priority principle), 2) when the thought is compatible with the action (consistency principle) and 3) when the thought is the only apparent cause of the action (exclusivity principle). Following that theoretical proposition, Wegner, Sparrow, and Winerman (2004) reported that priming effects (i.e., providing predictive information of what will happen next) is a good way to influence or simulate prior thoughts, and therefore creates a sense of agency even in the absence of any movement (see also Moore, Wegner, and Haggard 2009; Sato 2009). In their study, participants watched themselves in a mirror while another person

behind them, hidden from their view, moved their hands forward on each side where participants' hands would normally appear, and performed a series of movements. When participants could hear instructions previewing the movements, they experienced a higher degree of agency for these movements. This priming effect could then be used to improve communication between automation systems and operators to increase the feeling of control in a supervision context.

The aim of the current research was precisely to test whether providing an information about what the system is about to do next leads to an increase in the level of user acceptance, concomitant with an increase in control and performance. In the present experiment, participants had to perform an aircraft supervision task together with a secondary target-tracking task. This secondary task was used to increase the participant's mental workload (Young, Brookhuis, Wickens, & Hancock, 2015) and to create a more ecological aeronautic situation. In the main supervision task, participants simply had to check that the autopilot was correctly controlling the aircraft when encountering obstacles. In some rare occasions, participants had to take over the automated system when a non-optimal action was proposed by the autopilot. Two automation systems were used, one providing both temporal and directional information about the action the autopilot was about to engage, and one providing only temporal information. To appraise the quality of the interactions with the automated system, we used a combination of objective measures with subjective ratings (Annett, 2002). Moreover, performance on a secondary task is closely associated with the spare capacity unused by the primary task (Young et al., 2015). Effective control and performance were measured by reaction times in taking over the system when it took a non-optimal decision. Ratings of user acceptance were also assessed in both conditions. Finally, a portable eye-tracking system was used to monitor the way participants allocated visual attention resources during both tasks (i.e., supervision and tracking). This last measure provides additional information about the operator's performance and the smoothness of its interactions with the automated

system.

Method

Participants

Twenty-four undergraduate students from Aix-Marseille University with normal or corrected-to-normal vision participated in this experiment (mean age = 24.5; SD = 5.1; 15 females). All participants were naive to the purposes of the study, and had no particular expertise in aircraft control.

Material

This experiment was conducted using ONERA's airplane simulator (LabSim, ONERA, French Aerospace Lab, Salon de Provence). Participants were expected to perform two tasks in parallel : an aircraft supervision task and a tracking task (see Figure VI.1). Adding a second task increased working memory load of participants, and therefore made the supervision task harder. This situation is also more ecological since, in standard aeronautic situations, human operators usually have more than one screen or one action to supervise and often need to switch quickly and efficiently between various tasks.

Supervision task

The simulator was composed of a navigation display representing the aircraft in the horizontal plane together with the surrounding traffic (see Figure VI.2). Participants had to supervise the progress of an aircraft on a predefined flight path, through two conditions : a "temporal priming condition" and a "directional priming condition". At the beginning of each trial, participants were informed about the condition. After a 3-second delay, an obstacle (another aircraft) appeared on their path. A message "CONFLICT AHEAD" appeared on the screen to warn the participant that this obstacle could pene-



FIGURE VI.1 – Schematic Representation of the Apparatus during the Experiment. The Supervision task is displayed on the left screen and the Tracking task on the right screen.

trate the safety envelope of the plane if no change in its direction were to occur. In both conditions, the autopilot system automatically detected and selected a new direction to avoid the potential conflict without any intervention of the participant. The two automated conditions were defined by the degree of information displayed by the autopilot to participants when a conflict was detected. In the 'temporal priming condition', when the autopilot selected a new direction, the message "AVOID" appeared on the screen indicating that the plane was about to avoid the obstacle. In the "directional priming condition", the message "AVOID" appeared together with an arrow and a mark on the large white circle representing the angular indication. The arrow and the mark indicated the direction selected by the system before implementing this command. The "temporal prime condition" was used to evaluate the effect of providing temporal information about the command to be executed. By adding angular indications in the "directional

prime condition", it was possible to test the effect of providing information about the subsequent action. Note that primes were always predictive of the action.

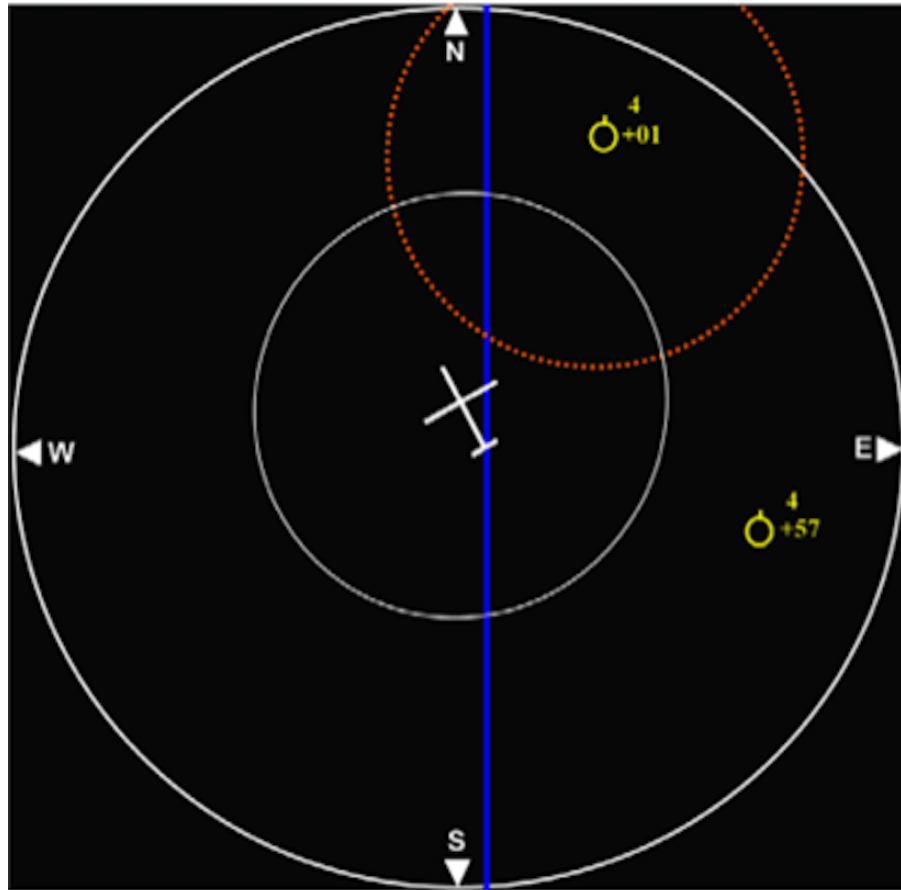


FIGURE VI.2 – Experimental set up with the navigation display : the large white circle represents the angular information. The smaller white circle surrounding the plane represents the aircraft safety envelope. A conflict appears when an obstacle (yellow circle) penetrates the safety envelope leading to the appearance of a red circle surrounding the obstacle.

If the decision chosen by the autopilot was perceived as optimal, participants were instructed to let the autopilot control the aircraft and execute the command. Otherwise, participants had to press the "e" key (for "error") as quickly as possible. These situations were relatively rare during the experiment (8% of the trials) and they simulated a typical

case in which a takeover had to be performed. These takeover situations always had the same structure : the autopilot selected a direction that prevented a collision with a first obstacle but this direction led to a collision with another obstacle. Regardless the condition and if a takeover was required, a green message "CLEAR" plus a sound indicated to the participant the end of a trial. The inter-trial interval lasted for 10 seconds.

The experiment was organized in continuous blocks during which the aircraft was supposed to follow a predefined trajectory. Each block lasted for about seven minutes. In each block, 10 potential conflicts with the surrounding traffic were implemented. For each potential conflict, the autopilot system indicated that it detected an incoming obstacle and it selected automatically a new direction to avoid the conflict. Participants had to judge the autopilot decision and to press the "e" key when they considered that the decision was inappropriate or not optimal. The non-optimal autopilot decisions would lead to collision with a second obstacle. Failure to detect a non-optimal decision lead to a red circle surrounding the aircraft. If they correctly pressed the "e" key, participants were told that their aircraft reduced its altitude to avoid collision and therefore, the red circle did not appear. There were 5 blocks in each experimental condition (i.e., in the "temporal priming" and the "directional priming" conditions) resulting in a total of 50 potential conflicts to supervise in each condition. Among these 100 potential conflicts, 60 were used as fillers (30 per condition). These filler trials were simple situations with few obstacles (1 or 3) leading systematically to a clear and unambiguous decision of the system. These simple conflicts were used to introduce some variability in the task. The 40 remaining potential conflicts were more complex with three fixed obstacles and two moving aircrafts. Among these 40 potential conflicts (20 per condition), eight required pressing the "e" key and taking over the autopilot's decision (four for each prime condition). Therefore, out of 100 conflicts, there were a total of eight takeover conflicts appearing in a pseudo-random order. The two constraints for the pseudo-randomization were to avoid an error at the beginning of a block and to avoid two errors in a row. Participants were exposed to each

condition alternately. Half of them began with the "temporal priming" condition and the other half with the "directional priming" condition. At the beginning of each block, the experimenter informed the participant about the type of blocks (s)he was about to perform.

The time elapsed between the conflict detection signal (indicating the presence of an obstacle) and the "e" key pressing was recorded (i.e., reaction time) in order to evaluate the performance of participants in these takeover situations.

Tracking Task

This task was performed on a second screen, at the same time as the supervision task and was similar to the tracking subtask from the Multi-Attribute Task Battery (MATB) (Comstock & Arnegard, 1992 ; Molloy & Parasuraman, 1996). This 2D compensatory tracking task took place above a lake which was represented by a "God's eye view" (Figure VI.3). Participants were instructed to keep the square cursor as close as possible to a moving target (a boat) by using a joystick. The color of the square cursor depended on its position relative to the target, in order to provide a visual feedback to the operator. This feedback was green when the cursor was on the target (i.e., when it covered at least 75% of the target's surface) and red otherwise. The target alternated between stationary position and movements. When the target moved, it was following a predefined straight trajectory. This second task was selected to be coherent with the supervision task and maintain the participant in an aeronautical framework. The two screens were positioned next to each other, so both could be monitored without changing posture. Participants were told that if a takeover in the supervision task was needed, this task became the priority even if the target moved outside the cursor.

During each block of the supervision task, 20 movements of the target were implemented in the tracking task. Each movement lasted about 8s. The trajectory was random as well as the interval between two movements (i.e., stationary position). Participant's



FIGURE VI.3 – Representation of the target and the cursor. If the coverage is higher than 75% of the target's surface, the cursor appears in green otherwise it appears in red.

performance during the tracking task was evaluated by a time ratio corresponding to the time spent "on the target" (i.e., when the cursor was green) over the "total time".

Additional measures

Participants wore eye-tracking glasses (SensoMotoric Instruments© Eye Tracking Glasses 2GW ; Sampling rate : 120Hz) to record gaze direction and fixation time. These measures served to assess the allocation of visual attention on each screen (and therefore each task) as well as the number of visual/attentional switches between the two computer

screens. The aim of these two measures was to evaluate the allocation of attentional resources in experimental conditions.

At the end of the experiment, the level of user acceptance of each automated system (i.e., the automated system with the temporal priming and the one with the directional priming) was assessed by asking participants to answer seven questions (voir la partie Matériel et Méthodes de cette thèse) on a 7-point Likert-type scale (1 = totally disagree, 7 = totally agree). Answers to these questions were averaged to form a reliable measure of user acceptance (Cronbach's Alpha = .87). Higher scores reflect more user acceptance of the automated system.

Results

The results concerning measures of response time in situations requiring a take-over and measures on the time spent on the tracking task are reported in Figure VI.4. Regarding the delay in pressing the "e" key, there was a significant difference between the two experimental conditions ($t(23) = 2.23, p < 0.05$). Participants detected a system error sooner in the directional prime condition (6.9 sec. on average) than in the temporal prime condition (7.6 sec; Figure VI.4, left panel).

Eye-tracking measures revealed that when information on the system's upcoming behavior was provided, participants tended to produce fewer switches between the two tasks ($M_{\text{temporal}} = 11.9$ and $M_{\text{directional}} = 11.4; t(23) = 1.8, p < 0.08$). This smaller number of task switches is concomitant with an increase in time spent on the tracking task (Figure VI.4, right panel). Participants spent more time on average paying attention to the tracking task during a trial in the directional prime condition (16186ms) compared to the temporal prime condition (15769ms) ($t(23) = 2.44, p < 0.05$). However, this difference in time spent on the tracking task did not lead to an increase in performance. A paired t -test revealed that there was no significant difference in the time ratio spent on the target

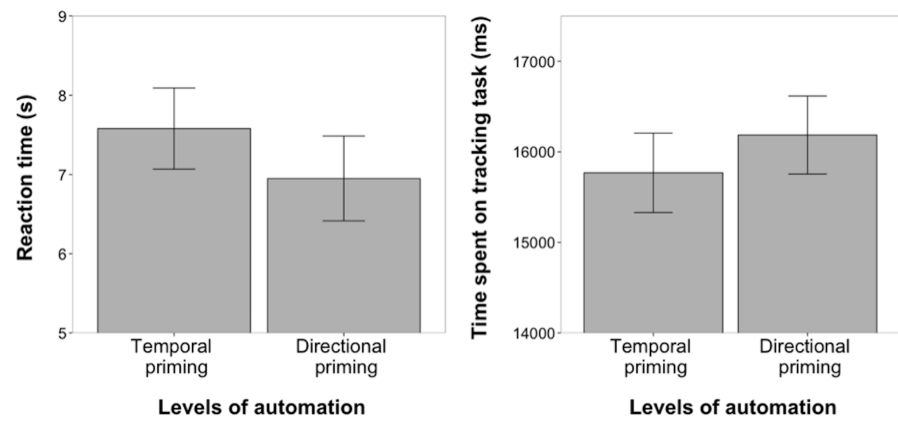


FIGURE VI.4 – Left : Mean response times to press the "e" key in the two experimental conditions. Right : Mean time spent on the tracking task during a trial. Error bars represent standard errors of the mean.

during the tracking task between the two experimental conditions ($t(23) = 0.39, p > 0.05$). Participants spent 77.98% of the time on the target in the temporal prime condition and 77.73% of the time in the directional prime condition. This absence of difference in the two experimental conditions might be due to a ceiling effect, the task being too easy to really help discriminating performances in the two conditions.

As hypothesized, a paired t -test on the acceptance scale revealed that participants rated the automated system that provided information about the direction of the action more acceptable ($M=5.28, SD=0.83$) than the automated system that only provided temporal information ($M=4.66, SD=0.99$) ($t(23)= 2.46, p < 0.05$). Particularly, the first system is evaluated more positively through items related to "trust dimension" (*I found this system reliable ; I trust this system*), "control dimension" (*This system increase my performance during the supervision task ; Using this system increases the safety level*) and perceived ease of use (*It is difficult for me to understand the system's behavior ; Using this system does not require high attentional resources ; Using this system is stressful*). These dimensions are known to be fundamental in user acceptance (Ghazizadeh, Lee, & Boyle, 2012 ; Lee & Moray, 1994 ; Parasuraman & Wickens, 2008 ; Venkatesh, 2000).

Discussion

In the present study, we tested whether providing an information about what the system is about to do next can compensate for the decrease in the sense of control and the user acceptance traditionally observed when operators are confronted to high levels of automation. Our results showed that providing additional direction information about the system's intentions led to (1) better performance in detecting the system's non-optimal decisions (i.e., shorter reaction times), (2) an increase of the operator's level of user acceptance and (3) a better allocation of attentional resources.

These results are in line with Norman's (1990) recommendations about the need to provide feedback on the state of the system, and previous studies that have shown a benefit of prime messages in the HCI domain (Deroo, Hoc, & Mars, 2012 ; Koo et al., 2015 ; Navarro et al., 2010, 2007). In this study, we go further in understanding the impact of such a prime message by assessing its benefit in a more ecological supervisory task, which is the main interaction mode for current technological systems, and frequently leads to an out-of-the-loop performance problem. The present data suggests that priming allows for a more efficient transfer of control from the system to the operators in case of system failure. More specifically, providing information about the nature of the action that the system is about to perform increased the operator's ability to stay in the control loop. This promising result obtained in an ecological aeronautic situation suggests that the theoretical framework of agency can help developing new interfaces between automated systems and human operators.

However, providing information about the system's upcoming behavior is certainly not sufficient. Such information may help, but could hurt, the operator's performance or user acceptance if too much, or inadequate, information is given at an inappropriate moment. For instance, Koo et al. (2015) demonstrated that the benefits of prime messages depended on the nature and the quantity of information displayed. In a semi-autonomous

driving task, they tested different messages that provided an explanation of the car's imminent autonomous action in advance (e.g., auto-braking function). They found that describing the car's future action ("how" : "the car is braking") and the reason for that action ("why" : "obstacle ahead") at the same time increased driving performance but decreased user acceptance. Similarly, Navarro et al. (2010) also observed dissociation between performance and acceptance. In a driving simulation task, they found that motor priming assistance (consisting of directional steering wheel vibrations) was more efficient compared to more traditional auditory (lateralized sound) warning devices. However, participants evaluated classic auditory signal more acceptable than the motor priming system.

In this context, using the theoretical framework of agency might help determine the relevant information necessary for both an efficient transfer of control and an appropriate level of acceptability. Our results showed that providing the system's intentions increases both performance and user acceptance. Particularly, results on performance are corroborated by a more positive evaluation of this system through items relating to the control dimension and perceived ease of use. However, there is a major difference between our study and Koo et al. and Navarro et al.'s work. Hoc, Young, and Blosseville (2009) proposed a four-level classification of human-machine cooperation modes : a perception mode, a mutual control mode, a function delegation mode and a fully automatic mode. According to this classification, the motor priming assistance developed by Navarro et al.'s corresponds to the mutual control mode (level 2). The auto-braking function provided by Koo et al.'s corresponds to the delegation function mode (level 3). In our experiment, participants only had to supervise the autopilot and to perform an action if the decision of the system was evaluated as non-optimal. This corresponds to the fully automatic mode (level 4). The benefit of providing priming information on performance and acceptance may therefore depend on the level of automation. In a supervision context of a fully automatic system, the use of prime messages might be interpreted as

communication rather than a form of intrusion. This result tends to indicate that to be fully efficient, interactions with automation have to be conceived by designers in accord with the degree of automation they are creating.

The interplay of increasing automation between operators and automated system tends to create a distance between operators and action outcomes, decreasing their feeling of control and the user acceptance of the system. Using prime messages allows the operators to follow the current goal of the automation and to check if their own goals are similar to those of the system. Verberne et al. (2012) found that sharing goals between automation and drivers led to increased trustworthiness and user acceptance of the systems. They argued that people respond to technology in a social manner, similarly to how people respond to other people (Reeves & Nass, 1996). However, in a supervision context, the system carries out all actions without any human intervention. The human mainly monitors the system and intervenes if necessary (Endsley, 1999). In that context, it is difficult for the operators to precisely follow the current goal of the system. This leads sometimes to "automation surprises" (Sarter & Woods, 1995 ; Sarter et al., 1997) occurring when automation behaves in ways that the operators do not expect. Recent aircraft accidents have been related to automation surprises, such as the Turkish Airlines TK1951 crash and the Air France AF447 crash, both in 2009 (DSB 2010 ; BEA 2012). Giving information about the system's intentions in the form of prime messages could be a solution to make the system a more collaborative agent.

Our results also show that providing the system's intentions do have an effect on the allocation of attention in a multi-task situation. Whilst many pilots agree that modern aircraft are easier to fly, closer inspection suggests that their workload has shifted, rather than reduced, so pilots now have to carry out more supervisory tasks (Baxter, Besnard, & Riley, 2007). Particularly, pilots can experience lack of knowledge on the current and future status of the autopilot (Rudisill, 1995 ; Sarter & Woods, 1995). This could lead pilots to spend more time trying to understand the aircraft's behaviour rather than

focusing their efforts on the primary task of piloting the aircraft. Our results indicate that making the system more transparent and more predictive reduces the attentional load.

In the present study, we found a reduced level of switching between the two tasks when spatial and temporal information were provided. This result is in line with the findings of Adler and Benbunan-Fich (2012) showing that high frequency of task switching tends to degrade overall performance. We also found an increase in the time spent on the tracking task together with a decrease in the time spent on the supervision task. The average time spent on each area of interest is traditionally considered as an indication of the difficulty to extract and interpret information from this area (Fitts et al., 1950 ; Jacob & Karn, 2003). Therefore, providing information about the system's intentions simplifies the supervision task by making the task more automatic, and leads to less cognitive effort in deciding whether a takeover is required, even if prime messages are additional information to process. This is consistent with the attentional resources approach of mental workload theory, as automatic processing releases attentional resources for other tasks, with a resulting decrease in mental workload (Young et al., 2015).

The fact that we did not observe any difference in the tracking task is certainly due to a ceiling effect, the task being too easy to really help discriminating the performances in the two experimental conditions. A more difficult tracking task or a supplementary task might be necessary to observe the real benefit of prime messages on the ability to perform multitasking.

Further studies are required to investigate if this decrease in attentional load could prevent operators of experiencing mental underload. As noted by Young et al. (2015), there is now a strong consensus that mental underload can be detrimental to performance just as it is to mental overload (Desmond & Hoyes, 1996 ; Hancock & Verwey, 1997 ; Young & Stanton, 2007). This underload could be a by-product of an overtrust on the autopilot. Studies have shown that what is required is an appropriate level of

trust (Lee & See, 2004 ; Parasuraman, Molloy, & Hilburn, 1993). More recently, Payre, Cestac, and Delhomme (2016) have shown in a driving simulator task, that participants with higher trust in automation had longer reaction times to recover a manual control and suggested that an appropriate training is required to overcome such problems. Although our participants evaluated the system more positively through items relating to trust dimension, we have to check that providing the system's intentions could not lead to an inappropriate level of trust or mental workload. In conclusion, this study showed how the framework of agency can be used to design more acceptable and more controllable automated systems. As suggested by this framework, providing more information about the system's intentions led to better performance in detecting when the system's decisions were non-optimal, a greater user acceptance and also a change in the way participants allocated their attentional resources. The present promising results should be complemented in future research by further studying the content of the prime message, the amount of information displayed and the timing of presentation in order to find optimal solutions to improve both performance and the sense of "we-agency" of the operator toward the automated system (Obhi & Hall, 2011 ; Pacherie, 2011).

Acknowledgements

We thank the "Direction Générale de l'Armement" (DGA) for its financial support to the first author and Arnd Rose (SensoMotoric Instruments) for his technical assistance with the eye-tracking technology.

Troisième Etude

Vicarious agency in human-machine interaction : a matter of time

Kevin Le Goff^{1,2}, Arnaud Rey², Patrick Haggard³, Bruno Berberian¹

¹ Systems Control and Flight Dynamics Department, ONERA, Salon de Provence, France

² Laboratoire de Psychologie Cognitive & CNRS, Aix-Marseille University, Marseille, France

³ Institute of Cognitive Neuroscience, University College London, United Kingdom

Cette étude est actuellement en cours de préparation

La seconde étude nous a confirmé la plus-value en termes d'acceptabilité et de performance de l'utilisation des messages amorces. Ainsi, toujours dans la démarche de prouver l'intérêt de l'utilisation du cadre théorique de l'agentivité pour améliorer les interactions homme-machine, nous voulions l'objectif de cette dernière étude était de montrer comment les principes proposés par Wegner et Wheatley (1999) pouvaient nous servir dans ce sens. Cette troisième étude avait donc pour but de vérifier si nous observions toujours la relation quadratique observée par Wegner et Wheatley entre la temporalité des messages amorces et le sentiment de contrôle dans le cadre de la supervision de systèmes. Le second objectif de cette étude était d'étudier si la temporalité avait un impact similaire sur l'acceptabilité des opérateurs et sur leur performance.

Abstract

The increasing presence of automation between operators and automated systems tends to disrupt operators from action outcomes leading them to leave the control loop. The theoretical framework of agency suggests that priming the operator about the system's upcoming behavior could help restoring an appropriate sense of control and increase the user acceptance of what the system is doing. In the present study, using an aircraft supervision task, we tested if priming the automated system's intentions at different time intervals had a similar impact (quadratic or linear trend) on performance, sense of control and user acceptance. Our results showed that the delay between prime messages and the system's actions had a linear impact on the operator's performance (in detecting an error or in the time spent on a secondary task) and a quadratic impact on the operator's levels of control and acceptance.

Introduction

Agency refers to the sense of controlling one's own actions and, through these actions, events in the outside world (Chambon et al., 2015). This form of self-awareness is important not only for motor control but also for social interactions and the ascription of causal responsibility, and serves as a key motivational force for human behaviour and coordination dynamics (Kelso, 2016).

Recently, the concept of agency has been applied to the Human-Computer Interaction (HCI) domain (Berberian et al., 2013, 2012 ; Le Goff, Rey, & Berberian, 2015 ; Limerick et al., 2014 ; McEneaney, 2009, 2013 ; Obhi & Hall, 2011). During the interactions with technology, it has been shown that the simple process of producing an action to cause an intended outcome depends upon several variables that can alter the agentic experience dramatically (Berberian et al., 2012 ; Limerick et al., 2014). In particular, automated systems can generate interesting ambiguities of agency : who is in control, the operator or the system? In some cases, this point can be quite unclear, leading to what is referred to as "operator confusion", poor performances and potential dangers depending on context. This ambiguity about who is in control has also an impact on user acceptance (Baron, 1988). Indeed, users reject systems which tends to be misunderstood, in favour of systems that are less efficient but more acceptable (Inagaki & Nagai, 2007). With the development of automation in aviation and self-driving cars, understanding and overcoming these difficulties remains a crucial challenge in the HCI domain.

A better understanding of the role played by agency may therefore provide a useful framework for thinking about interactions with automated technology. Notably, three principles of agency proposed by Wegner and Wheatley (1999) are relevant to improve human-machine interfaces. These authors claim that an action is perceived as willed 1) when the thought precedes the action at a proper interval (called the *priority* principle), 2) when the thought is compatible with the reaction (*consistency* principle) and 3) when

the thought is the only apparent cause of the action (*exclusivity* principle) (Wegner & Wheatley, 1999). Following that theoretical proposition, they created an experiment to demonstrate that people can experience control of events that they do not in fact cause. A subject and a confederate simultaneously used a single mouse to control the position of a pointer on a screen. Subject heard through headphones the name of an object on the screen shortly before the pointer stopped near that object, then the subject frequently believed that he had intentionally moved towards the object even though in reality his arm had been moved passively by the confederate. This was the first experiment showing that priming effects (i.e., providing predictive information of what will happen next) is a good way to influence or simulate prior thoughts and therefore creates a sense of agency.

Wegner, Sparrow and Winerman (2004) reported that this priming effects could also raise a sense of agency even in the absence of any movement (see also Sato, 2009). In their study, participants watched themselves in a mirror while another person behind them, hidden from their view, moved their hands forward on each side where participants' hands would normally appear, and performed a series of movements. When participants could hear instructions previewing the movements, they experienced a higher degree of agency for these movements. Therefore, when the action is clearly occurring to another person's body, for example, apparent mental causation might still prompt some experience of agency. Wegner et al. (2004) defined this feeling of authorship for the actions of others as *vicarious* agency.

Previous studies already demonstrated the benefit of prime messages while supervising automation (Le Goff, Rey, Haggard, & Berberian, 2016) and more generally in the HCI domain (Koo et al., 2015 ; Navarro et al., 2010, 2007). However, providing information about the system's upcoming behavior is certainly not sufficient. Such information may help, but could also hurt the user's acceptance if too much, or inadequate, information is given at an inappropriate moment (Koo et al., 2015 ; Navarro et al., 2007). The question of timing appears to have an impact on the strength of the agency felt by the

participant as demonstrated by the quadratic curve obtained by Wegner and Wheatley (1999). To be optimal, these prior thoughts should occur neither too far nor too close of an action. Interestingly, supervising automation might be seen as a particular case of vicarious agency in which "other" is an automated system. In the present study, we tested if priming the automated system's intentions at different time intervals had a similar impact (quadratic or linear trend) on performance, sense of control and user acceptance.

Method

Participants

Sixteen participants with normal or corrected-to-normal vision participated in this experiment (mean age = 27.07 ; SD = 4.8 ; 7 females). All participants were naive to the purposes of the study, and had no particular expertise in aircraft control.

Material

This experiment was conducted using ONERA's airplane simulator (LabSim, ONERA, French Aerospace Lab, Salon de Provence). Participants were expected to perform two tasks in parallel : an aircraft supervision task and the Multi-Attribute Task Battery II (MATB, Santiago-Espada et al., 2011). Adding a second task increased working memory load of participants, and therefore made the supervision task harder. This situation is also more ecological since, in standard aeronautic situations, human operators usually have more than one screen or one action to supervise and often need to switch quickly and efficiently between various tasks.

Supervision task

The simulator was composed of a navigation display representing the aircraft in the horizontal plane together with the surrounding traffic (see Figure VII.1). Participants

had to supervise the progress of an aircraft on a predefined flight path. The autopilot system automatically detected and selected a new direction to avoid a potential conflict (i.e., another aircraft) without any intervention of the participant anytime an obstacle was about to enter the aircraft safety envelope. Before implementing a new path, the message "AVOID" appeared together with an arrow and a mark on the large white circle representing the angular indication. The arrow and the mark indicated the future direction selected by the system. This prime message appeared either at 13s, 7s, 4s or 1s before the implementation of the new direction, defining the four experimental conditions. Note that primes were always predictive of the action.

If the decision chosen by the autopilot was perceived as optimal, participants were instructed to let the autopilot control the aircraft and execute the command. Otherwise, when they considered that the decision was inappropriate or not optimal, participants had to press the "e" key (for "error") as quickly as possible. These critical situations always had the same structure : the autopilot selected a direction that prevented a collision with a first obstacle but this direction led to a collision with another obstacle. Failure to detect a non-optimal decision lead to a red circle surrounding the aircraft (see Figure VII.2). If they correctly pressed the "e" key, participants were told that the aircraft reduced its altitude to avoid collision and therefore, the red circle did not appear. Regardless of the four conditions and if a takeover was required, a green message "CLEAR" plus a sound indicated to the participant the end of a trial. The inter-trial interval lasted 10s.

The experiment was organized in continuous blocks during which the aircraft was supposed to follow a predefined trajectory. Each block lasted about six minutes. In each block, seven potential conflicts with the surrounding traffic were implemented. There were three blocks in each experimental condition (i.e., 13s, 7s, 4s, 1s conditions) resulting in a total of 21 potential conflicts to supervise in each condition. Among these 84 potential conflicts, 12 required pressing the "e" key to report a non-optimal decision from the autopilot (three for each prime condition). Participants were exposed to each condition

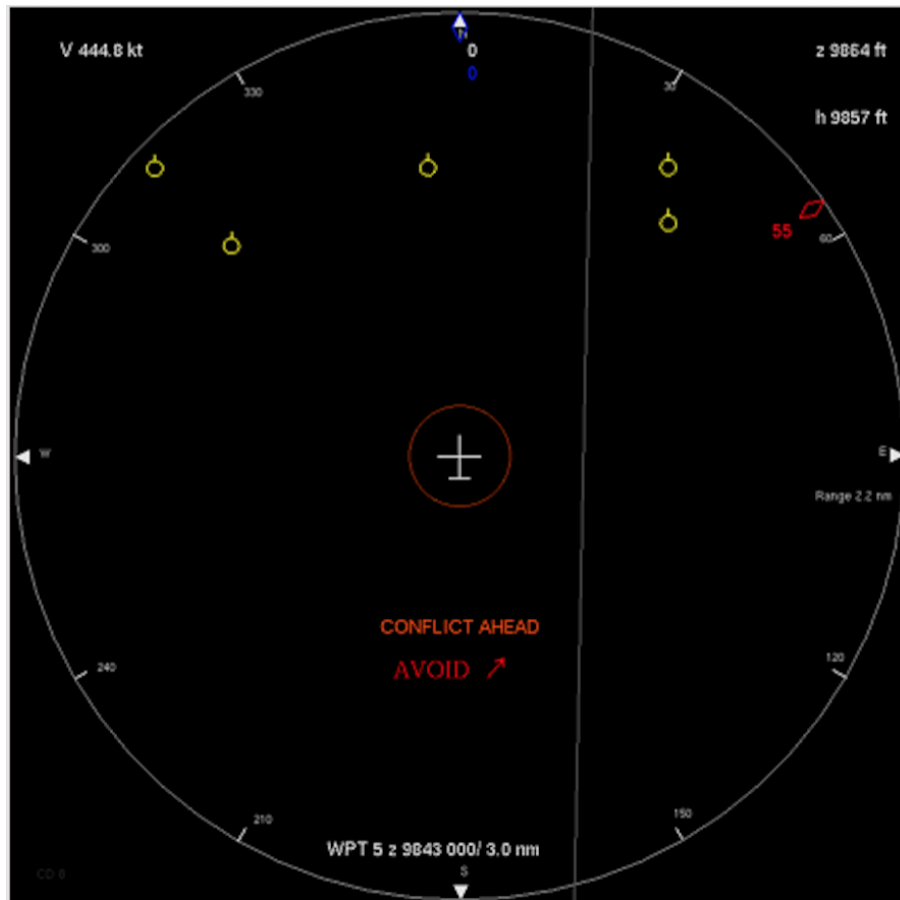


FIGURE VII.1 – Experimental set up with the navigation display : the large white circle represents the angular indication. The smaller white circle surrounding the plane represents the aircraft safety envelope. Before implementing a new path, the message "AVOID" appeared together with an arrow and a mark on the large white circle indicated the future direction selected by the system.

in a complete random order.

The time elapsed between the prime message (indicating the future direction) and the "e" key press was recorded in order to evaluate the performance of participants in these critical situations.

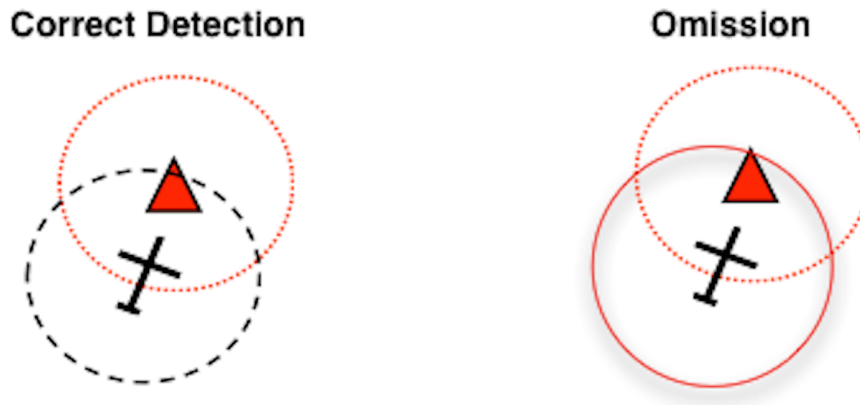


FIGURE VII.2 – Description of the visual feedback provided to the participant when the system takes an inappropriate decision (8% of cases). Left : feedback after participant correctly detected the error by pressing the "e" key. Right : feedback displayed if participants did not press the "e" key showing the aircraft dangerously approaching another obstacle. In this case, a red circle surrounded the aircraft.

Multi-Attribute Task Battery II

MATB II is a multi-tasking research platform designed to assess performance on four main, concurrent tasks (tracking, monitoring, resource management, and communications, see Figure VII.3). MATB II is an updated version of the original MATB (Comstock & Arnegard, 1992 ; Santiago-Espada et al., 2011). This task was performed on a second screen, at the same time as the supervision task. We did not use the tracking task given that the context of the study is the supervision of a system and not direct control over an action.

The monitoring task had two components : operators responded to red and green lights, and to scales that go out of range by registering either too high or too low. The resource management task represented fuel management aboard an aircraft. Operators maintained fuel in two tanks that continually deplete below target levels. Tanks are connected by pumps, which direct resource flow into or out of each of the tanks, and are

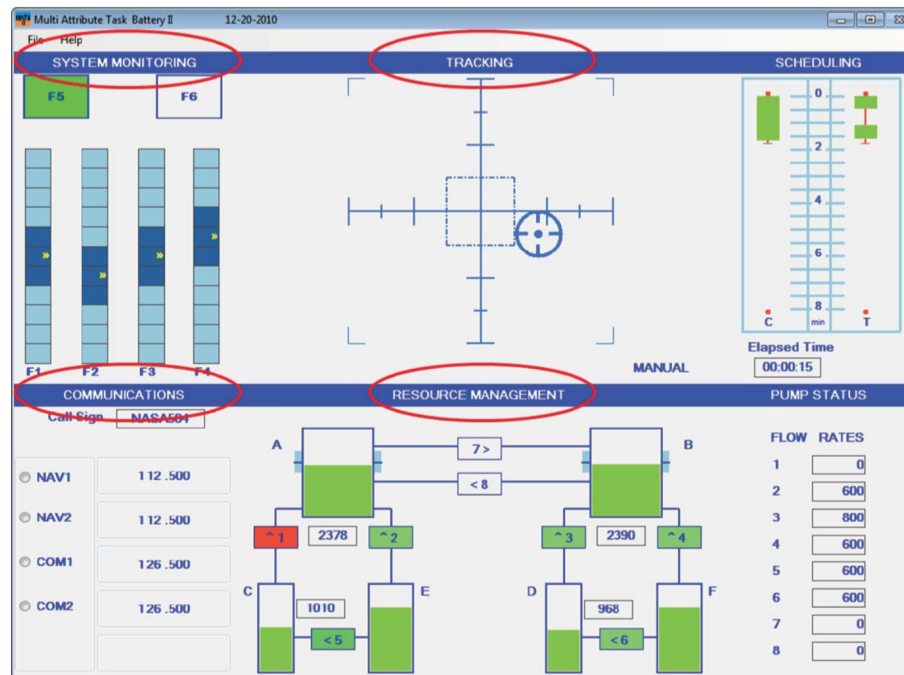


FIGURE VII.3 – Representation of the tasks for the Multi-Attribute Task Battery II.

controlled by the operator to regulate tank levels. Events in resource management are failures of interconnecting pumps. In each test trial, pumps failed once in randomized order, and were repaired automatically. The communication task simulated pilot interaction with air traffic controller requests. Operators heard command messages to change radio and frequency on one of four communications radios. The instructions were directed either to an "other" ship or to the operator's own ship by using a callsign designated during training. Requests could be ignored unless the own sign was called.

Additional measures

Three cameras tracked participants' eyes in order to evaluate the time spent on each task. The aim of this measure was to evaluate the allocation of attentional resources depending on the experimental condition. We hypothesized that priming system's intentions has a different impact on our measures of interest depending on the timing.

At the end of each block, participants were asked to evaluate their sense of control over the automated system and their acceptability. Sense of control was measured by a classic question in the agency literature (*I felt myself in control when I use this system*) on a 7-point Likert-type scale (1 = totally disagree, 7 = totally agree).

Then, acceptability of the automated system from the supervision task was measured by seven questions on a similar Likert-type scale. The questionnaire contains items related to "trust dimension" (*I found this system reliable ; I trust this system*), "safety dimension" (*This system increase my performance during the supervision task ; Using this system increases the safety level*) and perceived ease of use (*It is difficult for me to understand the system's behavior ; Using this system does not require high attentional resources ; Using this system is stressful*). These dimensions are known to be fundamental for user acceptance (Ghazizadeh et al., 2012 ; Lee & Moray, 1994 ; Parasuraman & Wickens, 2008 ; Venkatesh, 2000). Answers to these questions were averaged to form a reliable measure of acceptability (Cronbach's alpha = .87). Higher scores reflect more acceptability of the automated system.

Data Analysis

For all three measures described in this paper, we were interested in knowing which trend best describes their relation with the delay between the prime message and the system's action. We tested both a linear and a quadratic contrast by using a bootstrapping sampling technique (Efron & Tibshirani, 1994) with $n = 2000$. Bootstrap methods provide a robust way of testing hypotheses and computing confidence intervals without assuming normality (Wilcox, 2010, 2011 ; Xiong, Boyle, Moeckli, Dow, & Brown, 2012). Assuming normality could be problematic here because of our small sample size. Statistical significance is reached if zero is not included in the confidence interval. We only reported the most significant trend for each measure.

Results

The results concerning measures of reaction time in situations requiring a take-over and measures on the time spent on the supervision task are reported in Figure VII.4. We performed a trend analysis. This analysis shows a significant linear trend ($\Psi = 7.75$, 95% CI [6.71, 8.80]), indicating that as the delay between prime message and system's action decreased, reaction time decreased proportionately. We observed a similar linear trend for the time spent on supervision task ($\Psi = 2.94$, 95% CI [1.70, 4.18]). Unsurprisingly, these results indicate that the reaction time and the time spent on supervision task are sensitive to the delay between prime message and system's action.

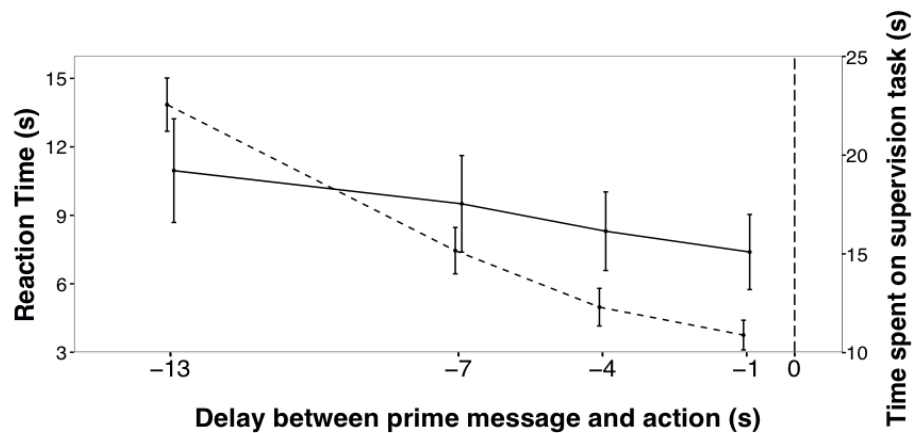


FIGURE VII.4 – Mean response time to press the "e" key (dashed line) and mean time spent on the supervision task during a trial (solid line) at each of the 4 delays. Errors bars are 95% CIs.

Concerning the sense of control, we observed similarly to Wegner and Wheatley (1999), a quadratic polynomial effect ($\Psi = -0.71$, 95% CI [-1.25, -0.17]). Compared with trials when prime messages were displayed 13s or 1s, there was an increase of sense of control when the prime messages were displayed 7s or 4s before the system's action (see Figure VII.5). Interestingly, we observed a similar quadratic polynomial effect regarding the evaluation of the user acceptance ($\Psi = -0.72$, 95% CI [-1.17, -0.28]). These

results replicate the previous claim of Wegner and Wheathley (1999) that the timing of the prior though in relation to the action is important to the sense of control. This claim can be extended to user acceptance in a supervision context.

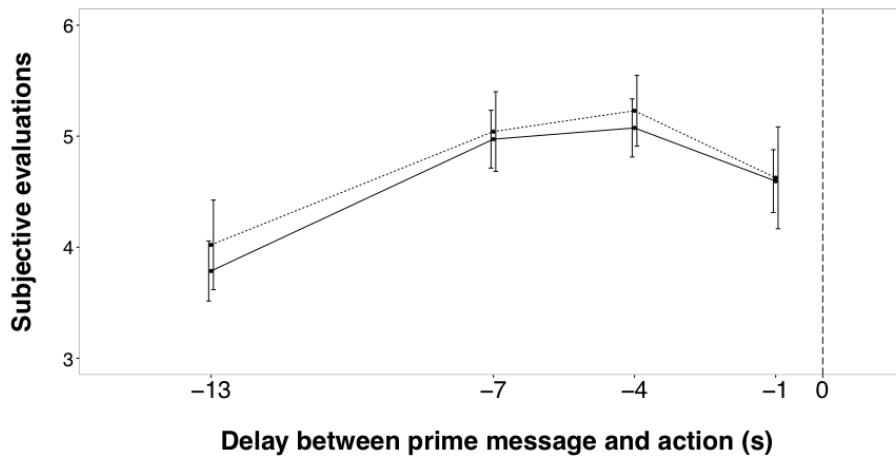


FIGURE VII.5 – Figure 5. Mean values for the sense of control (dashed line) and the user acceptance (solid line) at each of the 4 delays. Errors bars are 95% CIs.

Discussion

Two main results were obtained in the present study. Our results showed that the delay between prime messages providing system's intentions and the system's actions had an impact on (1) three major components of human-machine interaction (i.e., performance, operator's levels of control and acceptance), and that (2) this impact differs qualitatively regarding these three components (i.e., linear or quadratic trend).

Concerning the performance, results showed a linear relationship between delay and reaction time to detect a non-optimal decision : longer delays led to longer reaction times. Recently, Gold et al. (2013) investigated the behaviour of drivers when a takeover was required due to an automation failure. Automation was in charge of the driving operations while drivers performed a secondary task. When the automation detected an

unexpected event (i.e., a car stopped on the road), an alarm warned the driver that a takeover was required. It was found that drivers who received shorter takeover request times reacted more quickly.

Our results are consistent with those of Gold et al. (2013) regarding reaction times. Priming system's intentions just before its action facilitates the creation of the intention-action-effect chain. This result is in line with the results of Chambon and Haggard (2012) that demonstrates that action priming affects the fluency of action selection processes and increases performance in reaction times.

In addition to better reaction times, we also found a decrease in the time spent on the supervision task. The average time spent on each area of interest is traditionally considered as an indication of the difficulty to extract and interpret information from this area (Fitts et al., 1950 ; Jacob & Karn, 2003). Therefore, shortening the delay between prime messages and system's actions, simplifies the supervision task, and leads to less cognitive effort in deciding whether we have to correct automation's action.

Previous studies explained that the ambiguity about who is the agent of the action has a major impact on user acceptance (Baron, 1988 ; Le Goff et al., 2016 ; Limerick et al., 2014). We hypothesized that a better understanding of the role played by agency may therefore provide a useful framework for thinking about interactions with automated technology (Le Goff et al., 2015). Notably, empirical studies demonstrated that the priming effect enables the respect of priority and consistency principles when people experiments vicarious agency (Wegner et al., 2004 ; Wegner & Wheatley, 1999), even when this *vicarious* agency takes place with automation (Le Goff et al., 2016).

Wegner and Wheatley (1999) showed a quadratic relation between the timing of the prime message and the sense of control. We also observed an inverted U-curve describing the impact of this delay on the user's sense of control. Although the timing delays used by Wegner and Wheatley (-30s, -5s, -1s, +1s) are slightly different from ours (-13s, -7s, -4s, -1s), both results suggest that prime messages should be displayed in a relevant temporal

window to optimize the priority and congruency principles and to raise a correct sense of control. It is likely that this temporal window strongly depends on the task people do (direct control, supervision ...).

Interestingly, we found that user acceptance followed the same quadratic pattern than the user's sense of agency. Although, there was a tendency overall for participants to evaluate the system more acceptable when the delay between prime messages and actions decreases, there was a marked fluctuation in this evaluation depending on when the prime occurred. Participants evaluated the system as more acceptable when prime messages occurred neither too soon (-13s) nor too late (-1s). Therefore, warnings have to be displayed but with a precise timing otherwise user acceptance might not be optimal.

Although, these results should be complemented in future research by further studying the content of the prime message, the present study confirms that the psychological approach of agency can help proposing design principles to improve human machine interaction when human operators are confronted with automated systems (Berberian et al., 2012 ; Le Goff et al., 2015 ; Limerick et al., 2014). This approach provides promising guidelines not only for improving the operator's performances but also the operator's sense of control and the system's acceptability.

Acknowledgments

We thank the "Direction Générale de l'Armement" (DGA) for its financial support to the first author.

Discussion Générale

Mise en perspective de nos résultats

Le but de cette thèse était d'étudier comment le cadre théorique de l'agentivité pouvait nous aider à identifier et évaluer les informations requises pour permettre à l'opérateur de rester dans la boucle de contrôle, c'est-à-dire rendre la supervision de systèmes fortement automatisés plus efficace, restaurer un sentiment de contrôle approprié et augmenter l'acceptabilité des systèmes. Pour y parvenir, nous avons réalisé quatre expériences lors desquelles les participants devaient superviser le vol d'un avion effectuant un déplacement sous pilote automatique. Dans ce dernier chapitre, nous rappellerons dans un premier temps le contexte dans lequel s'inscrit cette thèse. Puis, nous résumerons les principaux résultats des expériences menées au cours de cette thèse. Ensuite, nous discuterons de l'implication de nos résultats dans le domaine des interactions homme-machine. Pour conclure, nous exposerons les perspectives ouvertes par les résultats de cette thèse.

Rappel du contexte

Il est maintenant établi que l'utilisation de systèmes de plus en plus automatisés tend à éloigner les pilotes des actions à effectuer des effets produits par ces actions (Limerick et al., 2014). Cette sortie de la boucle de contrôle est identifiée depuis longtemps (Billings, 1991 ; Billings, Lauber, Funkhouser, Lyman, & Huff, 1976 ; Wiener, 1988 ; Wiener & Curry, 1980). Ce phénomène n'est pas propre à l'aviation et se retrouve également dans d'autres secteurs tels que celui l'automobile (par exemple, Banks & Stanton, 2016 ; Louw

& Merat, 2016).

Cette sortie de la boucle de contrôle se traduit par une forme de désengagement dans la tâche, ce qui peut devenir problématique lors de situations imprévues (pannes, erreurs, etc.) pour lesquelles l'opérateur doit reprendre en main le système rapidement et efficacement. Les études empiriques ont montré, entre autres, une diminution de la compréhension de la situation par l'opérateur (Kaber & Endsley, 1997), un excès de confiance des opérateurs vis-à-vis des systèmes automatisés (Parasuraman & Manzey, 2010) et une charge de travail inégalement répartie au cours de la tâche, avec une charge insuffisante pour être efficace dans une tâche de monitoring (Young et al., 2015 ; Young & Stanton, 2007) et une forte augmentation de la charge de travail, accompagnée d'un stress accru, lors des reprises en main non anticipées (Sarter et al., 1997) ainsi qu'une augmentation de leurs temps de réaction (Merat & Jamson, 2009 ; Young & Stanton, 2007). Maintenir les opérateurs dans la boucle de contrôle est donc devenu un enjeu majeur pour l'ensemble de la communauté des facteurs humains, quel que soit le domaine d'application.

Parmi les différentes solutions proposées (entraînements, allocation de fonctions, etc.), de nombreuses études se sont intéressées à la transformation du système automatisé en agent collaboratif (Christoffersen & Woods, 2002 ; Dragan et al., 2013 ; Klein et al., 2004 ; Zimmermann et al., 2014). Une des pistes explorées pour l'amélioration de cette collaboration est celle de l'utilisation de messages amorces pour transmettre des informations aux opérateurs. Toutefois, certains résultats obtenus dans ces études ont montré que la performance et l'acceptabilité pouvaient être décorréliées (Koo et al., 2015 ; Navarro et al., 2010), confirmant le fait que la conception d'un système plus collaboratif n'est pas forcément une chose aisée car de nombreux paramètres sont à prendre en compte.

A travers cette thèse, nous avançons l'idée selon laquelle le cadre théorique de l'agentivité pourrait aider à concevoir de tels systèmes. En effet, les concepteurs de systèmes automatisés pourraient largement bénéficier des avancées récentes de la recherche en

sciences cognitives sur le sentiment d'agentivité, à la fois en termes de mesures, de recommandations et dans la compréhension des mécanismes mis en œuvre lors de l'utilisation de systèmes automatisés.

Pour appuyer notre propos, nous avons choisi d'utiliser la technique d'amorçage (supraliminal) afin de favoriser l'agentivité des participants dans un contexte de supervision. En effet, Wegner et al. (2004) ont montré que l'effet d'amorçage pouvait être une bonne manière d'influencer ou de simuler les pensées a priori et par conséquent de créer un sentiment d'agentivité même en l'absence de mouvement. Ces pensées permettent de respecter les principes de priorité et de cohérence¹ établis par Wegner et Wheatley (1999). Nous faisons l'hypothèse que ces principes peuvent être utiles pour structurer les messages générés par le système automatisé et pour déterminer leur temporalité d'apparition afin de maintenir un sentiment de contrôle optimal de l'opérateur malgré une autonomie élevée de l'automatisme.

Discussion des résultats obtenus

Augmentation du sentiment de contrôle

Il est intéressant de constater que le sentiment de contrôle des opérateurs ou utilisateurs est loin d'être la variable la plus étudiée dans la communauté des facteurs humains. Pourtant, la prise en compte du sentiment de contrôle des utilisateurs pour la conception de nouvelles interfaces n'est pas nouvelle. En effet, dans leurs huit "règles d'or" pour la conception d'interfaces, Shneiderman et Plaisant (2004) conseillent aux concepteurs de créer des interfaces qui maintiennent un lieu de contrôle interne. Ce conseil est basé sur l'idée que les utilisateurs "désirent fortement se sentir en charge du système et que le

1. Le principe d'exclusivité est le troisième principe établi par Wegner et Wheatley. Toutefois, nous n'en parlerons pas ici car l'identité de l'agent qui est en charge de l'action (i.e., le système) n'est pas ambiguë contrairement à l'étude initiale de Wegner et Wheatley (1999). Le principe d'exclusivité pourrait être étudié en faisant l'hypothèse que dans les situations de supervision, l'opérateur et lui-seul (exclusivité) possède un veto sur les actions du système.

système réponde à leurs actions" (Shneiderman & Plaisant, 2004).

Bien que notre thèse s'appuie sur le cadre théorique du sentiment d'agentivité, il est important d'éclaircir ici un point important. Nous parlons de sentiment de contrôle et de sentiment d'agentivité de façon interchangeable. Ce choix de l'interchangeabilité des termes est facilement critiquable en fonction des définitions que l'on en donne. Par exemple, Pacherie (2007, 2008) subdivise le sentiment d'agentivité en sentiment d'intentionnalité, d'initiation et de contrôle. Le sentiment de contrôle apparaît donc comme une composante du sentiment d'agentivité et non un équivalent. Pourtant, nous avons fait ce choix car le sentiment de contrôle est l'expression couramment utilisée dans le domaine des interactions homme-machine et le sentiment d'agentivité tel qu'il est défini actuellement dans la littérature correspond à une action effectuée par l'agent lui-même (Chambon et al., 2015 ; Haggard et al., 2002 ; Haggard & Tsakiris, 2009). Notre thèse n'a pas pour but de recréer l'illusion que l'opérateur produit lui-même l'action comme dans l'expérience de Wegner et Wheatley (1999) et comme suggéré par Wegner (2004) avec sa théorie de l'illusion de la volonté consciente (*illusion of conscious will*) mais seulement de maintenir un sentiment de contrôle optimal chez l'opérateur grâce au respect des principes de priorité et de cohérence afin non seulement d'augmenter l'acceptabilité des systèmes automatisés mais également afin de permettre un investissement plus important de l'opérateur dans la tâche en cours.

En utilisant l'effet d'amorçage dans l'Expérience 1, nous avons observé une augmentation subjective de la confiance des participants dans leur capacité de reprise en main en cas d'erreur du système traduisant ainsi une augmentation de leur sentiment de contrôle vis-à-vis de l'appareil. Cette augmentation n'était visible que pour la condition *directional prime*, condition pour laquelle le système donnait une indication aux participants sur la temporalité avant l'implémentation de l'action et la direction sélectionnée. A l'inverse, dans la condition *temporal prime*, le fait de connaître le temps avant l'implémentation de l'action par le système n'a pas aidé les participants à ressentir plus de contrôle par

rapport à la condition *no prime*, c'est-à-dire sans information directionnelle ni temporelle.

Ces résultats confirment que l'utilisation de messages amorces a un impact positif sur le sentiment de contrôle des opérateurs lorsque ces messages contiennent suffisamment d'informations. Lorsque nous discuterons des résultats sur l'acceptabilité, nous verrons que ce résultat n'est pas aussi trivial qu'il n'y paraît.

La performance

En plus de la mesure subjective du sentiment de contrôle, il nous fallait également prendre en compte des mesures objectives comme la performance. La performance est nécessairement un paramètre important, à la fois d'un point de vue opérationnel mais également au niveau du sentiment d'agentivité. En effet, Wen et al. (2015) ont montré que la réussite dans la tâche à réaliser était également un facteur influençant le développement du sentiment d'agentivité.

Les Expériences 2 et 3 ont montré que les participants bénéficiaient davantage du système fournissant les informations temporelles et directionnelles que du système fournissant uniquement les informations temporelles. Le temps de réaction des participants était raccourci lorsqu'ils supervisaient le système dans la condition *directional prime*. Nous pourrions être tentés d'expliquer cette augmentation de la performance par la simple présence de plus d'informations en présence d'un prime à la fois temporel et directionnel par rapport aux deux autres conditions. Pourtant, force est de constater que l'augmentation d'information entre le prime temporel et la condition sans prime n'engendre aucune augmentation de la performance. Autrement dit, la nature même de l'information amenée à l'opérateur semble prépondérante dans l'augmentation de la performance.

Dans un souci de simplification que nous avons justifié dans la partie "Matériel et Méthodes de l'Introduction", nous avons préféré nous intéresser au temps de réaction pour la détection d'une erreur et non pas aux procédures mises en place lors de la reprise en main. Nous sommes conscients que cela représente une des limites importantes de nos

études. En effet, de nombreuses études ont montré une dégradation de la qualité de la reprise en main après des phases de contrôle laissées au système automatique (Hoc et al., 2006 ; Stanton, Young, Walker, Turner, & Randle, 2001 ; Young et al., 2015).

Par conséquent, seules des études supplémentaires se concentrant sur le comportement de l'opérateur lorsqu'il doit revenir à un contrôle manuel et sur l'évaluation en détails de la qualité de sa reprise en main (e.g., analyse des procédures) et de la perception de son environnement (e.g., mesures de la conscience de situation) permettraient de conclure efficacement sur l'apport du cadre théorique de l'agentivité dans la compensation du phénomène de sortie de boucle de contrôle.

L'attention

Une mesure complémentaire de la performance a été obtenue grâce à l'oculométrie. En effet, nous nous sommes intéressés à la gestion des ressources attentionnelles des participants. Les résultats des Expériences 2 et 3 montrent que le système donnant à la fois une indication temporelle et directionnelle permettait aux participants d'allouer plus de ressources attentionnelles à la tâche annexe tout en augmentant leur performance de détection sur la tâche principale. Ces résultats sont corroborés par l'évaluation subjective des participants lorsqu'ils devaient estimer eux-mêmes le temps passé sur chaque tâche après chaque bloc de l'Expérience 2.

Ces résultats confirment nos hypothèses sur les bénéfices en termes d'allocation de ressources attentionnelles apportés par la supervision d'un système automatisé moins opaque (Adler & Benbunan-Fich, 2012 ; Fitts et al., 1950 ; Jacob & Karn, 2003). Les opérateurs de tels systèmes doivent souvent pouvoir gérer plusieurs tâches en parallèle (e.g., les pilotes d'avion qui doivent monitorer plusieurs instruments de vol tout en supervisant le pilote automatique et gérer les communications avec la cabine et avec le sol). Ainsi, diminuer l'opacité du système grâce à des messages amorces permettrait aux opérateurs de mieux gérer les tâches annexes et tout aussi importantes.

Toutefois, il est également important de s'interroger sur l'effet inverse de la libération des ressources attentionnelles. De nombreuses études ont montré qu'une diminution de l'investissement de ces ressources pouvait également avoir des effets néfastes lors d'interactions avec des systèmes automatiques (*mental underload*, Desmond & Hoyes, 1996 ; Hancock & Verwey, 1997 ; Young et al., 2015 ; Young & Stanton, 2007). En particulier, De Winter, Happee, Martens, et Stanton (2014) expliquent que la libération des ressources attentionnelles va faciliter le passage à des tâches annexes qui n'ont pas de lien avec la tâche principale de l'opérateur (e.g., l'utilisation du téléphone portable au volant). Casner et Schooler (2013, 2015) ont quant à eux montré le lien entre l'automatisation croissante et les périodes de divagation attentionnelle (ou *mind-wandering*) des pilotes lors des tâches de pilotages. Dans une étude récente, Merat, Jamson, Lai, Daly, et Carsten (2014) ont montré dans une tâche de conduite avec une voiture autonome que la réponse des conducteurs face à des situations critiques était similaire lorsque leur attention n'était pas portée sur une autre tâche. À l'inverse, les performances étaient dégradées lorsque les conducteurs étaient distraits par une autre activité.

Par ailleurs, cette diminution de la charge attentionnelle pourrait également traduire un excès de confiance vis-à-vis du système automatisé. En analysant spécifiquement les réponses à la question portant sur le niveau de fiabilité perçu entre les deux systèmes, les résultats montrent que pour un même ratio "bonnes/mauvaises décisions", les participants évaluent le système donnant les indications temporelles et directionnelles comme étant plus fiable que l'autre système. Un tel résultat apparaît comme bénéfique pour l'acceptabilité des automatismes par les opérateurs à condition qu'il ne traduise pas un excès de confiance vis-à-vis des actions implémentées par le système. En effet, nous avons vu dans l'introduction que cet excès de confiance était délétère dans le cadre des IHM car elle diminuait la capacité des opérateurs à remettre en cause les actions du système et augmentait également le temps nécessaire à la détection des erreurs commises par le système automatisé (Lee & See, 2004 ; Parasuraman, Molloy, & Singh, 1993 ; Payre,

Cestac, & Delhomme, 2016).

Il est donc important pour notre propos de vérifier lors de futures études qu'un système favorisant l'agentivité des opérateurs entraînent bien une libération bénéfique des ressources attentionnelles pour la gestion de tâches annexes liées à la tâche principale et non pas un désengagement de la tâche ainsi qu'un excès de confiance. L'amélioration des performances observée dans nos études est un premier argument en ce sens.

L'acceptabilité

A première vue, les résultats concernant la performance et le sentiment de contrôle peuvent sembler évidents (i.e., donner plus d'informations à l'opérateur l'aide à mieux superviser un système automatisé), pourtant il est important de rappeler que plusieurs études ont montré que le simple fait de donner plus d'informations n'était pas toujours une solution pour l'amélioration des interactions entre l'humain et les systèmes automatiques notamment en termes d'acceptabilité de l'utilisateur. De plus, il faut rappeler qu'un système qui n'est pas acceptable par les opérateurs ne sera pas utilisé malgré la plus-value au niveau de la sécurité ou de la performance (Inagaki & Nagai, 2007).

Koo et al. (2015) ont montré que les bénéfices apportés par les messages amorces dépendaient en partie de la nature de l'information transmise. Dans leur étude, ils ont montré que les participants préféraient superviser une voiture autonome qui leur indique pourquoi la voiture devait freiner (i.e., obstacle en face) plutôt que l'action que la voiture va implémenter (i.e., la voiture freine). De plus, la combinaison des deux types de messages (i.e., la voiture freine car il y a un obstacle en face) entraînait de meilleures performances de conduite mais diminuait l'acceptabilité du système par les conducteurs. Cette étude nous montre donc que la nature et la quantité des informations transmises à l'opérateur sont cruciales pour augmenter à la fois la performance mais également l'acceptabilité du système.

Navarro et al. (2010) ont également obtenu des résultats montrant une divergence entre performance et acceptabilité. L'amorce motrice utilisée au niveau du volant pour indiquer la direction à prendre par le conducteur était jugée moins acceptable que l'alarme sonore latéralisée malgré une qualité de conduite améliorée. Ce résultat montre que la modalité à travers laquelle le message amorce va être présentée est également à prendre en compte.

Contrairement aux études de Koo et al. (2015) et de Navarro et al. (2010), nous nous sommes placés dans un cadre de supervision. Nous avons donc posé dans l'Introduction, l'hypothèse que dans un tel cadre d'interactions homme-machine, ces amorces ne seraient pas perçues comme intrusives car l'opérateur n'est pas celui qui produit l'action, faisant ainsi écho aux études réalisées par Sato (2009) ; Wegner et al. (2004). Nous pensons au contraire que ces informations seraient perçues comme une forme de communication de la part du système.

Les résultats de l'Expérience 2 et 3 ont confirmé cette hypothèse. A travers les questionnaires d'acceptabilité des deux expériences, les participants ont évalué le système donnant des informations directionnelles et temporelles plus acceptable à superviser que le système n'indiquant que les informations temporelles. Ainsi, la diminution du temps de réaction pour détecter l'erreur du système et l'augmentation du sentiment d'agentivité s'accompagnent également d'une augmentation de l'acceptabilité des participants pour le système automatique supervisé.

Le lien temporel

L'étude de Koo et al. (2015) nous indique que la nature et la quantité d'informations transmises à l'opérateur est importante. L'étude de Navarro et al. (2010) nous montre que la modalité à travers laquelle l'information est véhiculée est aussi un paramètre à prendre en compte. Récemment Gold et al. (2013) dans une étude sur les automobiles autonomes ont montré que le délai des messages amorces était également critique. Dans cette étude,

les auteurs se sont intéressés au temps nécessaire au conducteur pour revenir dans la boucle de contrôle lors de la détection d'un obstacle. Ils ont fait varier le délai entre l'apparition d'un message amorce, indiquant au conducteur la nécessité d'une reprise en main, et le temps estimé avant la collision avec l'obstacle s'il n'y a pas de changement de trajectoire. Les résultats ont montré qu'un délai de 5s entraînait une reprise en main plus rapide de la part du conducteur mais aussi une dégradation de la qualité de sa conduite. A l'inverse un délai de 7s permettait au conducteur une reprise en main plus fiable (e.g., vérification de l'angle mort) mais augmentait également son temps de réaction.

La temporalité du message amorce est également un paramètre important pour respecter le sentiment d'agentivité du participant. Dans leur étude de 1999, Wegner et Wheatley ont montré l'existence d'une relation quadratique entre l'intervalle de temps séparant l'apparition du message amorce et l'action de l'expérimentateur et le sentiment d'agentivité du participant. Plus l'intervalle temporel entre le mot présenté dans le casque et l'arrêt sur cet objet était court et plus le participant reportait une forte volonté d'action pour l'arrêt réalisé. Le but de l'Expérience 4 était donc de vérifier l'impact de la temporalité des messages amorces sur les trois variables principales de cette thèse : le sentiment de contrôle, la performance et l'acceptabilité.

Les résultats de notre étude ont confirmé l'existence d'une relation quadratique entre l'intervalle de temps et le sentiment de contrôle du participant, répliquant ainsi le résultat obtenu par Wegner et Wheatley (1999) dans une étude plus complexe sur les interactions homme-machine. De façon intéressante, cette relation quadratique se retrouve également dans l'évaluation de l'acceptabilité des participants. Cela confirme donc l'importance de la prise en compte du sentiment d'agentivité des opérateurs lors de la conception d'interfaces automatisées pour améliorer le sentiment d'acceptabilité des opérateurs.

Quant à la performance, elle suit non pas une relation quadratique mais linéaire, rejoignant ainsi les résultats de Gold et al. (2013) montrant que plus le temps entre l'amorce et l'action est court et plus cela oblige l'opérateur à agir rapidement. Le simple fait

d'ajouter des messages amorces n'est donc pas suffisant pour améliorer le sentiment de contrôle de l'opérateur, sa performance et son acceptabilité. Des études supplémentaires sont nécessaires afin de déterminer la temporalité optimale permettant le meilleur compromis possible entre le sentiment de contrôle, d'acceptabilité et le niveau de performance des opérateurs.

Liage intentionnel

Un autre résultat important concerne l'absence d'effet de notre condition *directional prime* sur le liage intentionnel. Son utilisation en tant que marqueur implicite de l'agentivité est à questionner. En effet, même si la performance et les marqueurs explicites tels que le sentiment d'agentivité et l'acceptabilité des opérateurs sont les aspects les plus importants au niveau opérationnel, l'utilisation de mesures implicites permettrait une validation supplémentaire de l'apport du cadre théorique de l'agentivité dans le domaine des interactions homme-machine.

Notre étude n'est pourtant pas la première à utiliser le liage intentionnel pour mesurer implicitement le sentiment d'agentivité d'un opérateur interagissant avec un système automatisé. Berberian et al. (2012) avaient par exemple montré que l'augmentation croissante du niveau d'automatisation entraînait une diminution du phénomène de liage intentionnel (i.e., une compression temporelle moins forte). L'utilisation du liage intentionnel en tant que mesure implicite du sentiment d'agentivité a également été testée pour évaluer l'impact de la modalité d'interaction entre les utilisateurs et un système. Coyle et al. (2012) ont montré qu'appuyer sur sa propre peau (*skinput*) pour déclencher un effet entraînait plus de compression temporelle qu'un appui classique sur un clavier. Dans une autre étude similaire, la même équipe a comparé le liage intentionnel entre un appui clavier et un déclenchement vocal. Cette fois, c'est l'appui clavier qui a entraîné plus de liage, conduisant les auteurs à conclure en classant l'appui sur la peau comme entraînant le plus d'agentivité, suivi de l'appui clavier et pour finir l'interface vocale

(Limerick et al., 2015).

A l'inverse, il existe également des études montrant une absence de liage intentionnel lorsque les participants interagissent avec des systèmes automatisés. En effet, dans leur étude comparant des interactions humain-humain par rapport à des interactions humain-machine, Obhi et Hall (2011) n'ont pas observé de liage intentionnel dans le cas où les participants interagissaient avec un système automatisé, à l'inverse des situations où les participants agissaient avec un autre humain. Les auteurs ont émis l'hypothèse que nous n'étions pas capables d'attribuer des intentions à un système automatisé et que notre mécanisme de neurones miroirs n'était pas activé lors d'une tâche en coopération avec une machine, condition sine-qua-none pour observer le phénomène de liage intentionnel lors de co-actions (Obhi & Hall, 2011 ; Wohlschläger et al., 2003). Nous discuterons de ce point dans la partie "Perspectives".

Les résultats contradictoires obtenus lors de l'utilisation du liage intentionnel comme mesure implicite de l'agentivité dans le cadre d'interactions avec un système automatisé semblent surtout être le reflet de la difficulté pour les chercheurs à définir clairement les mécanismes sous-tendant ce phénomène de compression temporelle. De futures études sont indispensables pour conclure sur la pertinence ou non de cette mesure implicite de l'agentivité en IHM. Il faut également se rappeler que le phénomène d'atténuation sensorielle est considéré à l'heure actuelle comme une autre mesure implicite de l'agentivité. Cette mesure peut également faire l'objet d'études afin de déterminer si elle pourrait être plus fiable que le liage intentionnel pour évaluer l'agentivité des opérateurs supervisant des systèmes automatisés.

Limites de notre approche

Limites méthodologiques

A travers les quatre expériences réalisées pendant cette thèse, il est important de citer quelques limites inhérentes aux choix méthodologiques que nous avons effectués. En effet, malgré l'utilisation du simulateur de page radar LIPS de l'ONERA ainsi que des tâches annexes ayant un lien avec les tâches aéronautiques (i.e., tâche de *tracking* pour les Expériences 2 et 3 ; *Multi-Attribute Task Battery II* pour l'Expérience 4), notre environnement expérimental reste très éloigné des situations opérationnelles. Il existe logiquement de nombreuses variables (stress, contexte opérationnel, durée de la tâche, charge de travail, etc.) que nous retrouvons dans une situation écologique et qui peuvent certainement modifier l'impact des messages amorces sur la performance, le sentiment de contrôle et l'acceptabilité des opérateurs.

Il faut également souligner le nombre d'essais critiques que nous avons dû maintenir assez bas² pour satisfaire des contraintes temporelles (certains essais pouvaient durer 40s) et garder un temps d'expérimentation inférieur à deux heures et pour ne pas biaiser la confiance des participants vis-à-vis du système (i.e., si le système se trompe trop souvent, il devient inutile). Ce niveau de fiabilité perçue est pourtant un paramètre majeur lors de l'utilisation de systèmes autonomes. Ainsi, il serait intéressant pour notre

2. Nous avons effectué des statistiques non paramétriques pour l'ensemble de nos analyses afin de vérifier que le faible échantillon n'était pas problématique. Toutefois, pour des raisons de simplification, nous avons décidé de reporter les résultats des tests paramétriques (à l'exception du dernier article).

propos de répliquer ces études empiriques avec un niveau de fiabilité plus faible (i.e., augmenter le nombre de mauvaises décisions), plus élevé (i.e., diminuer le nombre de mauvaises décisions) ou encore variable, car nous avons vu en Introduction, avec l'étude de Parasuraman, Molloy, et Singh (1993), qu'un niveau variable de fiabilité permettait de diminuer l'excès de confiance vis-à-vis de l'automatisme.

Il serait également intéressant de se concentrer sur la difficulté des tâches annexes. En effet, nous avons volontairement déterminé un niveau de complexité assez bas afin que les participants ne soient pas absorbés par cette tâche et se concentre également sur la tâche de supervision. Dans un cadre opérationnel toutefois, il n'est pas rare que les opérateurs maintiennent leur attention sur d'autres tâches plutôt que de monitorer les automatismes. Ainsi, augmenter la complexité de la tâche annexe ou augmenter le nombre de tâches à gérer simultanément par l'opérateur pourraient nous permettre de montrer s'il existe réellement un bénéfice en termes d'allocation des ressources attentionnelles et surtout si la réalisation de ces tâches annexes n'augmente pas le phénomène de persévération pouvant amener les opérateurs à négliger des événements critiques dans leur tâche de supervision et se focaliser uniquement sur les tâches annexes.

Concernant la modalité d'interaction, nous avons privilégié la modalité visuelle pour le message amorce pour une question de facilité de mise en place. Dans un contexte opérationnel, le nombre d'informations visuelles et sonores présentes dans un cockpit par exemple, rend impératif des études sur la perception de cette information dans un environnement déjà saturé et donc l'impact sur l'acceptabilité de l'opérateur d'un indice supplémentaire. De nombreuses études s'orientent donc vers la modalité tactile pour la transmission d'informations supplémentaires (Navarro et al., 2010, 2007 ; Van Veen & Van Erp, 2001). Il est important de rappeler que le choix de la modalité d'interaction n'est pas anodin. En effet, les études de Coyle et al. (2012) et Limerick et al. (2015) ont montré que les modalités d'interactions avaient également un impact sur le sentiment d'agentivité. Même si l'utilisation du liage intentionnel doit être étudié davantage pour

en faire un marqueur fiable, l'agentivité fournit des mesures et un cadre théorique avec la théorie de l'intégration multiple pour nous permettre de mieux appréhender l'impact des modalités d'interaction sur l'utilisateur et comment les optimiser en fonction de la pondération des différents indices par rapport à la situation dans laquelle se trouve l'opérateur.

Expertise

En plus des différents choix méthodologiques, il faut également discuter de nos résultats par rapport au niveau d'expertise des sujets expérimentaux. En effet, les opérateurs de systèmes fortement automatisés correspondent généralement à une population experte (e.g. les pilotes, les opérateurs de centrales nucléaires, etc.). La longue période de formation et d'entraînements ainsi que la compréhension des systèmes qu'ils doivent superviser permettent aux opérateurs de développer des heuristiques que n'ont pas les sujets expérimentaux que nous avons choisi pour nos études (Rasmussen, 1986).

Il serait donc intéressant de comparer l'impact de ces messages amorces en fonction du niveau d'expertise des opérateurs afin de pouvoir adapter leur temporalité et leur nature. Il est fort probable que les résultats obtenus en termes d'agentivité et d'acceptabilité varient en fonction du niveau d'expertise des opérateurs. Toutefois, le caractère ubiquitaire des systèmes automatisés ainsi que l'arrivée prochaine des voitures autonomes sur nos routes rendent pertinentes les études sur le sentiment d'agentivité appliquées à des populations non expertes.

Niveau d'automatisation

Nous avons rappelé en introduction que de nombreuses taxonomies avaient été proposées afin de rendre compte du niveau d'automatisation d'un système interagissant avec un opérateur humain. Nous avons choisi de nous focaliser sur une tâche de supervision ce

qui correspond à un niveau d'automatisation très élevé de la machine (*supervision control* d'après la nomenclature d'Endsley du Tableau 1). Nous avons fait ce choix d'après les résultats de Berberian et al. (2012) qui ont montré un impact graduel de l'automatisation sur l'agentivité. Du fait des limites méthodologiques, nous voulions utiliser le niveau d'interaction susceptible de produire le plus de rupture au niveau du sentiment d'agentivité des participants.

Des résultats différents pourront être ainsi obtenus lorsque l'opérateur et la machine coopèrent davantage dans la réalisation d'une tâche (e.g., conduite automobile assistée) comme suggéré par l'expérience de Damen et al. (2014). En effet, dans leur expérience sur l'effet d'amorçage, les auteurs ont montré que les amorces congruentes supraliminales diminuaient le jugement d'agentivité. Or, dans cette étude, les effets étaient produits directement par l'action des participants. Ces résultats peuvent nous permettre d'expliquer en partie les résultats obtenus par Navarro et al. (2010).

Une autre problématique apparaît également lorsque le niveau d'automatisation est moins élevé (i.e., système moins autonome). En effet, il faut déterminer si l'amorce (quelles que soient sa nature et sa modalité) intervient soit ponctuellement lors de situations critiques (Navarro et al., 2010, 2007), soit de façon continue (Mars, Deroo, & Hoc, 2014), sachant qu'un niveau de contrôle partagé variable pourrait entraîner des confusions chez l'utilisateur comme souligné par Mars et al. (2014).

Il serait donc intéressant pour de futures investigations de tester l'impact de ces messages amorces en fonction du degré d'autonomie des systèmes automatisés. Il ne serait pas surprenant qu'un contrôle partagé entre l'opérateur et la machine nécessite des messages amorces d'une nature, quantité, modalité et temporalité différentes que dans le cadre de la supervision.

Perspectives

Outre les réponses pouvant être apportées aux limites exposées ci-dessus, deux perspectives de recherche relative au concept d'agentivité nous semblent particulièrement pertinentes.

Une nouvelle identité commune avec des agents artificiels

Toujours dans cette volonté de créer des systèmes autonomes de plus en plus collaboratifs, les concepteurs cherchent à développer des systèmes permettant d'interagir de la même manière qu'un autre opérateur humain. Le cadre théorique de l'agentivité peut aider à la compréhension des mécanismes mis en place lors de ces interactions.

Sebanz, Bekkering, et Knoblich (2006) définissent une co-action comme "toute forme d'interaction sociale lors de laquelle deux individus ou plus coordonnent leurs actions dans un espace et un temps donnés pour modifier leur environnement". De manière intéressante, Pacherie (2011) a récemment introduit l'idée que les différents mécanismes sous-tendant le sentiment d'agentivité pour des actions individuelles sont du même ordre que ceux sous-tendant des actions effectuées en coopération. L'agentivité des actions coproduites pourrait donc être basée sur le même principe de congruence entre les conséquences prédites et réelles des actions et notre capacité à prédire les actions de nos partenaires et leurs effets.

De plus, l'un des critères requis pour la réalisation d'une action conjointe est la

capacité à prédire l'action d'autrui et le résultat associé. La capacité de prédire les actions des autres met en jeu des régions impliquées dans le système de contrôle de nos propres actions (Sebanz & Frith, 2004). Cette capacité est primordiale pour une coordination efficace lors d'une action conjointe (Kilner, Vargas, Duval, Blakemore, & Sirigu, 2004 ; Verfaillie & Daems, 2002 ; Vesper, van der Wel, Knoblich, & Sebanz, 2011). Vesper et al. (2011) ont démontré que les participants impliqués dans une tâche conjointe se rendaient eux-mêmes plus facilement prédictibles en réduisant la variabilité temporelle afin de faciliter la coordination.

De manière intéressante, Obhi et Hall (2011) ont montré que lorsque deux individus interagissent ensemble sur une même tâche, le phénomène de liage intentionnel est présent pour les deux. Les auteurs ont interprété ce résultat comme le reflet de la formation d'une nouvelle identité commune (*we-identity*, voir également Crivelli & Balconi, 2010 ; Poonian & Cunnington, 2013 ; Wohlschläger et al., 2003). Cette identité commune semble également être présente lorsque deux individus possèdent des rôles asymétriques (i.e., un leader et un suiveur, van der Wel, 2015).

Nous avons donc fait l'hypothèse qu'augmenter la prédictibilité du système automatisé en donnant des informations sur ses actions futures favoriserait la création d'une nouvelle identité commune entre l'opérateur et le système, formation se traduisant par le phénomène de liage intentionnel. Pourtant, comme nous l'avons présenté pour expliquer notre absence d'effet pour le liage intentionnel, il semble que la création de cette identité commune n'apparaît pas lors d'interactions avec un agent artificiel (Obhi & Hall, 2011 ; Poonian & Cunnington, 2013 ; Wohlschläger et al., 2003), ce qui appuie le point de vue de Sheehan et Sosna (1991) qui explique que la distance physique et cognitive entre l'humain et un système automatisé entraîne naturellement un éloignement de notre part face aux agents artificiels comparés aux autres humains ou avatars.

Pourtant, à l'inverse, Reeves et Nass (1996) nous explique dans leur fameux livre *The Media Equation* que les humains traitent naturellement les ordinateurs comme des êtres

sociaux. Cette idée se base sur l'observation de personnes se comportant socialement de la même façon vis-à-vis d'un ordinateur que vis-à-vis d'un autre être humain. Ils font également l'hypothèse que ces comportements pourraient être développés par l'anthropomorphisation des systèmes (Lim & Reeves, 2010 ; Nowak & Biocca, 2003 ; Reeves & Nass, 1996). Il serait donc intéressant de voir comment l'augmentation de l'anthropomorphisation des systèmes grâce à des différents paramètres (e.g., mouvements suivant des lois biologiques, communications verbales, ressemblances physiques, émotions, etc.) facilitent le développement d'une identité commune entre l'opérateur et le système artificiel et surtout si cette nouvelle identité favorise la performance et l'acceptabilité dans des tâches de supervision.

Ethique et Responsabilité

A travers cette thèse, nous nous sommes intéressés à l'impact de l'automatisation sur l'acceptabilité des opérateurs. Toutefois, le cadre théorique de l'agentivité peut également nous aider à étudier les problèmes de responsabilité. En effet, pour Haggard et Tsakiris (2009) et Frith (2014), le sentiment d'agentivité joue un rôle central dans le mécanisme d'attribution de la responsabilité. Ainsi, des études récentes comme celle de Caspar, Christensen, Cleeremans, et Haggard (2016) au cours de laquelle les participants ont rejoué la célèbre expérience de Milgram (1963) nous montre, à travers des mesures implicites comme le liage intentionnel et des marqueurs cérébraux, que les personnes qui agissent sous la contrainte ont un niveau d'agentivité qui se rapproche davantage d'actions passives que d'actions purement volontaires. Cette étude illustre habilement l'idée selon laquelle la notion de responsabilité est étroitement liée à la notion d'agentivité et donc à l'engagement dans une tâche.

Pourtant, nous avons vu que l'un des effets néfastes de l'augmentation de l'automatisation des systèmes pouvait être le désengagement des opérateurs dans leurs tâches.

Ce désengagement apparaît comme problématique dans la mesure où l'opérateur reste le responsable d'un point de vue légal du système et de ses actions. Cette responsabilité pénale part en effet du postulat que l'opérateur reste en contrôle des actions du système.

Avec l'autonomie grandissante des systèmes, la notion de responsabilité devient floue. Si l'automatisation diminue les performances de l'opérateur humain en termes de supervision, qu'en est-il de la responsabilité de l'opérateur en cas d'erreur ? Ce paradoxe déjà présent dans l'aéronautique devrait également apparaître dans l'automobile avec les futures voitures autonomes.

Un problème éthique se pose également concernant la distance de plus en plus grande entre l'opérateur et les actions implémentées par le système. Norman (1990) soulignait que les pilotes étaient isolés physiquement et mentalement des activités du système. Cette isolation de l'opérateur tend à l'éloigner du résultat de ses actions médiées par le système. Il arrive à certains pilotes de se plaindre du fait que l'automatisation a réduit leur statut à celui de "presse-boutons" (Baron, 1988).

Cette forme de désengagement vis-à-vis des résultats de l'action peut perturber les mécanismes classiquement mis en œuvre dans la régulation du comportement humain. En effet, différents travaux ont montré que l'intégration des conséquences de nos propres actions était une condition nécessaire pour agir avec éthique et jugement moral (Bandura, Barbaranelli, Caprara, & Pastorelli, 1996 ; Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006 ; Bratman, 2007). L'utilisation dans le secteur militaire de robots est la parfaite illustration de ce problème éthique. L'autonomie de ces robots ou drones s'accroît au fur et à mesure que la technologie progresse. Si la décision d'entrer en combat appartient toujours à l'opérateur, la distance (physique et cognitive) entre lui et ses actions pose clairement des questions au regard des travaux sur la perte de responsabilité lié à la diminution de l'agentivité. Ainsi, de la même manière que la notion d'agentivité peut nous aider à concevoir des systèmes plus acceptables, ce cadre théorique peut également de nous aider à réaliser des interfaces permettant aux opérateurs

de rester engagés dans leur tâche et de se sentir toujours responsables de leurs actions, même si celles-ci sont médiées par un système automatisé.

Conclusion

L'augmentation de l'efficacité des systèmes automatisés tend à faire sortir les opérateurs de la boucle de contrôle, diminuant ainsi leur performance lorsqu'une reprise en main est nécessaire. De nombreuses études se sont penchées sur ce phénomène de sortie de boucle et de nombreuses pistes ont déjà été proposées. Notre thèse avait pour but de venir compléter les efforts déjà fournis en démontrant que les concepteurs de systèmes fortement automatisés peuvent bénéficier du cadre théorique du sentiment d'agentivité, concept issu des sciences cognitives. En effet, ce cadre théorique fournit à la fois des mesures (i.e., explicites et implicites), des principes (comme ceux de Wegner & Wheatley, 1999) et des explications (i.e., intégration des différentes indices) qui peuvent être utiles à la communauté des facteurs humains. Il est important de rappeler que l'intérêt de la thèse ne se situe pas dans l'utilisation de messages amorces, technique largement utilisée à l'heure actuelle en HCI, mais plutôt de fournir un cadre théorique permettant de compléter les explications sur les résultats obtenus lors de l'utilisation de ces amorces.

De nombreux efforts sont encore à faire afin de valider les bénéfices d'un tel cadre théorique en s'intéressant à d'autres modalités d'interactions, à d'autres niveaux d'automatisation (e.g., contrôle mutuel) et à des niveaux d'expertise différents. Toutefois, les premiers résultats sont prometteurs et nous montrent l'importance du respect de certains principes et paramètres (e.g., la temporalité) pour permettre à un opérateur de se sentir en contrôle du système qu'il supervise.

Bibliographie

- Aarts, H., Custers, R., & Wegner, D. M. (2005). On the inference of personal authorship : Enhancing experienced agency by priming effect information. *Consciousness and cognition*, 14(3), 439–458.
- Adler, R. F., & Benbunan-Fich, R. (2012). Juggling on a high wire : Multitasking effects on performance. *International Journal of Human-Computer Studies*, 70(2), 156–168.
- Amalberti, R. (1993). Safety in flight operations. *Reliability and safety hazardous work situations. Approaches to analysis and design*. Lawrence Erlbaum Hove UK, 171–194.
- Amalberti, R. (2013). Human error at the centre of the debate on safety. In *Navigating safety* (pp. 19–52). Springer.
- Annett, J. (2002). Subjective rating scales : science or art ? *Ergonomics*, 45(14), 966–987.
- Arnold, R. (2015). Role of pilot lack of manual control proficiency in air transport aircraft accidents. *Procedia Manufacturing*, 3, 3142–3146.
- Bahner, J. E., Hüper, A.-D., & Manzey, D. (2008). Misuse of automated decision aids : Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies*, 66(9), 688–699.
- Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6), 775–779.
- Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of

- moral disengagement in the exercise of moral agency. *Journal of personality and social psychology*, 71(2), 364.
- Banks, V. A., & Stanton, N. A. (2016). Keep the driver in control : Automating automobiles of the future. *Applied ergonomics*, 53, 389–395.
- Barlas, Z., & Obhi, S. S. (2013). Freedom, choice, and the sense of agency.
- Baron, S. (1988). Pilot control. *Human factors in aviation(A 89-34431 14-54)*. San Diego, CA, Academic Press, Inc., 1988,, 347–385.
- Baxter, G., Besnard, D., & Riley, D. (2007). Cognitive mismatches in the cockpit : Will they ever be a thing of the past ? *Applied ergonomics*, 38(4), 417–423.
- Baxter, G., Rooksby, J., Wang, Y., & Khajeh-Hosseini, A. (2012). The ironies of automation : still going strong at 30 ? In *Proceedings of the 30th european conference on cognitive ergonomics* (pp. 65–71).
- BEA. (2012). *Rapport Final sur l'accident du vol AF 447 Rio de Janeiro - Paris* (Rapport technique). Consulté sur <https://www.bea.aero/fileadmin/documents/docspa/2009/f-cp090601/pdf/f-cp090601.pdf>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate : a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289–300.
- Benjamini, Y., & Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of educational and Behavioral Statistics*, 25(1), 60–83.
- Berberian, B., Le Blaye, P., Schulte, C., Kinani, N., & Sim, P. R. (2013). Data transmission latency and sense of control. In *International conference on engineering psychology and cognitive ergonomics* (pp. 3–12).
- Berberian, B., Sarrazin, J.-C., Le Blaye, P., & Haggard, P. (2012). Automation technology and sense of control : a window on human agency. *PLoS One*, 7(3), e34075.
- Billings, C. (1991). Human-centered aircraft automation : A concept and guidelines.

- Billings, C. (1996). Human-centered aviation automation : Principles and guidelines.
- Billings, C., Lauber, J., Funkhouser, H., Lyman, E., & Huff, E. (1976). Nasa aviation safety reporting system.
- Blakemore, S.-J., & Frith, C. (2003). Self-awareness and action. *Current opinion in neurobiology*, 13(2), 219–224.
- Blakemore, S.-J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? *Neuroreport*, 11(11), R11–R16.
- Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature neuroscience*, 1(7), 635–640.
- Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends in cognitive sciences*, 6(6), 237–242.
- Bliss, J. P., & Acton, S. A. (2003). Alarm mistrust in automobiles : how collision alarm reliability affects driving. *Applied ergonomics*, 34(6), 499–509.
- Borg, J. S., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments : An fmri investigation. *Journal of cognitive neuroscience*, 18(5), 803–817.
- Bratman, M. E. (2007). *Structures of agency : Essays*. Oxford University Press.
- Buehner, M. J., & Humphreys, G. R. (2009). Causal binding of actions to their effects. *Psychological Science*, 20(10), 1221–1228.
- Cardoso-Leite, P., Mamassian, P., Schütz-Bosbach, S., & Waszak, F. (2010). A new look at sensory attenuation action-effect anticipation affects sensitivity, not response bias. *Psychological science*.
- Carmody, M. A., & Gluckman, J. P. (1993). Task specific effects of automation and automation failure on performance, workload and situational awareness. In *Proceedings of the seventh international symposium on aviation psychology* (Vol. 1, pp. 167–171).
- Carpentier, J. (2011). *Cent vingt ans d'innovation en aéronautique*. Hermann.

- Carr, N. (2014). *The Glass Cage : Automation and Us*. W. W. Norton & Company.
- Casner, S. M., & Schooler, J. W. (2013). Thoughts in flight automation use and pilots? task-related and task-unrelated thought. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, 0018720813501550.
- Casner, S. M., & Schooler, J. W. (2015). Vigilance impossible : Diligence, distraction, and daydreaming all lead to failures in a practical monitoring task. *Consciousness and cognition*, 35, 33–41.
- Caspar, E. A., Christensen, J. F., Cleeremans, A., & Haggard, P. (2016). Coercion changes the sense of agency in the human brain. *Current biology*, 26(5), 585–592.
- Chambon, V., & Haggard, P. (2012). Sense of control depends on fluency of action selection, not motor performance. *Cognition*, 125(3), 441–451.
- Chambon, V., Sidarus, N., & Haggard, P. (2015). From action intentions to action effects : how does the sense of agency come about? *Sense of Agency : Examining Awareness of the Acting Self*.
- Chambon, V., Wenke, D., Fleming, S. M., Prinz, W., & Haggard, P. (2013). An online neural substrate for sense of agency. *Cerebral Cortex*, 23(5), 1031–1037.
- Chidambaram, V., Chiang, Y.-H., & Mutlu, B. (2012). Designing persuasive robots : how robots might persuade people using vocal and nonverbal cues. In *Proceedings of the seventh annual acm/ieee international conference on human-robot interaction* (pp. 293–300).
- Chopra, V., Bovill, J., Spierdijk, J., & Koornneef, F. (1992). Reported significant observations during anaesthesia : a prospective analysis over an 18-month period. *British journal of anaesthesia*, 68(1), 13–17.
- Christoffersen, K., & Woods, D. D. (2002). How to make automated systems team players. *Advances in human performance and cognitive engineering research*, 2, 1–12.
- CIAIAC. (2011). *Final Report on the accident of the flight SpanAir*

- 5022 Madrid - Las Palmas (Rapport technique). Consulté le 2016-10-22, sur http://www.fomento.gob.es/NR/rdonlyres/EC47A855-B098-409E-B4C8-9A6DD0D0969F/107087/2008_032_A_ENG.pdf
- Comstock, R., & Arnegard, R. (1992). The multi-attribute task battery for human operator workload and strategic behavior research.
- Cook, M. B., Rieth, C. A., & Ngo, M. K. (2015). Displays for effective human-agent teaming : the role of information availability and attention management. In *International conference on virtual, augmented and mixed reality* (pp. 174–185).
- Cook, M. B., Smallman, H. S., & Rieth, C. A. (2014). Increasing the effective span of control : advanced graphics for proactive, trend-based monitoring. *IIE Transactions on Occupational Ergonomics and Human Factors*, 2(3-4), 137–151.
- Cooper, J., Newbower, R., & Kitz, R. (1984). An analysis of major errors and equipment failures in anesthesia management. *Survey of Anesthesiology*, 28(5), 376–377.
- Corallo, G., Sackur, J., Dehaene, S., & Sigman, M. (2008). Limits on introspection distorted subjective time during the dual-task bottleneck. *Psychological Science*, 19(11), 1110–1117.
- Coyle, D., Moore, J., Kristensson, P. O., Fletcher, P., & Blackwell, A. (2012). I did that ! measuring users' experience of agency in their own actions. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 2025–2034).
- Crivelli, D., & Balconi, M. (2010). Agency and inter-agency, action and joint action : Theoretical and neuropsychological evidence. In *Neuropsychology of the sense of agency* (pp. 107–122). Springer.
- Damen, T. G., van Baaren, R. B., & Dijksterhuis, A. (2014). You should read this ! perceiving and acting upon action primes influences one's sense of agency. *Journal of Experimental Social Psychology*, 50, 21–26.
- DeBruine, L. M. (2002). Facial resemblance enhances trust. *Proceedings of the Royal Society of London B : Biological Sciences*, 269(1498), 1307–1312.

- Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., ... Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, *395*(6702), 597–600.
- Dekker, S. W., & Woods, D. D. (2002). Maba-maba or abracadabra? progress on human–automation co-ordination. *Cognition, Technology & Work*, *4*(4), 240–244.
- Deroo, M., Hoc, J.-M., & Mars, F. (2012). Influence of risk expectation on haptically cued corrective manoeuvres during near lane departure. *Ergonomics*, *55*(4), 465–475.
- Desmond, P. A., & Hoyes, T. W. (1996). Workload variation, intrinsic risk and utility in a simulated air traffic control task : evidence for compensatory effects. *Safety Science*, *22*(1), 87–101.
- DeSteno, D., Breazeal, C., Frank, R. H., Pizarro, D., Baumann, J., Dickens, L., & Lee, J. J. (2012). Detecting the trustworthiness of novel partners in economic exchange. *Psychological science*, 0956797612448793.
- Dewey, J. A., & Knoblich, G. (2014). Do implicit and explicit measures of the sense of agency measure the same thing? *PloS one*, *9*(10), e110118.
- De Winter, J. C., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness : A review of the empirical evidence. *Transportation research part F : traffic psychology and behaviour*, *27*, 196–217.
- Dragan, A. D., Lee, K. C., & Srinivasa, S. S. (2013). Legibility and predictability of robot motion. In *2013 8th acm/ieee international conference on human-robot interaction (hri)* (pp. 301–308).
- DSB. (2010). *Dutch safety board, final Report on the accident of the flight TK1951 Istanbul-Amsterdam* (Rapport technique). Consulté le 2016-10-22, sur <http://www.aaiu.ie/sites/default/files/DSB%20Accident%20%20on%20approach%20TC-JGE%20Amsterdam%20Schipol%202009-02-25.pdf>
- Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

- Eimer, M., & Schlaghecken, F. (1998). Effects of masked stimuli on motor activation : behavioral and electrophysiological evidence. *Journal of Experimental Psychology : Human Perception and Performance*, 24(6), 1737.
- Eimer, M., & Schlaghecken, F. (2003). Response facilitation and inhibition in subliminal priming. *Biological psychology*, 64(1), 7–26.
- Endsley, M. R. (1987). The application of human factors to the development of expert systems for advanced cockpits. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 31, pp. 1388–1392).
- Endsley, M. R. (1996). Automation and situation awareness. *Automation and human performance : Theory and applications*, 163–181.
- Endsley, M. R. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3), 462–492.
- Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, 37(2), 381–394.
- Engbert, K., Wohlschläger, A., & Haggard, P. (2008). Who is causing what? the sense of agency is relational and efferent-triggered. *Cognition*, 107(2), 693–704.
- Ephrath, A. R., & Young, L. R. (1981). Monitoring vs. man-in-the-loop detection of aircraft control failures. In *Human detection and diagnosis of system failures* (pp. 143–154). Springer.
- Eriksson, A., & Stanton, N. A. (2015). When communication breaks down or what was that?—the importance of communication for successful coordination in complex systems. *Procedia Manufacturing*, 3, 2418–2425.
- Faro, D. (2010). Changing the future by reshaping the past : the influence of causal beliefs on estimates of time to onset. *Journal of Consumer Research*, 37(2), 279–291.
- Faro, D., McGill, A. L., & Hastie, R. (2013). The influence of perceived causation on judgments of time : an integrative review and implications for decision-making.

- Farrer, C., Bouchereau, M., Jeannerod, M., & Franck, N. (2008). Effect of distorted visual feedback on the sense of agency. *Behavioural neurology*, *19*(1, 2), 53–57.
- Farrer, C., Franck, N., Georgieff, N., Frith, C. D., Decety, J., & Jeannerod, M. (2003). Modulating the experience of agency : a positron emission tomography study. *Neuroimage*, *18*(2), 324–333.
- Fitts, P. M. (1951). Human engineering for an effective air-navigation and traffic-control system.
- Fitts, P. M., Jones, R. E., & Milton, J. L. (1950). Eye movements of aircraft pilots during instrument-landing approaches. *Ergonomics : Psychological mechanisms and models in ergonomics*, *3*, 56.
- Franck, N., Farrer, C., Georgieff, N., Marie-Cardine, M., Daléry, J., d'Amato, T., & Jeannerod, M. (2001). Defective recognition of one's own actions in patients with schizophrenia. *American Journal of Psychiatry*, *158*(3), 454–459.
- Frith, C. D. (2014). Action, agency and responsibility. *Neuropsychologia*, *55*, 137–142.
- Funk, K., Lyall, B., Wilson, J., Vint, R., Niemczyk, M., Suroteguh, C., & Owen, G. (1999). Flight deck automation issues. *The International Journal of Aviation Psychology*, *9*(2), 109–123.
- Gaines, B. R. (1981). The technology of interaction—dialogue programming rules. *International Journal of Man-Machine Studies*, *14*(1), 133–150.
- Galan, R. (2012). *Si l'aviation m'était contée : Encyclopédie de poche de l'aviation*. Privat.
- Gallagher, S. (2000). Philosophical conceptions of the self : implications for cognitive science. *Trends in cognitive sciences*, *4*(1), 14–21.
- Gallagher, S. (2012). Multiple aspects in the sense of agency. *New Ideas in Psychology*, *30*(1), 15–31.
- Geiselman, E. E., Johnson, C. M., & Buck, D. R. (2013). Flight deck automation invaluable collaborator or insidious enabler ? *Ergonomics in Design : The Quarterly*

- of *Human Factors Applications*, 21(3), 22–26.
- Gentsch, A., Kathmann, N., & Schütz-Bosbach, S. (2012). Reliability of sensory predictions determines the experience of self-agency. *Behavioural brain research*, 228(2), 415–422.
- Gentsch, A., & Schütz-Bosbach, S. (2011). I did it : unconscious expectation of sensory consequences modulates the experience of self-agency and its functional signature. *Journal of Cognitive Neuroscience*, 23(12), 3817–3828.
- Ghazizadeh, M., Lee, J. D., & Boyle, L. N. (2012). Extending the technology acceptance model to assess automation. *Cognition, Technology & Work*, 14(1), 39–49.
- Gillen, M. W. (2010). Diminishing skills? *AeroSafety world*, 5.
- Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). “take over!” how long does it take to get the driver back into the loop? In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 57, pp. 1938–1942).
- Gratch, J., Hill, S., Morency, L.-P., Pynadath, D., & Traum, D. (2015). Exploring the implications of virtual human research for human-robot teams. In *International conference on virtual, augmented and mixed reality* (pp. 186–196).
- Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature neuroscience*, 5(4), 382–385.
- Haggard, P., & Tsakiris, M. (2009). The experience of agency feelings, judgments, and responsibility. *Current Directions in Psychological Science*, 18(4), 242–246.
- Hancock, P. (2014). Automation : how much is too much? *Ergonomics*, 57(3), 449–454.
- Hancock, P., & Verwey, W. B. (1997). Fatigue, workload and adaptive driver systems. *Accident Analysis & Prevention*, 29(4), 495–506.
- Helmreich, R. L., & Foushee, H. C. (1993). *Why crew resource management? empirical and theoretical bases of human factors training in aviation*. Academic Press.
- Hoc, J.-M. (2000). From human–machine interaction to human–machine cooperation. *Ergonomics*, 43(7), 833–843.

- 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.
- Hoc, J.-M., Mars, F., Milleville-Pennel, I., Jolly, É., Netto, M., & Blosseville, J.-M. (2006). Human-machine cooperation in car driving for lateral safety : delegation and mutual control. *Le travail humain*, 69(2), 153–182.
- Horberry, T., Stevens, A., & Regan, M. A. (2014). *Driver acceptance of new technology : Theory, measurement and optimisation*. Ashgate Publishing, Ltd.
- Humphreys, G. R., & Buehner, M. J. (2010). Temporal binding of action and effect in interval reproduction. *Experimental brain research*, 203(2), 465–470.
- Inagaki, T., & Nagai, Y. (2007). Support by warning or by action : Which is appropriate under mismatches between driver intent and traffic conditions ? *IEICE transactions on fundamentals of electronics, communications and computer sciences*, 90(11), 2540–2545.
- Jacob, R., & Karn, K. (2003). Eye Tracking in Human-Computer Interaction and Usability Research : Ready to Deliver the Promises. In J. Hyona, R. Radach, & H. Deubel (Eds.), *The mind's eye : cognitive and applied aspects of eye movement research*. Amsterdam ; Boston : Amsterdam : Elsevier.
- Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration–exploitation trade-off : evidence for the adaptive gain theory. *Journal of cognitive neuroscience*, 23(7), 1587–1596.
- Jones, C. M. (2013). What do we know about high-frequency trading ? *Columbia Business School Research Paper*(13-11).
- Kaber, D. B., & Endsley, M. R. (1997). Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Progress*, 16(3), 126–131.
- Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive

- automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*, 5(2), 113–153.
- Kaber, D. B., Onal, E., & Endsley, M. R. (2000). Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. *Human Factors and Ergonomics in Manufacturing*, 10(4), 409–430.
- Kelso, J. S. (2016). On the self-organizing origins of agency. *Trends in cognitive sciences*.
- Kessel, C. J., & Wickens, C. D. (1982). The transfer of failure-detection skills between monitoring and controlling dynamic systems. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, 24(1), 49–60.
- Kilner, J. M., Vargas, C., Duval, S., Blakemore, S.-J., & Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nature neuroscience*, 7(12), 1299–1301.
- Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems*, 19(6), 91–95.
- Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (2015). Why did my car just do that? explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 9(4), 269–275.
- Kumar, D., & Srinivasan, N. (2013). Hierarchical control and sense of agency : differential effects of control on implicit and explicit measures of agency. In *Proceedings of 35th annual meeting of the cognitive science society*.
- Ky, P., & Miaillier, B. (2006). Sesar : towards the new generation of air traffic management systems in europe. *Journal of Air Traffic Control*, 48(1).
- Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International journal of human-computer studies*, 40(1), 153–184.
- Lee, J. D., & See, K. A. (2004). Trust in automation : Designing for appropriate

- reliance. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, 46(1), 50–80.
- Lee, J. D., & Seppelt, B. D. (2009). Human factors in automation design. In *Springer handbook of automation* (pp. 417–436). Springer.
- Le Goff, K., Rey, A., & Berberian, B. (2015). Toward a model for effective human-automation interaction : The mediated agency. In *International conference on digital human modeling and applications in health, safety, ergonomics and risk management* (pp. 274–283).
- Le Goff, K., Rey, A., Haggard, P., & Berberian, B. (2016). Automation technology and sense of agency. *En préparation*.
- Lester, J. C., Towns, S. G., & Fitzgerald, P. J. (1998). Achieving affective impact : Visual emotive communication in lifelike pedagogical agents. *International Journal of Artificial Intelligence in Education (IJAIED)*, 10, 278–291.
- Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). *Brain*, 106(3), 623–642.
- Lim, S., & Reeves, B. (2010). Computer agents versus avatars : Responses to interactive game characters controlled by a computer or other player. *International Journal of Human-Computer Studies*, 68(1), 57–68.
- Limerick, H., Coyle, D., & Moore, J. W. (2014). The experience of agency in human-computer interactions : a review. *Sense of Agency : Examining Awareness of the Acting Self*.
- Limerick, H., Moore, J. W., & Coyle, D. (2015). Empirical evidence for a diminished sense of agency in speech interfaces. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (pp. 3967–3970).
- Lingnau, A., & Vorberg, D. (2005). The time course of response inhibition in masked priming. *Perception & Psychophysics*, 67(3), 545–557.

- Linser, K., & Goschke, T. (2007). Unconscious modulation of the conscious experience of voluntary control. *Cognition*, *104*(3), 459–475.
- Louw, T., & Merat, N. (2016). A methodology for inducing the out of the loop phenomenon in highly automated driving. In *Presentation at international conference : International conference on traffic and transport psychology, brisbane, australia. doi* (Vol. 10).
- Mackworth, N. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, *1*(1), 6–21.
- Mars, F. (2001). *La désorientation spatiale en aéronautique : apport des neurosciences intégratives à la conception des postes de pilotage* (Thèse de doctorat non publiée). Université de la Méditerranée-Aix-Marseille II.
- Mars, F., Deroo, M., & Hoc, J.-M. (2014). Analysis of human-machine cooperation when driving with different degrees of haptic shared control. *IEEE Transactions on Haptics*, *7*(3), 324–333.
- Marti, S., Sackur, J., Sigman, M., & Dehaene, S. (2010). Mapping introspection's blind spot : Reconstruction of dual-task phenomenology using quantified introspection. *Cognition*, *115*(2), 303–313.
- McEneaney, J. E. (2009). Agency attribution in human-computer interaction. In *International conference on engineering psychology and cognitive ergonomics* (pp. 81–90).
- McEneaney, J. E. (2013). Agency effects in human-computer interaction. *International Journal of Human-Computer Interaction*, *29*(12), 798–813.
- Merat, N., & Jamson, A. H. (2009). How do drivers behave in a highly automated car. In *Proceedings of the 5th international driving symposium on human factors in driver assessment, training and vehicle design* (pp. 514–521).
- Merat, N., Jamson, A. H., Lai, F. C., Daly, M., & Carsten, O. M. (2014). Transition to manual : Driver behaviour when resuming control from a highly automated vehicle. *Transportation research part F : traffic psychology and behaviour*, *27*, 274–282.

- Metzger, U., & Parasuraman, R. (2005). Automation in future air traffic management : Effects of decision aid reliability on controller performance and mental workload. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, 47(1), 35–49.
- Midden, C., Verberne, F., & Ham, J. (2012). Trusting automation technology for safer roads : The effect of shared driving goals. In *Persuasive technology : Design for health and safety; the 7th international conference on persuasive technology* (pp. 57–60).
- Milgram, S. (1963). Behavioral study of obedience. *The Journal of abnormal and social psychology*, 67(4), 371.
- Mohleji, S. C., Lamiano, D. F., & Massimini, S. V. (2009). Air transportation system automation. In *Springer handbook of automation* (pp. 1181–1213). Springer.
- Molloy, R., & Parasuraman, R. (1996). Monitoring an automated system for a single failure : Vigilance and task complexity effects. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, 38(2), 311–322.
- Moore, J. W. (2016). What is the sense of agency and why does it matter? *Frontiers in Psychology*, 7.
- Moore, J. W., & Fletcher, P. (2012). Sense of agency in health and disease : a review of cue integration approaches. *Consciousness and cognition*, 21(1), 59–68.
- Moore, J. W., Middleton, D., Haggard, P., & Fletcher, P. C. (2012). Exploring implicit and explicit aspects of sense of agency. *Consciousness and cognition*, 21(4), 1748–1753.
- Moore, J. W., & Obhi, S. S. (2012). Intentional binding and the sense of agency : a review. *Consciousness and cognition*, 21(1), 546–561.
- Moore, J. W., Wegner, D. M., & Haggard, P. (2009). Modulating the sense of agency with external cues. *Consciousness and cognition*, 18(4), 1056–1064.
- Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. (1998). Automation bias : Deci-

- sion making and performance in high-tech cockpits. *The International journal of aviation psychology*, 8(1), 47–63.
- Nass, C., & Moon, Y. (2000). Machines and mindlessness : Social responses to computers. *Journal of social issues*, 56(1), 81–103.
- Navarro, J., Mars, F., Forzy, J.-F., El-Jaafari, M., & Hoc, J.-M. (2010). Objective and subjective evaluation of motor priming and warning systems applied to lateral control assistance. *Accident Analysis & Prevention*, 42(3), 904–912.
- Navarro, J., Mars, F., & Hoc, J.-M. (2007). Lateral control assistance for car drivers : a comparison of motor priming and warning systems. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, 49(5), 950–960.
- Nielsen, J., & Levy, J. (1994). Measuring usability : preference vs. performance. *Communications of the ACM*, 37(4), 66–75.
- Norman, D. A. (1990). The 'problem' with automation : inappropriate feedback and interaction, not 'over-automation'. *Philosophical Transactions of the Royal Society of London B : Biological Sciences*, 327(1241), 585–593.
- Norman, D. A. (2013). *The design of everyday things : Revised and expanded edition*. Basic books.
- Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence*, 12(5), 481–494.
- NTSB. (1986). *Final Report on the accident of the flight China Airlines 006 Taipei - Los Angeles* (Rapport technique).
- NTSB. (2010). *Loss of Control on Approach, Colgan Air, Inc., Operating as Continental Connection Flight 3407, Bombardier DHC-8-400, N200wq, Clarence Center, New York, February 12, 2009*. (Rapport technique N° NTSB/AAR-10/01). Washington, DC. Consulté le 2016-08-29, sur <http://www.nts.gov/investigations/AccidentReports/Reports/AAR1001.pdf>

- Obhi, S. S., & Hall, P. (2011). Sense of agency in joint action : Influence of human and computer co-actors. *Experimental brain research*, 211(3-4), 663–670.
- Pacherie, E. (2007). The sense of control and the sense of agency. *Psyche*, 13(1), 1–30.
- Pacherie, E. (2008). The phenomenology of action : A conceptual framework. *Cognition*, 107(1), 179–217.
- Pacherie, E. (2011). 14 the phenomenology of joint action : Self-agency versus joint agency. *Joint attention : New developments in psychology, philosophy of mind, and social neuroscience*, 343.
- Parasuraman, R. (1979). Memory load and event rate control sensitivity decrements in sustained attention. *Science*, 205(4409), 924–927.
- Parasuraman, R., & Davies, D. (1977). A taxonomic analysis of vigilance performance. In *Vigilance* (pp. 559–574). Springer.
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation : An attentional integration. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, 52(3), 381–410.
- Parasuraman, R., Molloy, R., & Hilburn, B. (1993). Adaptive function allocation reduces performance cost of static automation. In *7th international symposium on aviation psychology* (pp. 37–42).
- Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced 'complacency'. *The International Journal of Aviation Psychology*, 3(1), 1–23.
- 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.
- 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.

- Parasuraman, R., & Wickens, C. D. (2008). Humans : Still vital after all these years of automation. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, 50(3), 511–520.
- Payre, W., Cestac, J., & Delhomme, P. (2016). Fully automated driving impact of trust and practice on manual control recovery. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, 58(2), 229–241.
- Plat, M., & Amalberti, R. (2000). Experimental crew training to deal with automation surprises. *Cognitive engineering in the aviation domain. Hillsdale (NJ) : Lawrence Erlbaum Associates*, 287–308.
- Poonian, S., & Cunnington, R. (2013). Intentional binding in self-made and observed actions. *Experimental brain research*, 229(3), 419–427.
- Rasmussen, J. (1986). *Information processing and human-machine interaction : An approach to cognitive engineering*. New York, NY, USA : Elsevier Science Inc.
- Reason, J. (1990). *Human error*. Cambridge university press.
- Reason, J. (2000). Human error : models and management. *Bmj*, 320(7237), 768–770.
- Reeves, B., & Nass, C. (1996). *How people treat computers, television, and new media like real people and places*. CSLI Publications and Cambridge university press Cambridge, UK.
- Roussel, C. (2013). *A preactivation theory of action effect prediction* (Thèse de doctorat non publiée). Paris 5.
- Rudisill, M. (1995). Line pilots' attitudes about and experience with flight deck automation : Results of an international survey and proposed guidelines.
- Rushby, J. (2011). New challenges in certification for aircraft software. In *Proceedings of the ninth acm international conference on embedded software* (pp. 211–218).
- Santiago-Espada, Y., Myer, R. R., Latorella, K. A., & Comstock, J. R. (2011). The multi-attribute task battery ii (matb-ii) software for human performance and workload research : A user's guide.

- Sarter, N. B., & Woods, D. D. (1995). How in the world did we ever get into that mode? mode error and awareness in supervisory control. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, 37(1), 5–19.
- Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. *Handbook of human factors and ergonomics*, 2, 1926–1943.
- Sato, A. (2009). Both motor prediction and conceptual congruency between preview and action-effect contribute to explicit judgment of agency. *Cognition*, 110(1), 74–83.
- Sato, A., & Yasuda, A. (2005). Illusion of sense of self-agency : discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition*, 94(3), 241–255.
- Sauer, J., Chavallaz, A., & Wastell, D. (2015). Experience of automation failures in training : effects on trust, automation bias, complacency and performance. *Ergonomics*, 1–14.
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action : bodies and minds moving together. *Trends in cognitive sciences*, 10(2), 70–76.
- Sebanz, N., & Frith, C. (2004). Beyond simulation? neural mechanisms for predicting the actions of others. *Nature neuroscience*, 7(1), 5–6.
- See, J. E., Howe, S. R., Warm, J. S., & Dember, W. N. (1995). Meta-analysis of the sensitivity decrement in vigilance. *Psychological Bulletin*, 117(2), 230.
- Sheehan, J. J., & Sosna, M. (1991). *The boundaries of humanity : Humans, animals, machines*. Univ of California Press.
- Sheridan, T. B., & Parasuraman, R. (2005). Human-automation interaction. *Reviews of human factors and ergonomics*, 1(1), 89–129.
- Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators* (Rapport technique). DTIC Document.
- Shneiderman, B., & Plaisant, C. (2004). *Designing the User Interface : Strategies for Effective Human-Computer Interaction* (Pearson Addison-Wesley éd.). Reading,

MA.

- Sidarus, N., Chambon, V., & Haggard, P. (2013). Priming of actions increases sense of control over unexpected outcomes. *Consciousness and cognition*, 22(4), 1403–1411.
- Sidarus, N., & Haggard, P. (2016). Difficult action decisions reduce the sense of agency : A study using the eriksen flanker task. *Acta psychologica*, 166, 1–11.
- Stanton, N. A., Dunoyer, A., & Leatherland, A. (2011). Detection of new in-path targets by drivers using stop & go adaptive cruise control. *Applied ergonomics*, 42(4), 592–601.
- Stanton, N. A., Young, M. S., Walker, G. H., Turner, H., & Randle, S. (2001). Automating the driver's control tasks. *International Journal of Cognitive Ergonomics*, 5(3), 221–236.
- Synofzik, M., Vosgerau, G., & Lindner, A. (2009). Me or not me—an optimal integration of agency cues? *Consciousness and cognition*, 18(4), 1065–1068.
- Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model : a multifactorial two-step account of agency. *Consciousness and cognition*, 17(1), 219–239.
- Synofzik, M., Vosgerau, G., & Voss, M. (2013). The experience of agency : an interplay between prediction and postdiction.
- Synofzik, M., & Voss, M. (2010). Disturbances of the sense of agency in schizophrenia. In *Neuropsychology of the sense of agency* (pp. 145–155). Springer.
- Taleb, N. N. (2012). *Antifragile : Things that gain from disorder* (Vol. 3). Random House Incorporated.
- Teichner, W. H. (1974). The detection of a simple visual signal as a function of time of watch. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, 16(4), 339–352.
- van der Wel, R. P. (2015). Me and we : Metacognition and performance evaluation of joint actions. *Cognition*, 140, 49–59.

- Van Veen, H. A., & Van Erp, J. B. (2001). Tactile information presentation in the cockpit. In *Haptic human-computer interaction* (pp. 174–181). Springer.
- Venkatesh, V. (2000). Determinants of perceived ease of use : Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information systems research*, 11(4), 342–365.
- Verberne, F. M., Ham, J., & Midden, C. J. (2012). Trust in smart systems sharing driving goals and giving information to increase trustworthiness and acceptability of smart systems in cars. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, 54(5), 799–810.
- Verberne, F. M., Ham, J., & Midden, C. J. (2015). Trusting a virtual driver that looks, acts, and thinks like you. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, 0018720815580749.
- Verfaillie, K., & Daems, A. (2002). Representing and anticipating human actions in vision. *Visual Cognition*, 9(1-2), 217–232.
- Verney, S. P., Granholm, E., & Marshall, S. P. (2004). Pupillary responses on the visual backward masking task reflect general cognitive ability. *International Journal of Psychophysiology*, 52(1), 23–36.
- Vesper, C., van der Wel, R. P., Knoblich, G., & Sebanz, N. (2011). Making oneself predictable : Reduced temporal variability facilitates joint action coordination. *Experimental brain research*, 211(3-4), 517–530.
- Wegner, D. M. (2004). Précis of the illusion of conscious will. *Behavioral and Brain Sciences*, 27(05), 649–659.
- Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency : experiencing control over the movements of others. *Journal of personality and social psychology*, 86(6), 838.
- Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation : Sources of the experience of will. *American psychologist*, 54(7), 480.

- Wen, W., Yamashita, A., & Asama, H. (2015). The sense of agency during continuous action : performance is more important than action-feedback association. *PloS one*, *10*(4), e0125226.
- Wenke, D., Fleming, S. M., & Haggard, P. (2010). Subliminal priming of actions influences sense of control over effects of action. *Cognition*, *115*(1), 26–38.
- Wickens, C. D., Clegg, B. A., Vieane, A. Z., & Sebok, A. L. (2015). Complacency and automation bias in the use of imperfect automation. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, 0018720815581940.
- Wiener, E. L. (1988). Cockpit automation.
- Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation : Promises and problems. *Ergonomics*, *23*(10), 995–1011.
- Wilcox, R. (2010). *Fundamentals of modern statistical methods : Substantially improving power and accuracy*. Springer Science & Business Media.
- Wilcox, R. (2011). *Modern statistics for the social and behavioral sciences : A practical introduction*. CRC press.
- Wilke, C., Synofzik, M., & Lindner, A. (2012). The valence of action outcomes modulates the perception of one's actions. *Consciousness and cognition*, *21*(1), 18–29.
- Wohlschläger, A., Haggard, P., Gesierich, B., & Prinz, W. (2003). The perceived onset time of self-and other-generated actions. *Psychological Science*, *14*(6), 586–591.
- Woods, D., & Dekker, S. (2000). Anticipating the effects of technological change : a new era of dynamics for human factors. *Theoretical issues in ergonomics science*, *1*(3), 272–282.
- Woods, D., Dekker, S., Cook, R., Johannesen, L., & Sarter, N. (2012). *Behind Human Error*. Ashgate Publishing, Ltd.
- Woods, D., & Tinapple, D. (1999). W3 : Watching human factors watch people at work. In *Presidential address, presented at the 43rd annual meeting of the human factors and ergonomics society, houston, tx. multimedia production available at <http://csel>*.

eng. ohiostate. edu/hf99/.

- Wright, D., Mackenzie, S., Buchan, I., Cairns, C., & Price, L. (1991). Critical incidents in the intensive therapy unit. *The Lancet*, *338*(8768), 676–678.
- Xiong, H., Boyle, L. N., Moeckli, J., Dow, B. R., & Brown, T. L. (2012). Use patterns among early adopters of adaptive cruise control. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, *54*(5), 722–733.
- Yoshie, M., & Haggard, P. (2013). Negative emotional outcomes attenuate sense of agency over voluntary actions. *Current Biology*, *23*(20), 2028–2032.
- Young, M., Brookhuis, K., Wickens, C., & Hancock, P. (2015). State of science : mental workload in ergonomics. *Ergonomics*, *58*(1), 1–17.
- Young, M., & Stanton, N. A. (2007). What's skill got to do with it? vehicle automation and driver mental workload. *Ergonomics*, *50*(8), 1324–1339.
- Zimmermann, M., Bauer, S., Lutteken, N., Rothkirch, I. M., & Bengler, K. J. (2014). Acting together by mutual control : Evaluation of a multimodal interaction concept for cooperative driving. In *Collaboration technologies and systems (cts), 2014 international conference on* (pp. 227–235).

Annexes

Publications liées à une communication

Prédictibilité des systèmes et sentiment de contrôle : vers des interfaces plus agentives
EPIQUE 2015 Aix-en-Provence.

Toward a Model for Effective Human-Automation Interaction : The Mediated Agency
HCI 2015 Los Angeles.

ÉPIQUE 2015

Prédicibilité des systèmes et sentiment de contrôle : vers des interfaces plus agentives

Kevin Le Goff^{1,2}

kevin.le_goff@onera.fr

Arnaud Rey²

arnaud.rey@univ-amu.fr

Bruno Berberian¹

Bruno.Berberian@onera.fr

¹ ONERA Salon de Provence

Département Commande des Systèmes et Dynamique du vol

BA 701 FR-13661 Salon Cedex AIR

²Laboratoire de Psychologie Cognitive UMR 7290 Aix-Marseille Université - CNRS

Bâtiment 9 Case D

3, place Victor Hugo, 13331 Marseille Cedex3

RÉSUMÉ

Cet article a pour objectif de proposer à la communauté de l'ergonomie un nouveau cadre théorique issu des neurosciences : l'agentivité. Ce concept correspond au sentiment que nous ressentons lorsque nous sommes en contrôle de nos actions. Nous défendons l'idée selon laquelle les principes de l'agentivité développés en laboratoire pourraient compléter les solutions existantes afin de concevoir des agents artificiels plus coopératifs. Nous illustrerons notre propos à travers la présentation de résultats mettant en évidence l'intérêt de l'application de ce concept à la problématique de la supervision de systèmes fortement automatisés.

Mots clefs : Automatisation, Prédicibilité, Agentivité, Contrôle, Acceptabilité

1 INTRODUCTION

Au cours des 50 dernières années, les mutations technologiques à l'œuvre ont profondément changé l'interaction de l'homme avec son environnement. L'influence de l'automatisation est aujourd'hui perçue dans tous les aspects de la vie quotidienne et non plus seulement dans le monde de l'industrie. Que ce soit au travail ou à domicile, nous sommes habitués à interagir avec des systèmes informatiques sophistiqués conçus pour nous assister dans nos activités. Cette automatisation a permis de rendre certains aspects de la vie plus *faciles* : en permettant aux personnes en situation de handicap d'être en mesure de se déplacer et de communiquer ; plus *rapides* : avec la généralisation de dispositifs informatisés et l'augmentation de la productivité ; et plus *sûrs* : le taux d'accidents dans l'aviation et l'industrie a baissé grâce à la mise en œuvre de ces systèmes (Norman, 1990). L'ingéniosité de ces systèmes associée aux bénéfices économiques perçus par la société masque cependant un problème fondamental :

l'interposition de systèmes automatisés entre les opérateurs et les différentes tâches à effectuer a profondément modifié l'activité de l'opérateur humain et ses modalités d'interaction avec l'environnement. Ce changement, loin d'être trivial, crée de nouvelles problématiques pour les personnes chargées de l'exploitation, du dépannage et de la gestion de tels systèmes.

2 AUTOMATISATION ET PERFORMANCE

L'automatisation des systèmes est classiquement considérée comme une simple substitution d'une activité humaine par la machine, raccourci nommé "*substitution myth*" (Woods & Tinapple, 1999). L'idée véhiculée par un tel raccourci est celle d'un statu quo qualitatif vis-à-vis de l'activité de l'opérateur humain. Le seul changement au regard de l'activité humaine serait d'ordre quantitatif, l'automatisation délestant l'opérateur d'une partie du travail à effectuer sans fondamentalement modifier la nature de son activité. Dans ce contexte, il ne fut pas surprenant d'observer que l'automatisation des systèmes pouvait en effet diminuer la charge de travail des opérateurs humains et augmentait leur productivité (Kaber, Onal, & Endsley, 2000).

Pourtant, de nombreux travaux semblent contredire cette affirmation et soulignent les conséquences négatives en matière de sécurité et de performance de l'approche traditionnelle de l'automatisation : diminution de la vigilance de l'opérateur (i.e., manque de sensibilité de l'opérateur à des signaux, Billings, 1991 ; Wiener, 1989), augmentation de la complaisance¹ envers le système (Parasuraman, Molloy, & Singh, 1993 ; Singh, Molloy, & Parasuraman, 1997) et diminution de la conscience de situation ("*situation awareness*", Endsley, 1996). Ce phénomène, mieux connu sous le terme de sortie de la boucle de contrôle (« *Out Of the Loop performance problem* », Endsley & Kiris, 1995 ; Kaber & Endsley, 1997), se traduit par une série de difficultés rencontrées par l'opérateur, difficultés particulièrement visibles lors de situations d'urgence, notamment lors d'une défaillance du système automatisé nécessitant une reprise en main par l'opérateur humain. Parmi ces difficultés, on peut citer un temps de réaction plus élevé pour diagnostiquer une panne, décider si une intervention est nécessaire et pour déterminer les étapes à suivre pour rétablir la situation (Billings, 1991).

Ce problème d'efficacité lors des situations de reprise en main est un point critique dans la conception de tels systèmes. Au regard de leur caractère imparfait (la fiabilité absolue reste une chimère pour l'heure), les automatismes requièrent non seulement une supervision constante par l'opérateur humain, mais également la capacité pour cet opérateur de reprendre en main le système lors d'une situation critique. L'opérateur humain constitue alors le dernier rempart en cas d'évènements imprévus. On se retrouve ainsi dans une situation pour le moins incohérente, dans laquelle les systèmes sont conçus d'une manière telle qu'ils engendrent une sortie de la boucle de contrôle de l'opérateur, diminuant ses capacités de reprise en main, tout en considérant ce même opérateur comme le garant ultime de la sécurité de ce système dans les situations les plus critiques auxquelles la technologie elle-même est incapable de répondre. En outre, l'augmentation croissante de l'autonomie de ces systèmes et de leur efficacité rendent ces situations critiques de plus en plus rares, diminuant ainsi l'occurrence de l'intervention des opérateurs humains, et renforçant par la même leur incapacité à réagir à un évènement imprévu par manque d'expérience et de pratique de ces situations. Enfin, la démocratisation de ces technologies (arrivée prochaine de véhicules complètement automatisés sur nos routes) semble promettre un essor important de cette problématique dans les prochaines décennies.

¹ Le phénomène de complaisance consiste à accorder à l'assistance automatique une confiance excessive au regard de sa fiabilité. L'opérateur a alors tendance à diminuer son degré de vigilance dans la tâche.

3. APPROCHES CLASSIQUES DU PHENOMENE

Une approche classique de ce phénomène consiste à étudier l'impact des caractéristiques et limites de l'opérateur humain sur l'utilisation des automatismes. Au regard des difficultés rencontrées par l'opérateur, il est classiquement accepté que l'Humain est peu adapté aux tâches de monitoring (Endsley, 1996 ; Parasuraman & Riley, 1997) et différentes méthodes ont été proposées pour compenser ces difficultés. On citera par exemple la méthode d'entraînement aux aléas, qui consiste à exposer l'opérateur humain à des situations nécessitant une reprise en main lors de phases d'entraînement permettant notamment de diminuer le niveau de complaisance de l'opérateur vis-à-vis de l'automatisme (Bahner, Hüper, & Manzey, 2008; Plat & Amalberti, 2000). La solution la plus utilisée reste cependant celle du partage d'autorité (méthodes MABA-MABA², "*adaptive function allocation*", voir notamment Hollnagel, 2000 ; Parasuraman, Mouloua, & Molloy, 1996). Cette méthode consiste à définir l'allocation statique (ou dynamique) des différentes fonctions à réaliser entre le ou les opérateurs et les systèmes de contrôle. Elle permet ainsi de manipuler le niveau d'automatisation des systèmes (initiée par Sheridan & Verplanck, 1978) et de partager les tâches à effectuer entre l'automatisme et l'opérateur. Cette méthodologie repose sur l'idée que le partage des tâches aura pour résultat de garder l'opérateur impliqué dans les actions du système, permettant ainsi d'améliorer les performances et la conscience de situation de cet opérateur par rapport à une simple supervision d'un système fortement automatisé (Endsley 1987 ; Endsley and Kiris, 1995). Depuis, des données empiriques sont venues confirmer cette hypothèse (Endsley, 1999 ; Kaber & Endsley, 2004 ; Kaber, Onal, & Endsley, 2000).

Si ces méthodes permettent sans conteste de limiter l'impact du phénomène de sortie de boucle, de nombreux chercheurs soulignent la nécessité de dépasser cette simple approche quantitative du problème posé par l'automatisation des systèmes (voir notamment Dekker et Woods, 2002). Dès 1990, Norman soulignait que le problème n'est pas tant la présence de l'automatisation mais plutôt son design inapproprié (voir également Hollnagel, 1999) pointant quelques années plus tard la nécessité de socialiser nos interactions avec les automatismes (Norman, 2010). Ajouter ou étendre le rôle de la machine, c'est avant tout introduire un nouvel agent dans la boucle de contrôle. Autrement dit, l'automatisation des systèmes peut être assimilée à l'ajout d'un nouveau « membre d'équipage », impliquant de nouvelles demandes en termes de coordination pour l'opérateur humain (Sarter, Woods, & Billings, 1997). De ce constat est née une seconde approche, plus récente et complémentaire à la première, qui s'est focalisée sur la coopération entre l'Homme et la Machine. La question posée par cette approche n'est plus "qui a le contrôle sur quoi et à quel point?" mais "comment pouvons-nous réaliser correctement cette tâche ensemble?". Le point crucial pour les designers est alors d'arriver à réaliser des interfaces permettant une coordination efficace entre l'opérateur et l'automatisme, pas seulement lors de situations standards et facilement prédictibles mais également lors de circonstances imprévues. Depuis, de nombreuses études se sont penchées sur la question de comment transformer l'automatisme en agent collaboratif (Christoffersen & Woods, 2002 ; Dragan, Lee, & Srinivasa, 2013 ; Hoc, 2007 ; Hoc & Carlier, 2002 ; Klein et al., 2004 ; Zimmermann et al., 2014). Nous souhaitons participer à cet effort à travers l'introduction des avancées récentes relatives aux mécanismes sous-tendant le contrôle et la compréhension de nos propres actions. En particulier, nous défendons l'idée que l'opérateur interprète les intentions et le

² MABA-MABA méthode - "Men are Better At – Machines Are Better At" - Méthode basée sur l'identification des « forces » et « faiblesses » supposées de chacun des agents (humain et machine) afin d'allouer à chacun les tâches qui lui conviendront le mieux.

résultat des actions produit par un système à travers son propre « outillage cognitif ». Dès lors, la compréhension de cet « outillage » pourrait permettre de proposer des principes de conception permettant de concevoir des systèmes plus coopératifs, et donc plus contrôlables.

4. CONCEVOIR DES AGENTS COLLABORATIFS : L'APPORT DE L'AGENTIVITE

Lorsque nous agissons, nous avons un sentiment clair que nous contrôlons nos propres actions et que nous pouvons ainsi agir sur l'environnement extérieur. Ce sentiment est décrit comme le "sentiment d'agentivité" ("*sense of agency*", Gallagher, 2000). Ce sentiment nous permet de discriminer naturellement les actions dont nous sommes les auteurs de celles dont nous ne le sommes pas. De nombreuses études en psychologie (Aarts, Custers, & Wegner, 2005 ; Moore, Wegner, & Haggard, 2009), psychopathologie (Franck et al., 2001) et neurosciences (Farrer et al., 2003) se sont penchées sur les mécanismes permettant de s'attribuer ou non une pensée, des intentions ou des actions. Bien que ces mécanismes ne soient pas encore parfaitement compris, les différentes approches proposent que nous dérivions un sentiment d'agentivité grâce à des processus cognitifs comparant les conséquences prédites de nos actions avec le résultat effectif de ces mêmes actions. De manière intéressante, Élisabeth Pacherie (2012) a récemment introduit l'idée que les différents mécanismes sous-tendant le sentiment d'agentivité pour des actions individuelles sont du même ordre que ceux sous-tendant des actions effectuées en coopération. L'agentivité des actions coproduites serait donc basée sur le même principe de congruence entre les conséquences prédites et réelles des actions et notre capacité à prédire les actions de nos partenaires et leurs effets.

Nous pouvons donc imaginer que de la même manière que deux personnes travaillent ensemble, un superviseur doit être capable de prédire les actions de l'automatisme et leurs effets afin de faciliter leur coopération. Cette proposition fait écho à celle de Norman (1990) lorsqu'il explique qu'il est nécessaire d'avoir un feedback continu sur l'état du système de la même façon que deux personnes discutent entre elles lors d'une co-action. Or, les systèmes automatisés tendent à développer une cascade d'actions qui va décroître voire même éliminer cette prédictibilité, ce que Taleb appelle l'*opacité* des systèmes (Taleb, 2012). Celle-ci se traduit par une difficulté pour l'opérateur humain à percevoir les intentions du système, à comprendre son état courant et plus encore à anticiper ses états futurs. Cette opacité, déjà identifiée comme un élément central du phénomène de sortie de boucle par Norman (1990), modifie fortement notre expérience de l'agentivité. Berberian, Sarrazin, Le Blaye et Haggard (2012) ont récemment confirmé de façon empirique cette dégradation. En manipulant le niveau d'automatisation dans une tâche de supervision de pilotage d'avion, ils montrent une diminution de l'agentivité (grâce à des mesures implicite et explicite) concomitante avec une augmentation de l'automatisation. Dans ce contexte, réduire l'opacité des systèmes constitue une étape clef dans la conception d'agents artificiels plus coopératifs.

5. AGENTIVITE ET DESIGN: ILLUSTRATIONS

Notre but est d'appliquer le concept d'agentivité au domaine des interactions homme-machine et de prendre en compte le rôle des informations fournies par les systèmes automatisés dans la compréhension et le contrôle du système par l'opérateur. Comme indiqué précédemment, l'opacité du système est certainement une des causes majeures du problème de sortie de boucle. Ce constat établi (et largement partagé), la difficulté réside dans l'élaboration de principes d'échanges permettant de diminuer cette opacité : quelles sont les informations nécessaires à l'opérateur, sous quelle forme, à quel moment ? Nous faisons l'hypothèse que l'application du cadre théorique de l'agentivité permettra d'atteindre plus facilement cet objectif.

La capacité de prédiction apparaît comme un élément central dans l'élaboration du sentiment de contrôle, qu'il s'agisse d'action individuelle ou d'action conjointe (Verfaillie & Daems, 2002 ; Vesper, van der Wel, Knoblich, & Sebanz, 2011). Cette capacité de prédiction repose elle-même sur un concept central, celui de l'intentionnalité. De récentes études en neurosciences (Chambon et al., 2014 ; Wenke, Fleming, & Haggard, 2010) montrent notamment que l'agentivité dépend de la facilité avec laquelle la réponse motrice est sélectionnée au niveau cérébral. Une fluidité au niveau de la sélection de la réponse traduit la congruence entre les intentions de l'opérateur et l'action qu'il effectue, ce qui a pour effet d'augmenter le sentiment de contrôle. En ce sens, rendre accessible les intentions du système nous apparaît essentiel en vue d'optimiser le contrôle exercé par l'opérateur humain sur les systèmes technologiques. Pour ce faire, nous proposons de nous appuyer sur la théorie de l' "*apparent mental causation*" de Daniel Wegner (voir Wegner et Weatherley, 1999) qui fournit des pistes pour déterminer la nature, la forme et le timing d'un feedback approprié. Particulièrement, Wegner et Weatherley proposent trois principes permettant de "faire naître un sentiment d'agentivité". Le principe de *priorité*, de *consistance* et d'*exclusivité* traduisent le fait que lorsque nous avons une pensée (ou intention) qui émerge avant une action, que cette pensée est consistante avec notre action et que l'action n'a pas d'autres causes alternatives possibles, alors nous ressentons le sentiment d'être l'auteur de cette action. Afin de simuler la priorité de ces pensées (ou intention), l'effet d'amorçage a souvent été utilisé. En particulier, de nombreux travaux ont montré que l'utilisation d'amorce congruente avec l'action observée entraîne une augmentation du sentiment de contrôle (voir notamment Aarts, Custers & Wegner, 2005 ; Sato, 2009).

Nous proposons d'utiliser cet effet d'amorçage dans le cadre de la supervision. Nous faisons l'hypothèse que dans un tel cadre, l'utilisation d'amorces permettrait de rendre les opérateurs "plus agents" et ainsi d'être moins affectés par ce phénomène de sortie de boucle de contrôle. Ainsi, ils seront plus rapides et plus confiants lors de reprise en main de systèmes automatisés en cas de panne. Nous avons donc réalisé deux expérimentations au cours de laquelle les participants devaient réaliser une tâche de supervision de pilotage d'avion (Le Goff, Haggard, Rey, & Berberian, submitted). Des messages "amorces" indiquant à l'opérateur quelle action allait réaliser le système (cap à gauche ou à droite) sont utilisés. Au regard des principes proposés par Wegner, les principes de priorité et de consistance sont respectés. De plus, de telles amorces devaient accroître la prédictibilité du système à superviser et ainsi augmenter le sentiment de contrôle de l'opérateur. Nous nous attendions donc à observer une augmentation du sentiment d'agentivité à travers des mesures subjectives et objectives traduisant une reconstitution de la boucle de contrôle et son efficacité à travers une meilleure reprise en main si nécessaire. Dans une première expérience, nous montrons que l'utilisation de ces amorces permet au sujet de se sentir subjectivement (reports verbaux) plus en contrôle du système à superviser. Dans la seconde expérience, nous avons testé si les sujets étaient objectivement plus performants lors de phase de reprise en main. Les résultats montrent effectivement que les sujets sont plus rapides et plus efficaces pour détecter une erreur du système. Des données oculométriques démontrent en outre une plus grande disponibilité des ressources cognitives pour réaliser une tâche en parallèle en présence de ces amorces. Ces premiers résultats montrent ainsi très clairement l'apport des principes de l'agentivité pour la conception d'interfaces plus prédictives et l'impact de cette prédictibilité sur la performance de reprise en main du système en cas d'erreur.

Le bénéfice de cet amorçage peut être rapproché des résultats déjà obtenus dans l'étude des interactions homme système pour des tâches de contrôle direct (Berberian & al. 2013 ; Koo, & al., 2014 ; Navarro & al., 2007 ; 2010). Les travaux de Navarro et al. (2007, 2010) ont

notamment montré comment le "*motor priming*" pouvait améliorer significativement la conduite des participants. De manière intéressante, ces travaux ont également montré une dissociation entre l'augmentation de la performance et la diminution de l'acceptabilité. Ce résultat rejoint ceux obtenus par Damen et al. (2014), qui ont montré que les amorces subliminales augmentaient effectivement le niveau d'agentivité mais pas les amorces supraliminales qui peuvent être perçues comme intrusives. Le "*motor priming*" serait donc plus efficace au niveau implicite car il permettrait une sélection plus rapide de l'action à effectuer (Chambon et al., 2014 ; Wenke et al., 2010) mais serait moins acceptable au niveau explicite car il peut être considéré comme trop intrusif par le conducteur. Ainsi, l'agentivité pourrait donner un cadre explicatif aux résultats obtenus par Navarro et al..

5 CONCLUSION

Nous avons vu qu'en dépit de tous les avantages de l'automatisation, les difficultés rencontrées par l'opérateur humain à superviser et reprendre en main de tels systèmes demeurent un problème majeur pour notre société. Il est clairement établi que certaines de ces difficultés résultent d'une diminution de la prédictibilité des interfaces automatisées due à un accroissement de leur complexité. La création d'interfaces plus facilement contrôlables n'est pas si aisée et reste encore aujourd'hui un challenge dans le monde de l'interaction homme-système. Nous proposons un cadre théorique, celui de l'agentivité, permettant de répondre à cette problématique. L'utilisation des principes, outils et mesures issus de ce cadre théorique nous semble bénéfique à différents égards (1) pour évaluer l'influence de l'automatisation des systèmes sur le sentiment de contrôle de l'opérateur, (2) pour établir des recommandations concrètes sur la façon de concevoir des interfaces maintenant l'opérateur dans la boucle de contrôle. Ainsi bien que de nombreuses solutions ont déjà été proposées pour améliorer le contrôle des interfaces, comme l'aide à l'anticipation ou encore une anthropomorphisation des systèmes, l'utilisation du cadre théorique de l'agentivité permettra d'évaluer leur efficacité, de les optimiser et de fournir un cadre explicatif de résultats déjà présents dans la littérature sur les interactions homme-système. Enfin, une meilleure compréhension de l'évolution de l'agentivité dans le cas des interactions avec des automatismes pourra certainement nous aider à raffiner nos modèles du sentiment d'agentivité et plus généralement nos modèles du contrôle.

6 BIBLIOGRAPHIE

- Aarts, H., Custers, R., & Wegner, D. M. (2005). On the inference of personal authorship: Enhancing experienced agency by priming effect information. *Consciousness and cognition*, 14(3), 439-458.
- Bahner, J. E., Hüper, A. D., & Manzey, D. (2008). Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies*, 66(9), 688-699.
- Berberian, B., Sarrazin, J. C., Le Blaye, P., & Haggard, P. (2012). Automation technology and sense of control: a window on human agency. *PLoS one*, 7(3).
- Berberian, B., Le Blaye, P., Schulte, C., Kinani, N., & Sim, P. R. (2013). Data transmission latency and sense of control. In *Engineering Psychology and Cognitive Ergonomics. Understanding Human Cognition* (pp. 3-12). Springer Berlin Heidelberg.
- Billings, C. E. (1991). Human-centered aircraft automation: A concept and guidelines. (NASA Tech. Memo. 103885). Moffet Field, CA: NASA-Ames Research Center.

- Chambon, V., Wenke, D., Fleming, S. M., Prinz, W., & Haggard, P. (2013). An online neural substrate for sense of agency. *Cerebral Cortex*, 23(5), 1031-1037.
- Christoffersen, K., & Woods, D. D. (2002). How to make automated systems team players. *Advances in human performance and cognitive engineering research*, 2, 1-12.
- Damen, T. G., van Baaren, R. B., & Dijksterhuis, A. (2014). You should read this! Perceiving and acting upon action primes influences one's sense of agency. *Journal of Experimental Social Psychology*, 50, 21-26.
- Dekker, S. W., & Woods, D. D. (2002). MABA-MABA or abracadabra? Progress on human-automation co-ordination. *Cognition, Technology & Work*, 4(4), 240-244.
- Dragan, A. D., Lee, K. C., & Srinivasa, S. S. (2013). Legibility and predictability of robot motion. In *Human-Robot Interaction (HRI), 2013 8th ACM/IEEE International Conference on* (pp. 301-308). IEEE.
- Endsley, M. R. (1987). The application of human factors to the development of expert systems for advanced cockpits. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 31(12), 1388-1392.
- Endsley, M. R. (1996). Automation and situation awareness. In Parasuraman, R., Mouloua, M. (Eds.) *Automation and Human Performance: Theory and Applications* (pp. 163–181). Mahwah, N.J.: Lawrence Erlbaum,
- Endsley, M. R. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3), 462-492.
- Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(2), 381-394.
- Farrer, C., Franck, N., Georgieff, N., Frith, C. D., Decety, J., & Jeannerod, M. (2003). Modulating the experience of agency: a positron emission tomography study. *Neuroimage*, 18(2), 324-333.
- Franck, N., Farrer, C., Georgieff, N., Marie-Cardine, M., Daléry, J., d'Amato, T., & Jeannerod, M. (2001). Defective recognition of one's own actions in patients with schizophrenia. *American Journal of Psychiatry*, 158(3), 454-459.
- Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in cognitive sciences*, 4(1), 14-21.
- Hoc, J. M. (2007). Human and automation: a matter of cooperation. In *HUMAN 07*, 277-285.
- Hoc, J. M., & Carlier, X. (2002). Role of a common frame of reference in cognitive cooperation: sharing tasks between agents in air traffic control. *Cognition, Technology & Work*, 4(1), 37-47.
- Hollnagel, E. (1999). From function allocation to function congruence. In S. W. A. Dekker & E. Hollnagel (Eds.), *Coping with computers in the cockpit*, pp. 29- 53.
- Hollnagel, E. (2000). Principles for Modeling function allocation, *International Journal of Human-Computer Studies*, 52, 253-265.
- Kaber, D. B., & Endsley, M. R. (1997). Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Progress*, 16(3), 126-131.

- Kaber, D. B., Onal, E., & Endsley, M. R. (2000). Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. *Human Factors and Ergonomics in Manufacturing*, 10(4), 409-430.
- Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*, 5(2), 113-153.
- Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems*, 19(6), 91-95.
- Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (2014). Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing (IJIDeM)*
- Le Goff, K., Haggard P., Rey, A., & Berberian B. (submitted). Automation technology and system predictability in supervisory task: Mediated agency : submitted to *Plos ONE*
- Moore, J. W., Wegner, D. M., & Haggard, P. (2009). Modulating the sense of agency with external cues. *Consciousness and cognition*, 18(4), 1056-1064.
- Navarro, J., Mars, F., & Hoc, J. M. (2007). Lateral control assistance for car drivers: a comparison of motor priming and warning systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49(5), 950-960.
- Navarro, J., Mars, F., Forzy, J. F., El-Jaafari, M., & Hoc, J. M. (2010). Objective and subjective evaluation of motor priming and warning systems applied to lateral control assistance. *Accident Analysis & Prevention*, 42(3), 904-912.
- Norman, D. A. (1990). The 'problem' with automation: inappropriate feedback and interaction, not 'over-automation'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 327(1241), 585-593.
- Norman, D. A. (2010). *Living with complexity*. MIT Press.
- Pacherie, E. (2012). The phenomenology of joint action: Self-agency vs. joint-agency. In Seemann, A. (Ed.), *Joint attention: New developments* (pp 343-389). MIT Press.
- Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced 'complacency'. *The International Journal of Aviation Psychology*, 3(1), 1-23.
- Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 38(4), 665-679.
- 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.
- Plat, M., & Amalberti, R. (2000). Experimental crew training to deal with automation surprises. In Sarter, N., Amalberti, R. (Eds), *Cognitive engineering in the aviation domain* (pp 287-308). Hillsdale (NJ): Lawrence Erlbaum Associates.
- Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. In Salvendy, G. (Ed), *Handbook of human factors and ergonomics second edition* (pp 1926-1943). Wiley.

- Sato, A. (2009). Both motor prediction and conceptual congruency between preview and action-effect contribute to explicit judgment of agency. *Cognition*, 110(1), 74-83.
- Singh, I. L., Molloy, R., & Parasuraman, R. (1997). Automation-induced monitoring inefficiency: role of display location. *International Journal of Human-Computer Studies*, 46(1), 17-30.
- Sheridan, T. B., & Verplank, W. L. (1978). Human and computer control of undersea teleoperators. MIT Cambridge Man-Machine System Lab.
- Taleb, N. N. (2012). *Antifragile: things that gain from disorder*. Random House Incorporated.
- Verfaillie, K., & Daems, A. (2002). Representing and anticipating human actions in vision. *Visual Cognition*, 9(1-2), 217-232.
- Vesper, C., van der Wel, R. P., Knoblich, G., & Sebanz, N. (2011). Making oneself predictable: Reduced temporal variability facilitates joint action coordination. *Experimental Brain Research*, 211(3-4), 517-530.
- Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, 54(7), 480.
- Wenke, D., Fleming, S. M., & Haggard, P. (2010). Subliminal priming of actions influences sense of control over effects of action. *Cognition*, 115(1), 26-38.
- Wiener, E.L. (1989). Cockpit Automation. In Wiener, E.L., Nagel, D.C. (Eds.), *Human Factors in Aviation* (pp. 433– 459). San Diego, CA: Academic Press.
- Woods, D.D., Tinapple, D. (1999). *W3: Watching Human Factors Watch People at Work*. Presidential Address, 43rd Annual Meeting of the Human Factors and Ergonomics Society, Houston, September 28.
- Zimmermann, M., Bauer, S., Lutteken, N., Rothkirch, I. M., & Bengler, K. J. (2014, May). Acting together by mutual control: Evaluation of a multimodal interaction concept for cooperative driving. In *Collaboration Technologies and Systems (CTS), 2014 International Conference on* (pp. 227-235). IEEE.

Toward a Model for Effective Human-Automation Interaction: The Mediated Agency

Kevin Le Goff^{1,2}, Arnaud Rey², and Bruno Berberian^{1(✉)}

¹ Systems Control and Flight Dynamics Department, ONERA, Salon de Provence, France
Bruno.Berberian@onera.fr

² Laboratoire de Psychologie Cognitive, CNRS and Aix-Marseille Université, Marseille, France
legoff.kev@gmail.com

Abstract. In our increasingly technological world, automation largely improved some aspects of our life. Nonetheless, automation can also have negative consequences. Indeed, operators seem often helpless to takeover an automated system in case of failure. This “out-of-the-loop” problem occurs when operator is unable to understand the intentions and to predict the outcome of actions of the system, causing a decrease of control. The following article illustrates how the psychological approach of agency can help (1) to better understand this OOTL performance problem and (2) to propose design principles to improve human machine interaction in case of system automation.

1 Introduction

In our increasingly technological world, the influence of automation is perceived in each aspect of everyday life. At work or at home, human beings are accustomed to interact with sophisticated computer systems designed to assist them in their activities. Automation certainly makes some aspects of life easier: by allowing people with disabilities to be able to move and communicate; faster: with the generalization of computerized devices and the increase of productivity; and safer: the accident rate in aviation or high-risk industry has dropped down thanks to the implementation of automated systems [1]. This is far from over, thus, no one would be surprised in future years to see a car without driver let her/him cross the road.

Automation is often a suitable solution for functions that humans cannot achieve safely or reliably. Previous studies demonstrated that high level of automation reduced human operator workload and increase the level of productivity [2]. Nonetheless, the interposition of automated systems between human operators and processes transforms the nature of human work. As a matter of fact, the role of the human actors tends to evolve from direct control to supervision. This change is far from trivial and creates new burdens and complexities for the individuals and teams of practitioners responsible for operating, troubleshooting and managing high-consequence systems.

2 Automation and OOL Performance Problem

When new automation is introduced into a system, or when there is an increase in the autonomy of automated systems, developers often assume that adding “automation” is a simple substitution of a machine activity for human activity (substitution myth, see [3]). Empirical data on the relationship between people and technology suggest that this is not the case and that traditional automation has many negative outcomes and safety consequences associated with it stemming from the human out-of-the-loop (OOL) performance problem [4, 5].

The OOL performance problem has been attributed to a number of underlying factors, including human vigilance decrements [6, 7], complacency [8, 9] and loss of operator situation awareness (SA) [10, 11]. Cognitive engineering literature has discussed at length the origins of vigilance decrements (e.g., low signal rates, lack of operator sensitivity to signals), complacency (e.g., over trust in highly reliable computer control) and the decrease in SA (use of more passive rather than active processing and the differences in the type of feedback provided) in automated system supervision and has established associations between these human information processing shortcomings and performance problems.

As a major consequence, the OOL performance problem leaves operators of automated systems handicapped in their ability to take over manual operations in the case of automation failure. Particularly, the OOL performance problem causes a set of difficulties including a longer latency to determine what has failed, to decide if an intervention is necessary and to find the adequate course of action [6]. The three following incidents from aviation, nuclear plant and finance domains illustrate such difficulties.

- Situation 1: Aviation

The first example concerns Flight 447 from Air France. On May 31, 2009, the Airbus A330 took off from Rio de Janeiro bound to Paris. Four hours after the departure and due to weather conditions, ice crystals obstructed the Pitot probes. Hence, speed indications were incorrect and lead to a disconnection of the autopilot. Likely following this disconnection, the crew was unable to diagnose the situation and apply the appropriate procedure. Alternating appearances and disappearances of some indicators and alarms coupled with high stress probably prevented the crew to correctly evaluate the state of the system and act appropriately (for the official report, see [12]).

- Situation 2: Nuclear Power Plant

The second example concerns the incident of the nuclear plant of Three Miles Island (Pennsylvania, USA), in 1979. A valve used to regulate the water inlet in the nuclear core was stuck open, although a light on the control interface indicated that the valve position was closed. However, this light did not indicate real position of the valve but instead that the closure order was given. Because of ambiguous information provided by the control interface, the operators were unable to correctly diagnose the problem for several hours [13]. During this period, a sequence of different failures and inappropriate actions lead to a partial meltdown of the nuclear core. Hopefully, the releases of radiations were not important enough to cause health and environmental damages. A major nuclear disaster was avoided.

- Situation 3: Stock Market

In a completely different domain, we can mention one of the costliest computer bug. Knight Capital is a firm specialized in high frequency trading (automated technic used to buy and sell stocks in fractions of a second). On August 1, 2012, the firm tested a new version of its trading algorithm. However, due to a bug, the algorithm started pushing erratic trades. Because supervisors were not up to date of the system behavior, it took a long hour to understand that the problem came from the algorithm and costed to Knight Capital about 400 million dollars [14].

Although these previous cases are from different domains, they highlight that when the automatic equipment fails, supervisors seem dramatically helpless for diagnosing the situation and determining the appropriate solution because they are not aware of the system state prior to the failure. Numerous experimental results confirm such difficulties. For example, Endsley and Kiris [4] provided evidence that performance during failure mode following a fully automated period were significantly degraded, as compared to a failure mode following a fully manual control. Merat and Jamson [15] reported similar conclusions. In a driving simulation task, they demonstrated that drivers' responses to critical events were slower in the automatic driving condition than in the manual condition. Because automation is not powerful enough to handle all abnormalities, this difficulty in takeover is a central problem in automation design. Moreover, with the development of autonomous cars, which should come onto our roads in a few years, everyone (not only expert operators) could be concerned by such difficulties, and the issue becomes universal.

These difficulties in takeover situations have been identified for a long time [1, 16] and different solutions have been proposed by the human factors society. Some of them consist in training human operator to produce efficient behavior in case of system failure. However, recent dramatic events indicate that such training does not ensure efficient takeover for trained situations, whereas the apparition of unexpected failure are not considered by such approach. Other solutions propose to manipulate the level of system automation, sharing the authority between the automation and the human operator (for example MABA-MABA methods, adaptive function allocation). Such approach rests on the hypothesis that new technologies can be introduced as a simple substitution of machines for people - preserving the basic system while improving it on some output measures. Unfortunately, such assumption corresponds to a vague and bleak reflection of the real impact of automation: automation technology transforms human practice and forces people to adapt their skills and routines [17].

If these traditional approaches have the virtue to partially decrease the negative consequences of automation technology, clear solutions are still missing to overcome these takeover difficulties [18, 19]. We argue that the key for designers is to focus on how automation technology transforms the human operator activity and what are the mechanisms of control involved in a supervisory task. We assume that the response to these questions remains a crucial challenge for successful design of new automated systems. In this paper, we propose a theoretical framework to explain this transformation.

3 OOL Performance Problem and System Predictability

As assumed by Norman [1], the lack of system predictability is certainly a central point in the comprehension of the OOL phenomenon and the associated takeover difficulties. With the advent in technology, current man-made complex systems tend to develop cascades and runaway chains of automatic reactions that decrease, even eliminate predictability and cause outsized events [20]. This is what we will call *system opacity*: the difficulty for a human operator to see the arrow from system intention to actual state and to predict the sequence of events that will occur.

However, such opacity is far from being a fatality. As pointed by Norman [1], the problem with automation is more its inappropriate design and application than over automation per se. Particularly, the lack of continual feedback and interaction appears as the central problem. Throughout the last years, computational and experimental evidences have proved the central place of feedback regarding the mechanisms that govern the control of our actions [21–25]. When people perform actions, feedback is essential for the appropriate monitoring of those actions. However, adequate feedback to the human operator is most of the time absent in case of system automation. Without appropriate feedback about the state of the system, people may not know if the actions are being performed properly or if problems are occurring. As a result, when an automatic equipment fails, people are not able to detect symptoms of troubles early enough to overcome them.

Interestingly, system engineers are blind to the paradox that we have never had more data than we have now, yet have less predictability than ever. To overcome such opacity, they have to propose adequate feedback about the state of the system. However, how to design a predictable system remains a difficult problem. A possible approach is to focus on how humans understand and control their own actions. Indeed, we can assume that operators interpret the intentions and the outcomes' actions of a system with their own "cognitive toolkit". Thus, understanding how this "cognitive toolkit" works could be relevant to propose design principles for potentially controllable systems.

4 Science of Agency as a Relevant Framework

When we act, we usually have a clear feeling that we control our own action and can thus produce effects in the external environment. This feeling has been described as "the sense of agency" [26], and is recognized as an important part of normal and human consciousness. Most people can readily sort many events in the world into those they have authored and those they have not. This observation suggests that each person has a system for authorship processing [27], a set of mental processes that monitors indications of authorship to judge whether an event, action, or thought should be ascribed to self as a causal agent [28]. Laboratory studies have attempted to shed more light on this mechanism and empirical data in recent psychology [29, 30], psychopathology [31, 32] and neuroscience [33, 34] have been accumulated. Interestingly, a variety of sources of information (e.g., one's own thoughts, interoceptive sensations, external feedback, etc.) could be involved in the authorship processing. Several indicators have

been already proposed, including body and environment orientation cues [35], direct bodily feedback [36, 37], direct bodily feedforward [38, 39], visual and other indirect sensory feedback [40], social cues [41], agent goal information [42] and own behavior relevant thought [43–45]. Although, the mental processes contributing to the sense of agency are not fully understood at this time, the different approaches propose that we derive a sense of being the agent for our own actions by a cognitive mechanism that computes the discrepancies between the predicted consequences of our own actions' actual consequences of these actions, similarly to action control models [22, 24, 43]. Thus, predictability appears as a key notion regarding the mechanism of agency and researchers demonstrated that efferent signals, re-afferent signals, higher order knowledge and beliefs influence it.

Interestingly, Pacherie [46] argued that the different mechanisms underlying sense of agency for individual actions are similar to those underlying sense of agency one experiences when engaged in joint action. That is, the sense of agency in joint action is based on the same principle of congruence between predicted and actual outcomes. Sebanz, Bekkering and Knoblich [47] defined a joint action as “any form of social interaction whereby two or more individuals coordinate their actions in space and time to bring about change in the environment”. Moreover, one of the criteria needed for the accomplishment of a joint action is the ability to predict other's actions and their outcomes. Predicting other's actions required areas involved in the human self-action control system [48]. This capability of prediction is crucial to achieve an efficient coordination in a joint action [49–51]. Vesper et al. [51] demonstrated that participants involved in a joint task were more predictable by reducing the temporal variability in order to facilitate the coordination.

If we consider that the predictability in human-system interactions and in human-human interactions operates in the same way, we can assume that the system opacity could dramatically change our experience of agency. This hypothesis receives an echo from the claim of Baron when he said:

“Perhaps the major human factors concern of pilots in regard to introduction of automation is that, in some circumstances, operations with such aids may leave the critical question, who is in control now, the human or the machine?” [52].

Recent empirical data have confirmed such degradation of our experience of agency in presence of automation [53]. Particularly, by manipulating the level of automation in an aircraft supervision task, we have demonstrated a decrease in agency (for both implicit and explicit measures) concomitant to the increase in automation. Consequently, we assume that a way to design a more controllable interface is to consider supervision as a joint action between a human operator and an artificial co-agent following the same principles as a biological co-agent. This proposition echoes that of Norman [1] when he assumed that *what is needed is continual feedback about the state of the system, in a normal natural way, much in the manner that human participants in a joint problem-solving activity will discuss the issues among themselves*. The use of the theoretical background of agency will make it easier to achieve this objective. This is why we argue, in this paper, for a mediated agency: an approach to HMI interactions that takes into account how the information provided by an automated system influences how an operator feels in control.

5 Agency Offers Tools and Measures

As previously assumed, system opacity is certainly a major cause of OOL performance problems. To overcome such system opacity, interactions designers could use the tools and measures, provided by the framework of agency (and by extension, the one of joint agency). Examples of such tools can be derived from the theory of apparent mental causation of Daniel Wegner [43]. His theory provided clues to determine the nature, the form and the timing of an appropriate feedback. Particularly, Wegner proposed that when a thought occurs prior to an action it is consistent with the action and the action has no plausible alternative cause, then we experience the feeling of consciously willing the action. This is what he called, priority, consistency and exclusivity principles. System engineers could use these principles to shape adequate feedbacks in order to make the automation more predictable. We already know that this capability of prediction is crucial to achieve an efficient coordination in a joint action [49–51]. Recently Berberian and colleagues provided evidence that these principles could be used to design Human-machine interfaces capable of compensating the negative effects of latency on action control [54]. We assumed that automated systems following such principles would make the operators more “agent”, and then they would not be affected by the OOL performance problem. Thus, they would be faster and more reliable to take over an automated system in case of failure.

Another contribution from this framework to the human-machine domain is provided by the use of measures of agency. Although quantifying the degree of agency remains difficult, several measures have been proposed. Interestingly, we can distinguish two different kinds of measures (Explicit *vs.* Implicit) referring potentially to two separable agency processing systems [26, 55]. The explicit level refers to the “Judgment of agency”, that is, the capability to attribute agency to oneself or another. This judgment is influenced by beliefs and external cues [56]. One can use classic declarative methods, such as surveys and self-reports to evaluate this aspect of agency. The implicit level refers to the “Feeling of agency”, a low-level feeling of being an agent, mainly based on sensory-motor cues. Implicit markers of agency have been proposed by Haggard and colleagues, namely the intentional binding effect (IB). In a key study, they noticed that human intentional actions produce systematic changes in time perception. In particular, the interval between a voluntary action and an outcome is perceived as shorter than the interval between a physically similar involuntary movement and the same outcome event [57]. This phenomenon has been widely reported (for a review, see [58]). Although our understanding of the underlying mechanisms is not clear, IB may provide an implicit window into human agency. For the last decade, a lot of studies about agency were published, some used a method or another and some used both but still in simple and easy to control paradigms (for example, a visual signal appeared and the participant had to push a button). Regarding the applicability of these methods to design processes, future work will have to determine the efficiency of such methods in more complex situations encountered in the human-system interaction field. Berberian et al. [53] used both measures to evaluate the variation of agency depending on the level of automatism in an aircraft supervision task with different autopilot settings. They provided evidences that intentional binding measures are sensitive to graded variations in control associated

with automation and was related to explicit judgments. This study demonstrates that IB could be used in a richer and more complex paradigm. These measures should therefore be combined to develop a framework for evaluating the OOL performance problem. Hence, ergonomists should bear in mind the different elements of the human sense of agency, and the different ways of measuring it, in order to correctly evaluate if an interface is sufficiently acceptable and controllable.

6 Conclusion

We have seen that, despite all the benefits of automation, there are still issues to be corrected (loss of control, less efficient monitoring...). This clearly establishes that some problems in human-machine interaction stem from a decreased predictability, due to an increase of complexity. However, designing relevant feedback and system-operator communication is clearly the key to avoiding these problems, but this remains a challenge. We proposed that using principles, tools and measures from the science of agency should lead to the introduction of a new methodology. "Being an agent" is a notion largely studied in neurosciences, psychology and philosophy. It would be also relevant, for the human-machine domain, to use a framework taking into account the difference between self-generated actions and those generated by other sources. This is why we argue for a mediated agency framework: an approach to HMI interactions that takes into account how the information provided by an automated system influence how an operator feels in control. We suggest that the science of agency in the field of human-machine interaction may be fruitful to elaborate concrete recommendations to design automatics system supervised by operators "in the loop" abating the negative consequence of OOL problem while maintaining the performance in the normal range. Measuring (explicitly or implicitly) the feeling of control may be important in evaluating different automated devices, and may also be relevant for evaluating operator's performances in supervisory tasks. In the end, a better understanding of how the sense of agency evolved in the case of interactions between humans and automated system would certainly refine the different models of agency and, more generally, models of control.

References

1. Norman, D.A.: The 'Problem' with automation: inappropriate feedback and interaction, not 'Over-Automation'. *Philos. Trans. R. Soc. Lond.* **327**, 585–593 (1990)
2. Kaber, D., Onal, E., Endsley, M.: Design of automation for telerobots and the effect on performance, operator situation awareness and subjective workload. *Hum. Factors Ergon. Manuf.* **10**, 409–430 (2000)
3. Woods, D.D., Tinapple, D.: W3: watching human factors watch people at work. Presidential Address, 43rd Annual Meeting of the Human Factors and Ergonomics Society (1999)
4. Endsley, M.R., Kiris, E.O.: The out-of-the-loop performance problem and level of control in automation. *Hum. Factors J. Hum. Factors Ergon. Soc.* **37**(2), 381–394 (1995)
5. Kaber, D.B., Endsley, M.R.: Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Saf. Prog.* **16**(3), 126–131 (1997)

6. Billings, C.E.: *Human-Centered Aircraft Automation: A Concept and Guidelines* (NASA Tech. Memo. 103885). Moffet Field, CA: NASA-Ames Research Center (1991)
7. Wiener, E.L.: Cockpit automation. In: Wiener, E.L., Nagel, D.C. (eds.) *Human Factors in Aviation*, pp. 433–459. Academic Press, San Diego (1988)
8. Parasuraman, R., Molloy, R., Singh, I.L.: Performance consequences of automation induced complacency. *Int. J. Aviat. Psychol.* **3**, 1–23 (1993)
9. Singh, I.L., Molloy, R., Parasuraman, R.: Automation-induced monitoring inefficiency: role of display location. *Int. J. Hum. Comput. Stud.* **46**(1), 17–30 (1997)
10. Carmody, M.A., Gluckman, J.P.: Task specific effects of automation and automation failure on performance, workload and situational awareness. In: Jensen R.S., Neumeister, D. (eds.) *Proceedings of the 7th International Symposium on Aviation Psychology*, pp. 167–171. Department of Aviation, Ohio State University, Columbus (1993)
11. Endsley, M.R.: Automation and situation awareness. In: Parasuraman, R., Mouloua, M. (eds.) *Automation and Human Performance: Theory and Applications*, pp. 163–181. Lawrence Erlbaum, Mahwah (1996)
12. B.E.A.: Final Report on the accident of the flight AF 447 Rio de Janeiro-Paris (2012). <http://www.bea.aero/fr/enquetes/vol.af.447/rapport.final.fr.php>
13. Norman, D.A.: *The Psychology of Everyday Things*. Basic Books, New York (1988)
14. Jones, C.M.: What do we know about high-frequency trading? Columbia Business School Research Paper No. 13–11 (2013)
15. Merat, N., Jamson, A.H.: How do drivers behave in a highly automated car. In: *Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, pp. 514–521 (2009)
16. Bainbridge, L.: Ironies of automation. *Automatica* **19**(6), 775–779 (1983)
17. Dekker, S.W., Woods, D.D.: MABA-MABA or abracadabra? progress on human–automation co-ordination. *Cogn. Technol. Work* **4**(4), 240–244 (2002)
18. Baxter, G., Rooksby, J., Wang, Y., Khajeh-Hosseini, A.: The ironies of automation: still going strong at 30? In: *Proceedings of the 30th European Conference on Cognitive Ergonomics*, pp. 65–71. ACM, New York (2012)
19. Norman, D.A.: *Living with Complexity*. MIT Press, Cambridge (2010)
20. Taleb, N.N.: *Antifragile: Things that Gain From Disorder*. Random House Incorporated, New York (2012)
21. Kawato, M.: Internal models for motor control and trajectory planning. *Curr. Opin. Neurobiol.* **9**(6), 718–727 (1999)
22. Wolpert, D.M.: Computational approaches to motor control. *Trends Cogn. Sci.* **1**(6), 209–216 (1997)
23. Wolpert, D.M., Ghahramani, Z., Jordan, M.I.: An internal model for sensorimotor integration. *Science* **269**(5232), 1880–1882 (1995)
24. Blakemore, S.J., Wolpert, D.M., Frith, C.D.: Abnormalities in the awareness of action. *Trends Cogn. Sci.* **6**(6), 237–242 (2002)
25. Frith, C.D., Blakemore, S.J., Wolpert, D.M.: Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. *Brain Res. Rev.* **31**(2), 357–363 (2000)
26. Gallagher, S.: Philosophical concepts of the self: implications for cognitive sciences. *Trends Cogn. Sci.* **4**, 14–21 (2000)
27. Wegner, D.M., Sparrow, B.: Authorship processing. In: Gazzaniga, M.S. (ed.) *The New Cognitive Neurosciences*, 3rd edn. MIT Press, Cambridge (2004)
28. Wegner, D.M., Sparrow, B., Winerman, L.: Vicarious agency: experiencing control over the movements of others. *J. Pers. Soc. Psychol.* **86**(6), 838 (2004)

29. Aarts, H., Custers, R., Wegner, D.M.: On the inference of personal authorship: enhancing experienced agency by priming effect information. *Conscious. Cogn.* **14**(3), 439–458 (2005)
30. Moore, J.W., Wegner, D.M., Haggard, P.: Modulating the sense of agency with external cues. *Conscious. Cogn.* **18**(4), 1056–1064 (2009)
31. Franck, N., Farrer, C., Georgieff, N., Marie-Cardine, M., Daléry, J., d’Amato, T., Jeannerod, M.: Defective recognition of one’s own actions in patients with schizophrenia. *Am. J. Psychiatry* **158**(3), 454–459 (2001)
32. Farrer, C., Franck, N., Georgieff, N., Frith, C.D., Decety, J., Jeannerod, M.: Modulating the experience of agency: a positron emission tomography study. *Neuroimage* **18**(2), 324–333 (2003)
33. Tsakiris, M., Haggard, P.: The rubber hand illusion revisited: visuotactile integration and self-attribution. *J. Exp. Psychol. Hum. Percept. Perform.* **31**(1), 80 (2005)
34. Vallacher, R.R., Wegner, D.M.: *A Theory of Action Identification*. Psychology Press, New York (2014)
35. Gandevia, S.C., Burke, D.: Does the nervous system depend on kinesthetic information to control natural limb movements? *Behav. Brain Sci.* **15**(04), 614–632 (1992)
36. Georgieff, N., Jeannerod, M.: Beyond consciousness of external reality: a “who” system for consciousness of action and self-consciousness. *Conscious. Cogn.* **7**(3), 465–477 (1998)
37. Blakemore, S.J., Frith, C.D.: Self-awareness and action. *Curr. Opin. Neurobiol.* **13**, 219–224 (2003)
38. Blakemore, S.J., Frith, C.D., Wolpert, D.M.: Spatio-temporal prediction modulates the perception of self-produced stimuli. *J. Cogn. Neurosci.* **11**(5), 551–559 (1999)
39. Daprati, E., Franck, N., Georgieff, N., Proust, J., Pacherie, E., Daléry, J., Jeannerod, M.: Looking for the agent: an investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition* **65**(1), 71–86 (1997)
40. Milgram, S.: *Obedience to Authority*. Harper & Row, New York (1974)
41. Langer, E., Roth, J.: Heads I win, tails it’s chance. *J. Pers. Soc. Psychol.* **32**(6), 951–955 (1975)
42. Wegner, D.M.: *The Illusion of Conscious Will*. MIT Press, Cambridge (2002)
43. Wegner, D.M.: The mind’s best trick: how we experience conscious will. *Trends Cogn. Sci.* **7**, 65–69 (2003)
44. Wegner, D.M., Wheatley, T.: Apparent mental causation: sources of the experience of will. *Am. Psychol.* **54**(7), 480–492 (1999)
45. Pacherie, E.: The phenomenology of joint action: self-agency vs. joint-agency. In: Seemann, A. (ed.) *Joint Attention: New Developments*, pp. 343–389. MIT Press, Cambridge (2012)
46. Sebanz, N., Bekkering, H., Knoblich, G.: Joint action: bodies and minds moving together. *Trends Cogn. Sci.* **10**(2), 70–76 (2006)
47. Sebanz, N., Frith, C.: Beyond simulation? Neural mechanisms for predicting the actions of others. *Nat. Neurosci.* **7**(1), 5–6 (2004)
48. Verfaillie, K., Daems, A.: Representing and anticipating human actions in vision. *Vis. Cogn.* **9**(1–2), 217–232 (2002)
49. Kilner, J.M., Vargas, C., Duval, S., Blakemore, S.-J., Sirigu, A.: Motor activation prior to observation of a predicted movement. *Nat. Neurosci.* **7**(12), 1299–1301 (2004)
50. Vesper, C., van der Wel, R.P., Knoblich, G., Sebanz, N.: Making oneself predictable: reduced temporal variability facilitates joint action coordination. *Exp. Brain Res.* **211**(3–4), 517–530 (2011)
51. Baron, S.: Pilot control. In: Wiener, E.L., Nagel, D.C. (eds.) *Human Factors in Aviation*, pp. 347–386. Academic Press, San Diego (1988)

52. Berberian, B., Sarrazin, J.C., Le Blaye, P., Haggard, P.: Automation technology and sense of control: a window on human agency. *PLoS ONE* **7**(3), 34075 (2012)
53. Berberian, B., Le Blaye, P., Schulte, C., Kinani, N., Sim, P.R.: Data transmission latency and sense of control. In: Harris, D. (ed.) *EPCE 2013, Part I. LNCS*, vol. 8019, pp. 3–12. Springer, Heidelberg (2013)
54. Synofzik, M., Vosgerau, G., Newen, A.: Beyond the comparator model: a multifactorial two-step account of agency. *Conscious. Cogn.* **17**(1), 219–239 (2008)
55. Moore, J.W., Middleton, D., Haggard, P., Fletcher, P.C.: Exploring implicit and explicit aspects of sense of agency. *Conscious. Cogn.* **21**(4), 1748–1753 (2012)
56. Haggard, P., Aschersleben, G., Gehrke, J., Prinz, W.: Action, binding, and awareness. In: Prinz, W., Hommel, B. (eds.) *Common Mechanisms in Perception and Action*, vol. 19, pp. 266–285. Oxford University Press, Oxford (2002)
57. Haggard, P., Clark, S., Kalogeras, J.: Voluntary action and conscious awareness. *Nat. Neurosci.* **5**(4), 382–385 (2002)
58. Moore, J.W., Obhi, S.S.: Intentional binding and the sense of agency: a review. *Conscious. Cogn.* **21**(1), 546–561 (2012)

CV et Production scientifique

LE GOFF KEVIN

11 Rue de la Chaîne
31000 Toulouse
legoff.kev@gmail.com
Tel : 06 43 64 41 80

LANGUES

- Français
- Anglais scientifique et technique (écrit et oral)

PROGRAMMATION

- Python
- Matlab
- R
- LaTeX

OUTILS

- Statistiques
- Eye-tracker
- EEG
- IRM

Expériences professionnelles

En cours : Thèse DGA en Neurosciences et Facteurs humains

ONERA & Laboratoire Psychologie Cognitive (13)

" Agentivité dans les systèmes fortement automatisés "

Le but de cette thèse est d'utiliser le cadre théorique du sentiment d'agentivité et des outils tels que l'eye-tracker pour proposer des interfaces améliorant le sentiment de contrôle, l'acceptabilité et la reprise en main de systèmes fortement automatisés.

2013 - En cours : Chargé d'enseignement

Kedge Business School (13) :

- U.E. Prise de Décision et Management (2ème Année)

Aix-Marseille Université (13) :

- U.E. Statistiques en Biologie (L2 Biologie)
- U.E. Introduction à la Programmation (L2 Neurosciences)
- U.E. Modélisation et Simulation (L3 Neurosciences)

Été 2012 : Consultant scientifique

Scalr (San Francisco, USA)

Start-up dans le Cloud computing, conseil en psychologie, prise de décision, ergonomie et neurosciences.

2011 - 2013 : Assistant de recherche

Institut for Mind and Biology (Chicago, USA, stage de M1, 2012)

Laboratoire de Psychologie Cognitive (Marseille, 13)

Etudes en comportement, EEG et IRM : protocoles, veille scientifique, analyses de données, écriture d'articles scientifiques.

Diplômes et Formation

En cours : Doctorat de Neurosciences et Facteurs humains

ONERA - Aix-Marseille Université (13)

2013 : Master de Neurosciences Intégratives et Cognitives

Aix-Marseille Université (13)

Obtention d'une bourse de thèse D.G.A.

2011 : Licence de Neurosciences Intégratives et Cognitives

Université de Provence (13)

2007 : Baccalauréat Scientifique, spécialité Mathématiques

Lycée Saint Louis Lorient (56)

Informations supplémentaires

Depuis 2014 :

- Co-fondateur et trésorier de l'Association des doctorants de mon laboratoire
- Représentant des doctorants au conseil de laboratoire

Production scientifique

Articles publiés

Basso, F., Bouillé, J., **Le Goff, K.**, Robert-Demontrond, P., & Oullier, O. (2016). Assessing the role of shape and label in the misleading packaging of food imitating products : from empirical evidence to policy recommendation. *Frontiers in psychology*, 7.

Madec, S., **Le Goff, K.**, Anton, J. L., Longcamp, M., Velay, J. L., Nazarian, B., Roth, M., Courrieu, P., Grainger, J., & Rey, A. (2016). Brain correlates of phonological recoding of visual symbols. *NeuroImage*, 132, 359-372.

Madec, S., **Le Goff, K.**, Riès, S. K., Legou, T., Rousselet, G., Courrieu, P., Alario, F.X., Grainger, J., & Rey, A. (2016). The time course of visual influences in letter recognition. *Cognitive, Affective, & Behavioral Neuroscience*, 1-9.

Articles soumis

Basso, F., Voyer, B., Petit, O., **Le Goff, K.**, & Oullier, O. (2016). In the ?I? of the beholder : Improving health promotion by using the actor ?s perspective. *Journal of Consumer Psychology*

Basso, F., Petit, O., Le Bellu, S., Lahlou, S., **Le Goff, K.**, Anton, J. L., Nazarian, B., Cancel, A., & Oullier, O. (2016). Taste at first (person) sight : Visual perspective modulates body mass index-dependent activations in brain regions contributing to taste and reward representations. *Cerebral Cortex*

Le Goff, K., Rey, A., Haggard, P., Oullier, O., & Berberian, B. (2016). Agency modulates interactions with automation technologies. *Ergonomics*

Articles en préparation

Le Goff, K., Rey, A., Haggard, P., & Berberian, B. (2016). Automation technology and sense of agency

Le Goff, K., Rey, A., Haggard, P., & Berberian, B. (2016). Vicarious agency in human-machine interaction : a matter of time

Proceedings

Basso, F., Bouillé, J., **Le Goff, K.**, Robert-Demontrond, P. & Oullier, O. (2013), A methodological approach of food imitating products using implicit association test, 29th International Conference of the French Marketing Association, La Rochelle, France, May 16-17.

Le Goff, K., Rey, A., & Berberian, B. (2015, August). Toward a Model for Effective Human-Automation Interaction : The Mediated Agency. In *International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management* (pp. 274-283). Springer International Publishing.

Le Goff, K., Rey, A., & Berberian, B. (2015, July). Prédicibilité des systèmes et sentiment de contrôle : vers des interfaces plus agentives EPIQUE 2015 Aix-en-Provence.

Communication orale

Basso, F., Bouillé, J., **Le Goff, K.**, Robert-Demontrond, P. & Oullier, O. (2013), A methodological approach of food imitating products using implicit association test, 29th International Conference of the French Marketing Association, La Rochelle, France, May 16-17.

Le Goff, K., Rey, A., & Berberian, B. (2015, August). Toward a Model for Effective Human-Automation Interaction : The Mediated Agency. In *International Conference*

on *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management* (pp. 274-283). Springer International Publishing.

Le Goff, K., Rey, A., & Berberian, B. (2015, July). Prédicibilité des systèmes et sentiment de contrôle : vers des interfaces plus agentives EPIQUE 2015 Aix-en-Provence.

Poster

Basso, F., Bouillé, J., **Le Goff, K.**, Robert-Demontrond, P. & Oullier, O. (2013). A methodological approach of *Food Imitating Products* using Implicit Association Test. 25th Association for Psychological Science Annual Convention, Washington, D.C., USA, May 23-26.

Le Goff, K., Rey, A., & Berberian, B. (2015, July). How can a science of agency inform automation design. 19th Annual Meeting of the Association for the Scientific Study of Consciousness, Paris, France, July 7-10.

Le Goff, K., & Rey, A. (2012). The unconscious order effect in complex decision making. 24th Association for Psychological Science Annual Convention, Chicago, USA, May 24-27.

Le Goff, K., Pacton, S., Manguin, J-L., & Rey, A. (2013). The interference of misspellings on spelling performance. 25th Association for Psychological Science Annual Convention, Washington, D.C., USA, May 23-26.

Abstract

In our increasingly technological world, automation has improved many aspects of our lives. But automation can also have negative sides. Indeed, human operators seem often helpless to takeover an automated system in case of failure. This "out-of-the-loop" problem occurs when an operator is unable to understand the intentions and to predict the outcome of actions of the system, causing a decrease of control. The goal of this thesis was to study how the theoretical framework of agency can help identifying and evaluating the information required to make supervision of fully automated systems more efficient, to restore an appropriate sense of control and to increase the user acceptance of what the system is doing. To achieve this goal, we tested participants in four aircraft supervision tasks. We first showed that providing greater information about the system's intentions increased the participants' sense of control. However, this condition did not produce any change in a frequently-used implicit marker of the sense of agency, the so-called "intentional binding effect" (Experiment 1). We also found that this information led to better performances in detecting when the system's decisions were non-optimal (Experiment 2, 3, and 4), to a change in the way participants allocated their attentional resources (Experiment 2 and 3) and also to increase the level of user acceptance (Experiment 3 and 4). Interestingly, we also showed in the last experiment that the delay between prime messages providing information about the system's intentions and the system's actions impacted differently the performance and the operator's levels of control and acceptance. Finally, in the last section, we discuss the implications of our results for the field of human-machine interaction. We claim that the science of agency may be useful to elaborate concrete recommendations for designing automatic systems in which operators remain "in the loop" of control.

Keywords : sense of agency ; human-machine interactions ; user acceptance ; supervision

Résumé

Le développement des systèmes automatisés a permis d'améliorer de nombreux aspects de notre vie quotidienne. Toutefois, les bénéfices engendrés par l'utilisation croissante de l'automatisation masquent un problème fondamental. En effet, en cas de pannes ou de situations imprévues, les opérateurs présentent souvent des difficultés dans la reprise en main de tels systèmes. Ce phénomène de « sortie de boucle de contrôle » se produit lorsque l'opérateur est incapable de comprendre les intentions du système et de prédire ses actions futures, entraînant ainsi une perte de contrôle. Le but de cette thèse était d'étudier comment le cadre théorique de l'agentivité pouvait nous aider à identifier et évaluer les informations requises pour rendre la supervision de systèmes fortement automatisés plus efficace, pour restaurer un sentiment de contrôle approprié et pour augmenter l'acceptabilité du système par les opérateurs. Pour y parvenir, nous avons réalisé quatre expériences lors desquelles les participants devaient superviser le vol d'un avion sous pilote automatique. Dans un premier temps, nous avons montré que fournir des informations sur les intentions du système augmentait le sentiment de contrôle des participants. Cependant, ce gain d'information ne nous a pas permis d'observer le phénomène de liage intentionnel (*Intentional Binding*), fréquemment utilisé comme marqueur implicite de l'agentivité (Expérience 1). Nous avons également trouvé que cette information entraînait de meilleures performances pour détecter des décisions non-optimales du système (Expériences 2, 3 et 4), conduisait à un changement dans la façon dont les participants allouaient leurs ressources attentionnelles (Expériences 2 et 3) et aussi augmentait le niveau d'acceptabilité du système (Expériences 3 et 4). De façon intéressante, nous avons également montré dans la dernière expérience que le délai entre l'apparition de l'information sur les intentions du système et l'implémentation de son action impactait différemment la performance et les sentiments de contrôle et d'acceptabilité des opérateurs. Pour finir, dans la dernière partie, nous avons discuté des implications de nos résultats dans le champ des interactions homme-machine. Nous avançons l'idée que le cadre théorique de l'agentivité peut être utile pour élaborer des recommandations concrètes pour la conception de systèmes fortement automatisés permettant à l'opérateur de rester dans la « boucle de contrôle ».

Mots-clés : sentiment d'agentivité ; interactions homme-machine ; acceptabilité ; supervision

Agentivité dans les systèmes fortement automatisés

Le développement des systèmes automatisés a permis d'améliorer de nombreux aspects de notre vie quotidienne. Toutefois, les bénéfices engendrés par l'utilisation croissante de l'automatisation masquent un problème fondamental. En effet, en cas de pannes ou de situations imprévues, les opérateurs présentent souvent des difficultés dans la reprise en main de tels systèmes. Ce phénomène de « sortie de boucle de contrôle » se produit lorsque l'opérateur est incapable de comprendre les intentions du système et de prédire ses actions futures, entraînant ainsi une perte de contrôle.

Le but de cette thèse était d'étudier comment le cadre théorique de l'agentivité pouvait nous aider à identifier et évaluer les informations requises pour rendre la supervision de systèmes fortement automatisés plus efficace, pour restaurer un sentiment de contrôle approprié et pour augmenter l'acceptabilité du système par les opérateurs. Pour y parvenir, nous avons réalisé quatre expériences lors desquelles les participants devaient superviser le vol d'un avion sous pilote automatique. Dans un premier temps, nous avons montré que fournir des informations sur les intentions du système augmentait le sentiment de contrôle des participants. Cependant, ce gain d'information ne nous a pas permis d'observer le phénomène de liage intentionnel (Intentional Binding), fréquemment utilisé comme marqueur implicite de l'agentivité (Expérience 1). Nous avons également trouvé que cette information entraînait de meilleures performances pour détecter des décisions non-optimales du système (Expériences 2, 3 et 4), conduisait à un changement dans la façon dont les participants allouaient leurs ressources attentionnelles (Expériences 2 et 3) et aussi augmentait le niveau d'acceptabilité du système (Expériences 3 et 4). De façon intéressante, nous avons également montré dans la dernière expérience que le délai entre l'apparition de l'information sur les intentions du système et l'implémentation de son action impactait différemment la performance et les sentiments de contrôle et d'acceptabilité des opérateurs.

Pour finir, dans la dernière partie, nous avons discuté des implications de nos résultats dans le champ des interactions homme-machine. Nous avançons l'idée que le cadre théorique de l'agentivité peut être utile pour élaborer des recommandations concrètes pour la conception de systèmes fortement automatisés permettant à l'opérateur de rester dans la « boucle de contrôle ».

Mots-clés : AUTOMATISATION ; SORTIE DE BOUCLE;AGENTIVITE

Agentivity in highly automated systems

In our increasingly technological world, automation has improved many aspects of our lives. But automation can also have negative sides. Indeed, human operators seem often helpless to takeover an automated system in case of failure. This "out-of-the-loop" problem occurs when an operator is unable to understand the intentions and to predict the outcome of actions of the system, causing a decrease of control.

The goal of this thesis was to study how the theoretical framework of agency can help identifying and evaluating the information required to make supervision of fully automated systems more efficient, to restore an appropriate sense of control and to increase the user acceptance of what the system is doing. To achieve this goal, we tested participants in four aircraft supervision tasks. We first showed that providing greater information about the system's intentions increased the participants' sense of control. However, this condition did not produce any change in a frequently-used implicit marker of the sense of agency, the so-called "intentional binding effect" (Experiment 1). We also found that this information led to better performances in detecting when the system's decisions were non-optimal (Experiment 2, 3, and 4), to a change in the way participants allocated their attentional resources (Experiment 2 and 3) and also to increase the level of user acceptance (Experiment 3 and 4). Interestingly, we also showed in the last experiment that the delay between prime messages providing information about the system's intentions and the system's actions impacted differently the performance and the operator's levels of control and acceptance.

Finally, in the last section, we discuss the implications of our results for the field of human-machine interaction. We claim that the science of agency may be useful to elaborate concrete recommendations for designing automatic systems in which operators remain "in the loop" of control.

Keywords : SENSE OF AGENCY ; HUMAN-MACHINE INTERACTIONS ; USER ACCEPTANCE ; SUPERVISION