



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse II – Jean Jaurès

Présentée et soutenue par

Eugénie AVRIL

Le 16 Décembre 2020

Automatisation de l'information et fiabilité : effets sur le comportement humain

École doctorale et discipline ou spécialité

ED 326 CLESCO : Ergonomie

Unité de recherche

SCoTE EA 7420

Directrice(s) ou Directeur(s) de Thèse

Julien Cegarra, Professeur des universités, INU Champollion

Jordan Navarro, Maître de conférences (HDR), Université Lumière Lyon 2

Jury

Sylvia Pelayo, Maître de conférences (HDR), Université de Lille, Rapporteur

Thierry Morineau, Professeur des universités, Université de Bretagne,
Rapporteur

Didier Gourc, Professeur, IMT Mines Albi, Examineur

Liên Wioland, Responsables d'études, INRS

Résumé

Face à la présence de plus en plus prépondérante des systèmes automatisés dans notre vie quotidienne, étudier les conséquences de problèmes de fiabilité de ces derniers sur le comportement humain est devenu une nécessité pour le domaine des facteurs humains.

Cette thèse a pour objectif d'étudier les effets de systèmes automatisés imparfaits, dont la fiabilité n'est pas irréprochable, sur le comportement humain. Elle se focalise plus précisément sur les systèmes automatisés prenant en charge la fonction information comme présentée dans le modèle de Parasuraman, Sheridan et Wickens (2000).

Ces effets sont étudiés à travers deux secteurs : le secteur aéronautique et le secteur du transport routier de marchandises. Croiser ces deux secteurs permet de mettre en place des méthodes complémentaires. Plus précisément, à travers une tâche de simulation de pilotage d'avion et une tâche d'évaluation de scénarios de planification, deux types de prise en charge de la fonction information présentant des problèmes de fiabilité ont été investigués. Ces études ont permis d'une part (1) d'approfondir les connaissances sur la fiabilité des systèmes automatisés selon les fonctions prises en charge et (2) d'apporter de nouvelles connaissances sur l'implémentation de nouveaux systèmes automatisés soutenant l'activité de planification où les études sur la fiabilité d'un système prenant en charge l'information sont inexistantes.

Mots-clés : Fiabilité, fonction information, systèmes automatisés, interaction homme-machine, oculométrie

Abstract

In our daily life it is easy to notice the overwhelming increase of automated systems. Studying the consequences of reliability problems of these systems on human behavior has become a necessity for different fields and especially for the human factors.

The purpose of this thesis is to study the effects of imperfect automated systems on human behavior, knowing that their reliability is not perfect.

More specifically, it focuses on automated systems that support the information function as presented in the model of Parasuraman, Sheridan and Wickens (2000).

These effects are studied across two main sectors: the aeronautical sector and the road freight transport sector. Using these two sectors allow to implement complementary methods. More precisely two types of support of the information function with reliability problems have been investigated. One is through an aircraft piloting simulation task and the second on a planning scenario evaluation task, These studies allowed on one hand (1) to deepen the knowledge on the reliability of automated systems according to the functions supported and (2) on another hand to bring new knowledge on the implementation of new automated systems supporting the planning activity where studies on the reliability of a system supporting information are non-existent.

Keywords: Reliability, information function, automated systems, human-machine interaction, eye tracking

Remerciements

A mes directeurs...

Julien Cegarra, merci d'avoir dirigé cette thèse. D'ailleurs, merci pour tout. Sans vous, il n'y aurait pas eu l'envie d'aller voir ce que c'est la recherche à travers une première expérience de mémoire, pas de souhait de poursuivre en thèse et donc pas trois belles années à travailler sur un sujet passionnant... Je remercie le Professeur qui m'a conseillé et qui m'a permis d'avancer dans ce travail de thèse mais aussi l'individu qui à travers sa bienveillance, sa gentillesse et son aide m'a permis d'évoluer à mon rythme et m'a donné l'envie de me surpasser.

Jordan Navarro, merci d'avoir accepté de co-diriger cette thèse. Merci pour nos échanges et pour ta présence constante. Malgré la distance, tu as toujours répondu présent dès que j'en ai eu besoin. Merci pour tes précieux et nombreux conseils, tes commentaires et le soutien dont tu as fait preuve, notamment sur l'écriture des articles. Je te suis également reconnaissante pour toute la bienveillance que tu as apporté en parallèle de ce travail de thèse.

Liên Wioland, merci d'avoir accepté d'encadrer cette thèse. Tes connaissances, tes conseils et tes retours m'ont permis d'y voir plus clair lorsque le flou gagnait du terrain. Autant professionnellement qu'humainement, tout en complémentarité, parfois même sans t'en rendre compte : tu as apporté à cet encadrement le cadre dont j'avais besoin pour mener à bien cette thèse. Merci beaucoup.

Aux membres du jury...

Merci à Sylvia Pelayo, Thierry Morineau et Didier Gourc d'avoir accepté d'évaluer ce travail de thèse, en cette période si compliquée et chargée de pandémie mondiale.

A l'équipe Smart-Planning...

Virginie, Didier, Antoine... Merci d'avoir rendu cette thèse possible. Merci de votre bonne humeur, bienveillance, ainsi que de votre aide constante dans la réflexion (et tellement plus encore !) des travaux de cette thèse.

Aux membres du laboratoire SCoTE...

Merci pour ces trois belles années. Merci de m'avoir montré ce que c'est de pouvoir faire sa thèse dans un environnement aussi favorable... Je suis sûre que c'est en partie parce que j'ai évolué près

de vous que l'envie de devenir un jour moi aussi enseignant-chercheur me semble si belle et si évidente.

A mes amis...

Charlotte, tu es sans aucun doute, *the* belle rencontre de cette aventure. J'ai rencontré une collègue et j'ai gagné une précieuse amie.

Marion, Kevin, Jonathan, Clara, Val... Merci d'avoir été et d'être encore et toujours là...

Aux doctorants et anciens doctorants... Merci pour ces soirées et pour les mots qui font sourire, toujours au bon moment...

A ma famille...

Il y a tant à dire... tellement d'émotions à retranscrire et si peu de mots qui existent pour vous décrire ce que je ressens. Alors je n'aurai qu'une phrase à vous citer et vous la connaissez... Votre présence m'a permis d'aller au bout de mes rêves... *Obobob tout au bout de mes rêves... Où la raison s'achève...*

A mon compagnon...

Manu... merci pour tout... ton aide, ton soutien et d'avoir toujours amené de la lumière lorsque parfois les jours étaient plus sombres...

Table des matières

PARTIE THEORIQUE	11
CHAPITRE 1 – LES SYSTEMES AUTOMATISES.....	12
1.1 L’EMERGENCE DES SYSTEMES AUTOMATISES : DEFINITIONS.....	13
1.2. ATTRIBUTION DES FONCTIONS SELON LES CAPACITES DU SYSTEME AUTOMATISE	14
1.2.1. Fitts (1951)	15
1.2.2. Sheridan et Verplank (1978)	16
1.3. ALLOCATION DES FONCTIONS ENTRE L’OPERATEUR HUMAIN ET LE SYSTEME AUTOMATISE.....	17
1.3.1. Endsley et Kaber (1999).....	18
1.3.2. Parasuraman, Sheridan et Wickens (2000).....	20
1.3.3. Hoc, Young & Blosseville (2009)	24
1.4. CONCLUSION INTERMEDIAIRE.....	26
CHAPITRE 2 : SYSTEME AUTOMATISE IMPARFAIT	27
2.1. LA FIABILITE D’UN SYSTEME AUTOMATISE:.....	28
2.1.1. Définitions	28
2.1.2. Exemple d’application : le secteur aéronautique.....	31
2.1.3. Exemple d’application : le secteur du transport routier de marchandises	34
2.2. CONCLUSION INTERMEDIAIRE.....	39
CHAPITRE 3 : INTERACTION HOMME-MACHINE.....	40
3.1. LES DIFFICULTES ENGENDREES PAR L’INTERACTION HOMME-MACHINE	41
3.1.1. Les ironies d’automatisation	41
3.1.2. Les conséquences d’un système automatisé imparfait.....	45
3.2. REPENSER LA CONCEPTION DES SYSTEMES AUTOMATISES POUR REDUIRE LES DIFFICULTES.....	50
3.3. CONCLUSION INTERMEDIAIRE.....	52
PARTIE EXPERIMENTALE	55
ARTICLE 1. EFFECT OF IMPERFECT INFORMATION AND ACTION AUTOMATION ON ATTENTIONAL ALLOCATION.....	56
ARTICLE 2. AUTOMATION TYPE AND RELIABILITY IMPACT ON VISUAL AUTOMATION MONITORING AND HUMAN PERFORMANCES	81
ARTICLE 3. DECISION SUPPORT SYSTEM RELIABILITY AND TRANSPARENCY: A QUALITATIVE ANALYSIS OF A PLANNING TASK IN ROAD FREIGHT TRANSPORT SECTOR	113
DISCUSSION GENERALE	140

Introduction générale

Suivre les propositions de navigation de son GPS, se fier aux informations données par les sondes dans son cockpit, utiliser la boîte automatique de son véhicule, travailler sur une chaîne de production industrielle automatisée : toutes ces situations ont en commun l'utilisation d'un système automatisé. De nos jours, ces systèmes ont pris une place centrale dans notre quotidien. Nous utilisons de plus en plus ces derniers comme des supports pour nous soutenir dans notre prise de décision et pour nous aider dans nos tâches récurrentes. Ils sont à la fois présents dans notre vie personnelle mais également dans notre vie professionnelle. Ainsi, avec l'émergence des nouvelles technologies, l'activité humaine a évolué et les systèmes automatisés sont apparus dans un grand nombre de métiers. Bien qu'il soit difficile de donner une date précise sur l'apparition de ces derniers, des exemples concrets tels que l'invention de la pompe à vapeur destinée à l'exploitation minière par Thomas Savery en 1698 ou encore l'invention du métier à tisser par Joseph-Marie Jacquard en 1801 témoignent de la présence de premiers systèmes automatisés dans le monde industriel il y a de cela quelques siècles. Plus récemment, la révolution industrielle a contribué à l'essor de ces systèmes comme en témoigne de façon satirique le célèbre film *les temps modernes* de Charlie Chaplin, indéniablement inscrit dans la culture cinématographique. Depuis celle-ci, les systèmes automatisés se sont développés et sont devenus plus élaborés. Ils ne sont plus seulement limités à fournir des informations automatisées comme peuvent le faire des alarmes ou encore exécuter des actions comme cela a été fait dans les chaînes de production avec l'automatisation¹ des gestes répétitifs (par exemple avec des bras automatisés). Aujourd'hui, les systèmes impliquent des systèmes automatisés sophistiqués sous forme informatisée ou algorithmique qui analysent des

¹ Nous considérons l'automatisation comme le processus qui va remplacer l'exécution de certaines tâches effectuée par un opérateur humain par un système automatisé. Dans l'exemple, les bras robotisés remplacent les bras humains dans l'exécution de gestes répétitifs. Cette notion est définie plus précisément dans le chapitre 1, partie 1.1. Émergence des systèmes automatisés : définitions.

informations, telle que peut le faire une aide au diagnostic dans le secteur médical ; ou encore choisir des actions à faire et les appliquer comme peuvent le faire les derniers modèles de véhicule (par exemple les voitures tesla avec l'autopilotage).

Ces systèmes ont été mis en place avec plusieurs objectifs tels que l'amélioration de la sécurité, l'augmentation de la productivité ou encore soulager les opérateurs dans des situations trop répétitives, voire considérées comme trop complexes (Wickens, Hollands, Bandury, & Parasuraman, 2013). Ces avantages peuvent être quantifiés objectivement : par exemple, l'évolution positive du taux annuels d'accidents d'avions dans le secteur aéronautique depuis 1987 montrée dans le rapport sur la sécurité aérienne (Gandil, 2018) ; ou encore l'amélioration du rendement dans des manufactures canadiennes suite à l'utilisation d'une technologie (Baldwin & Sabourin, 2004). Ces avantages ont contribué à l'accroissement rapide de ces systèmes dans notre quotidien dès lors que leur implantation était techniquement possible. Cependant, outre ces avantages, des conséquences imprévues, parfois plus préjudiciables qu'avantageuses ont fait leur apparition découlant directement de l'interaction homme-machine. De nombreux accidents sont les témoins des difficultés engendrées par cette interaction tels que : l'accident nucléaire de *Three Mile Island* en 1979 dans lequel une partie du réacteur a fondu suite à un enchaînement de défaillances techniques et d'erreurs humaines ; ou encore plus récemment l'accident de vol Air France 447 entre Rio et Paris le 1^{er} Juin 2009 impliquant le givrage des sondes de vitesse entraînant une incohérence des vitesses mesurées et une déconnexion du pilotage automatique. La trajectoire de l'avion n'étant pas totalement maîtrisée par les pilotes, la collision de l'appareil avec la mer est survenue quelques minutes après la reprise de contrôle manuel par les pilotes.

De fait, mieux comprendre les causes et les conséquences de cette interaction est un enjeu important pour les spécialistes des facteurs humains puisque celle-ci connaît une évolution croissante due à l'évolution rapide des systèmes automatisés. Son étude est notamment essentielle afin d'en tirer des recommandations pour la conception de prochains systèmes automatisés. De nombreuses études ont ainsi souligné l'implication des fonctions d'allocations entre l'opérateur

humain et les systèmes automatisés sur les performances humaines (Lorenz, Di Nocera, Röttger & Parasuraman, 2001; Rovira, Zinni, & Parasuraman, 2002; Onnasch, Wickens, Li & Manzey, 2014). Parasuraman, Sheridan et Wickens (2000) ont proposé un cadre pour examiner les problèmes de conception et évaluer ce qui doit être automatisé ou non. Ces auteurs ont fourni une série d'étapes et de procédures à suivre en se focalisant sur deux critères : le premier se référant à l'impact des fonctions et niveaux d'allocations sur les performances humaines et le second incluant la fiabilité des systèmes automatisés et les conséquences engendrées directement par celle-ci. En effet, il peut arriver que ces systèmes automatisés soient imparfaits, c'est-à-dire que des défaillances peuvent se produire ou se produisent. Ces dernières peuvent engendrer des conséquences plus ou moins préjudiciables selon ces fonctions et niveaux d'allocations (Onnasch, 2015; Parasuraman & Manzey, 2010; Wickens & Dixon, 2007).

Dans la continuité de ces travaux, cette thèse vise à approfondir les connaissances sur les effets des systèmes automatisés imparfaits sur le comportement humain, tout en se basant sur l'approche des fonctions et niveaux d'automatisation² proposée par Parasuraman, Sheridan et Wickens (2000). Plus précisément, il s'agit de mettre en évidence les effets de la fiabilité des systèmes automatisés sur les performances humaines et le comportement humain en se focalisant sur la fonction information des systèmes automatisés. Le concept de fiabilité sera également discuté.

La contribution ergonomique de cette thèse est d'investiguer les effets de la fiabilité des systèmes automatisés en se focalisant plus particulièrement sur un seul type de prise charge du système automatisé (information). Premièrement, les études expérimentales contrôlées de cette thèse permettent une généralisation des résultats sur les systèmes automatisés prenant en compte la fonction information plutôt que celle de la décision. En effet, de nombreux travaux ont surtout

² Les niveaux d'automatisation correspondent à différents niveaux de prise en charge des tâches par les systèmes automatisés. Ces niveaux sont détaillés selon les différents modèles présentés dans le chapitre 1 de cette thèse.

investigé la fiabilité de la fonction décision. Deuxièmement, une étude appliquée à l'activité de planification interroge la fiabilité d'un système automatisé. Elle permet également d'enrichir les connaissances dans une activité où les études impliquant à la fois la fonction information des systèmes automatisés et la fiabilité sont quasi-inexistantes.

Ce manuscrit comprend deux parties principales, une partie théorique constituée de trois chapitres et une partie expérimentale organisée autour de trois contributions empiriques.

Le premier chapitre de la partie théorique revient sur les principales approches impliquant les systèmes automatisés de ces dernières années utilisées dans la littérature sur les facteurs humains et issues de perspectives différentes telles que l'ingénierie ou la psychologie ergonomique. Ce chapitre vise également à introduire le modèle de Parasuraman, Sheridan et Wickens (2000) impliquant différents niveaux d'automatisation, croisé à différentes fonctions de prise en charge des systèmes automatisés. Ce chapitre se termine par la proposition de ces mêmes auteurs, invitant les concepteurs à considérer plusieurs critères lors de la conception d'un nouveau système d'aide automatisé : les fonctions et niveaux d'automatisation puis la fiabilité de l'aide automatisée et leur conséquences respectives sur les opérateurs humains.

Le deuxième chapitre vise à interroger la notion de système automatisé imparfait (dont la fiabilité peut être remise en question) en présentant plusieurs types de prise en charge de l'information pouvant être considérés lors de la conception d'un système automatisé. A travers ces différentes considérations, ce chapitre vise également à élargir la vision d'application de la fiabilité afin d'amener les prochaines études à considérer l'impact des fonctions et niveaux de ces systèmes automatisés dans des secteurs spécifiques où la prise en charge de l'information est différente de celle que l'on peut voir dans la majorité des études citées dans la littérature. Dans ces secteurs, un système automatisé peut bel et bien aussi être implanté avec la mise en œuvre des mêmes fonctions et niveaux d'automatisation.

Le troisième chapitre porte sur l'interaction homme-machine et vise à faire un point sur les conséquences potentielles émanant de cette interaction. Ce chapitre énoncera les principales

différentes conséquences de l'interaction homme-machine présentes dans la littérature, notamment en cas de systèmes automatisés imparfaits.

A la suite de la partie théorique, nous présenterons la problématique de recherche qui encadre la partie expérimentale. Cette partie expérimentale est composée de trois grandes contributions empiriques s'appuyant sur ce cadre théorique.

Dans la première, nous avons évalué l'impact d'un système automatisé imparfait selon les deux fonctions considérées dans l'approche de Parasuraman, Sheridan et Wickens (2000). L'objectif de cette première étude a été de clarifier les effets de quatre niveaux de fiabilité selon deux fonctions de prise en charge du système automatisé et d'investiguer la présence simultanée de ces deux fonctions.

Dans la deuxième contribution, les effets d'un système automatisé imparfait ont été étudiés selon les mêmes quatre niveaux de fiabilité mais avec deux types de système automatisé. L'objectif de cette étude a été de tester les effets d'un système automatisé imparfait adaptatif ou statique selon quatre niveaux de fiabilité.

La troisième contribution a pour objectif d'investiguer les conséquences d'un système automatisé pouvant être considéré comme imparfait à travers l'activité de planification dans le secteur du transport routier de marchandises. Cette activité a permis d'évaluer plus en profondeur la prise en charge de la fonction information d'un système automatisé, dans une activité où cette information est cruciale pour la prise de décision.

Pour finir, une discussion générale viendra compléter ce manuscrit et fera le bilan des différentes études. Elle mentionnera également les perspectives émanant de ce travail de recherche.

PARTIE THEORIQUE

CHAPITRE 1 – LES SYSTEMES AUTOMATISES

Lorsque nous posons côte à côte deux photographies d'une chaîne d'assemblage de véhicules avec 70 années d'intervalle, les situations semblent similaires : un opérateur humain assemble les pièces d'un véhicule sur une chaîne d'assemblage. Pourtant, si nous nous attardons sur la figure 1, nous pouvons nous rendre compte que la situation a largement évolué : une soixante-dizaine d'années après la photographie de gauche, l'opérateur utilise un système automatisé sous forme de bras automatisé pour l'aider dans son activité d'assemblage. Cette situation témoigne de la progressive automatisation de tous les secteurs professionnels.



Figure 1. Des opérateurs humains travaillant sur une chaîne d'assemblage de véhicules automobiles en 1934 (à gauche) et en 2010 (à droite).

L'objectif de ce premier chapitre est de revenir sur les principales approches proposant des répartitions de fonctions entre l'opérateur humain et les systèmes automatisés au fil des années.

Nous présenterons d'abord les approches centrées sur les capacités se focalisant sur les compétences de l'un ou de l'autre pour cette répartition ; puis nous finirons sur les approches centrées sur les fonctions prenant en compte les fonctions de traitement de l'information humaine dans leur conception en vue d'une meilleure interaction.

1.1 L'ÉMERGENCE DES SYSTÈMES AUTOMATISÉS : DÉFINITIONS

Les systèmes automatisés se sont développés au fur et à mesure des dernières années dans une multitude de domaines, remplaçant souvent l'humain dans l'accomplissement d'une partie ou de toutes ses tâches. Cette automatisation des tâches auparavant effectuées par l'opérateur humain s'est vue attribuée de nombreuses définitions telles que « l'exécution par une machine (généralement un ordinateur) d'une fonction qui était auparavant exécutée par un opérateur

humain » (Parasuraman & Riley, 1997, p.231) ou encore la « technologie qui sélectionne activement des données, transforme des informations, prend des décisions ou contrôle des processus » (Lee & See, 2004, p.50). Moray, Iganaki et Ithoh (2000) ont proposé une définition reprenant les idées principales de ces précédentes définitions en se basant sur les étapes des modèles de traitement de l'information (Parasurama, Sheridan, & Wickens, 2000). Cette définition -que nous adopterons dans cette thèse- stipule qu'un système automatisé correspond à la prise en charge de « toute action de détection, de traitement de l'information, de prise de décision ou de contrôle qui pourrait être effectuée par l'homme mais qui est en fait effectuée par la machine » (Moray, Inagaki, & Itoh, 2000, p. 44). Ainsi, selon cette définition, un système automatisé peut correspondre à la fois à (1) la mécanisation et à l'intégration de la détection des variables environnementales par des capteurs artificiels ; (2) au traitement des données et à la prise de décision par le système automatisé ; (3) à l'action mécanique (par des moteurs ou des dispositifs qui exercent une force sur l'environnement) et/ou (4) à la communication d'informations traitées à l'opérateur humain (Sheridan & Parasuraman, 2005).

De nombreuses approches de conception se sont développées pour proposer des répartitions de tâches ou fonctions entre l'opérateur humain et le système automatisé. Dans la partie suivante, nous rappelons les principales approches. Nous finirons par le modèle de Parasuraman, Sheridan et Wickens (2000), modèle mobilisé dans cette thèse et intégrant ses prédécesseurs.

1.2. ATTRIBUTION DES FONCTIONS SELON LES CAPACITES DU SYSTEME AUTOMATISE

Dans un premier temps, dans de nombreuses situations professionnelles, le « tout automatique » était envisagé comme une solution idéale par les concepteurs de systèmes.

1.2.1. Fitts (1951)

D'après Dekker et Woods (2002), il fut un temps où la question « qu'est-ce qu'il faut automatiser ? » n'avait qu'une réponse qui était : tout ce qui est techniquement automatisable. Le système automatisé étant plus efficace que l'opérateur humain dans certaines fonctions, il était alors envisagé d'allouer les fonctions par substitution (Hollnagel, 1999). L'opérateur humain devait donc prendre en charge ce qui n'avait pas été automatisé ou ce que le système automatisé n'avait pas pu faire.

Cependant, l'introduction d'un système automatisé n'est pas sans conséquences puisqu'elle change la situation de travail et amène l'opérateur à interagir avec ce nouveau système. Bien que cette interaction puisse bien se passer et amener des avantages (cf. gain de temps, augmentation de la production, baisse des gestes répétitifs, etc.), celle-ci peut être caractérisée par des difficultés (surcharge d'informations qui entraîne un désengagement de l'opérateur humain, manque de souplesse du système automatisé) entraînant des conséquences négatives (cf. des accidents coûteux ou mortels). Ces conséquences parfois négatives ont amené un autre point de vue basé sur une division des tâches liées aux capacités du système automatisé ou de l'opérateur humain (Fitts, 1951). La liste de Fitts est considérée comme étant l'une des premières approches décrivant les répartitions de fonctions³ entre l'opérateur humain et le système automatisé. Dans cette perspective, une activité est découpée en fonctions élémentaires qui sont ensuite distribuées à l'agent – humain ou système automatisé – qui est le plus efficace (figure 2). Les fonctions qui sont mieux effectuées par le système automatisé sont automatisées, alors que les autres sont laissées à la responsabilité de l'opérateur humain.

³ Les notions de fonctions et de tâches sont différentes. Selon Hoc et Chauvin (2011), une tâche est souvent considérée comme une structure plus macroscopique qu'une fonction. Elle est également plus concrète, alors qu'une fonction est plus abstraite, et peut contribuer à plusieurs tâches (Fuld, 2000). Une tâche est aussi associée à un objectif précis. Alors qu'une fonction est : « un moyen sans aucun but en soi, mais juste lié à un objectif super-ordonné » (Hoc & Debernard, 2002, p.4).

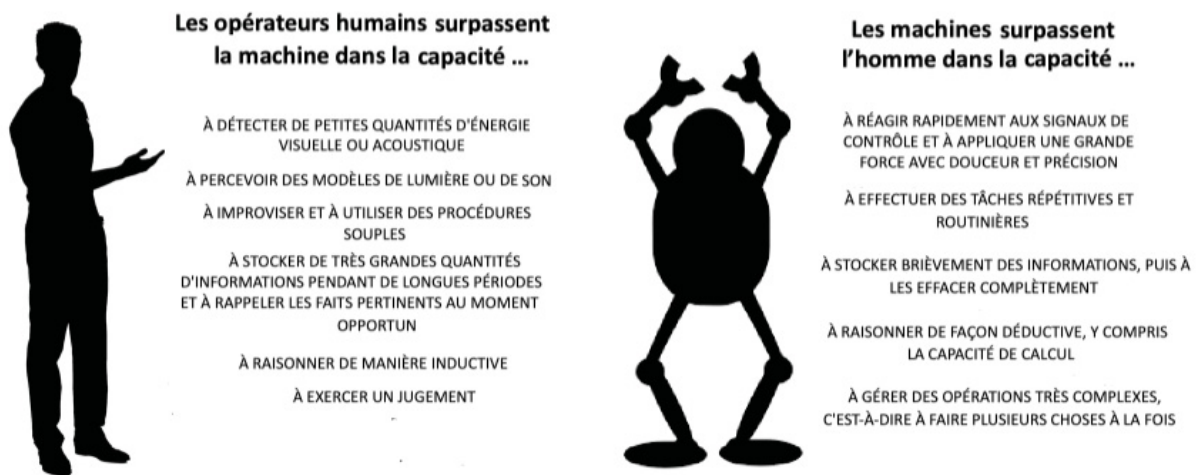


Figure 2. La liste de Fitts (1951).

Cette répartition communément nommée MABA-MABA (*Men are better at – Machines are better at*) part du principe que l'opérateur humain et le système automatisé ont des faiblesses et des forces qui sont fixes dans le temps. De ce point de vue-là, le rôle du concepteur va être de distribuer et de compenser les forces et des faiblesses de chacun.

1.2.2. Sheridan et Verplank (1978)

Une vingtaine d'années après, Sheridan et Verplank (1978) ont proposé une échelle en dix niveaux avec l'idée qu'automatiser ce n'est pas « tout ou rien » mais pouvoir décider de l'étendue de cette automatisation. Les auteurs ont identifié et proposé une échelle composée de dix niveaux allant d'un contrôle complètement manuel de l'opérateur à un contrôle entièrement automatisé, appliqués aux décisions et aux actions du système automatisé (figure 3). Le concepteur n'a plus qu'à choisir à quel niveau le système automatisé doit être implanté.

Dans le niveau 1, le système automatisé ne propose aucune assistance et l'humain doit donc accomplir la tâche sans aucune aide. Au fur et à mesure que les niveaux augmentent, de 2 à 9, l'autonomie du système automatisé dans la solution à apporter, la prise de décision et réalisation de

la tâche augmentent. Au dernier niveau, le système automatisé agit totalement en autonomie (figure 3).

L'objectif de Sheridan et Verplank (1978) était de définir « quel agent » (l'opérateur humain ou le système automatisé) a le contrôle et non de décrire explicitement comment un opérateur et un système automatisé pourraient partager leurs fonctions. Cette taxonomie est donc centrée sur les fonctions du système automatisé et ne permet pas de décrire comment la répartition des tâches est régulée entre l'opérateur humain et le système automatisé (Navarro, 2017).



















NIVEAUX	DESCRIPTION DES NIVEAUX		
1		UN CONTRÔLE ENTIÈREMENT MANUEL	
2		L'ORDINATEUR OFFRE UN ENSEMBLE COMPLET D'ALTERNATIVES DE DÉCISION/ACTION	
3		L'ORDINATEUR RÉDUIT LA SÉLECTION À QUELQUES ALTERNATIVES	
4		L'ORDINATEUR PROPOSE UNE ALTERNATIVE	
5		L'ORDINATEUR EXÉCUTE CETTE SUGGESTION SI L'OPERATEUR HUMAIN L'APPROUVE	
6		L'ORDINATEUR ACCORDE À L'OPERATEUR HUMAIN UN TEMPS LIMITE POUR METTRE SON VETO AVANT EXÉCUTION AUTOMATIQUE	
7		L'ORDINATEUR EXÉCUTE AUTOMATIQUEMENT, PUIS INFORME NÉCESSAIREMENT L'OPERATEUR HUMAIN	
8		L'ORDINATEUR N'INFORME L'OPERATEUR HUMAIN QUE S'IL LE DEMANDE	
9		L'ORDINATEUR N'INFORME L'OPERATEUR HUMAIN QUE S'IL LE DÉCIDE	
10		LE CONTRÔLE EST ENTIÈREMENT AUTOMATISÉ	

Figure 3. Les niveaux d'automatisation proposés par Sheridan et Verplank (1978).

1.3. ALLOCATION DES FONCTIONS ENTRE L'OPERATEUR HUMAIN ET LE SYSTEME AUTOMATISE
























































Ultérieurement, Billings (1991) a proposé de remettre l'opérateur humain au centre de la conception des systèmes automatisés. Comme l'ont souligné Sheridan et Parasuraman (2005) en

s'inspirant toujours de multiples accidents dans le secteur aéronautique, concevoir un système automatisé et attendre de l'homme qu'il s'y adapte pourrait être une recette conduisant à un désastre.

Les prochaines approches tendent donc à satisfaire la proposition de Billings (1991) avec la volonté de s'éloigner d'une conception des systèmes automatisés basée strictement sur les capacités de l'un ou de l'autre et d'intégrer plus de lien avec les théories cognitives (Endsley, 2018). Bien que ces approches soient encore largement « techno-centrées », l'intégration de fonctions liées aux théories cognitives témoignent d'un avancement dans le processus de conception. Dans ces approches, une vision plus collaborative dans l'interaction homme-machine est encouragée alors que précédemment, ils se partageaient les tâches, le contrôle... en fonction de leur performance ; implicitement, ils étaient « concurrents ».

1.3.1. Endsley et Kaber (1999)

Dans leur approche, les auteurs décrivent avec précision "quel agent" (l'opérateur humain ou le système automatisé) est censé faire "quoi" (tâche) selon quatre rôles majeurs identifiés dans l'interaction homme-machine et selon chaque niveau d'automatisation (figure 4). En représentant comment la surveillance du système, la génération, la prise de décision et l'implémentation des réponses peuvent être attribuées à un opérateur humain ou à un système automatisé, ou partagées entre les deux, cette taxonomie présente des avantages par rapport à celles proposées précédemment. Elle permet d'identifier de nombreuses combinaisons de niveau d'automatisation et de rôles partagés entre humain et le système, qui n'étaient pas incluses dans d'autres approches (Kaber & Endsley, 2004). Elle est également applicable à un plus large éventail de domaines (Vagia, Transeth, & Fjerdingen, 2016).

		RÔLES			
NIVEAUX D'AUTOMATISATION		SURVEILLANCE	GÉNÉRATION	SÉLECTION	IMPLÉMENTATION
Contrôle manuel					
Soutien à l'action		 			 
Traitement par lots		 			
Contrôle partagé		 	 		 
Aide à la décision		 	 		
Prise de décision mixte		 	 	 	
Système rigide		 			
Prise de décision automatisée		 	 		
Contrôle supervisé		 			
Automatisation totale					



 Opérateur humain
 Système automatisé

Figure 4. La taxonomie proposée par Endsley et Kaber (1999)

Quatre fonctions typiquement retrouvées dans des situations complexes dynamiques ont été identifiées par les auteurs : (1) la surveillance qui inclut la collecte d'informations pertinentes pour appréhender l'état du système automatisé ; (2) la génération qui formule des options ou des stratégies en vue de la réalisation d'un objectif ; (3) la sélection d'une option ou d'une stratégie particulière ; (4) l'implémentation des options choisies (Endsley & Kaber, 1999).

Les dix niveaux d'automatisation ont ensuite été attribués à l'opérateur humain ou au système automatisé (voir Kaber et Endsley, 2004, pour une description précise de chaque niveau d'automatisation).

1.3.2. Parasuraman, Sheridan et Wickens (2000)

Parasuraman, Sheridan et Wickens (2000) ont proposé une taxonomie en deux dimensions (figure 5). La première dimension s'inspire des dix niveaux impliquant la prise de décision et d'action du système automatisé proposés par Sheridan et Verplank (1978) à travers une échelle simplifiée d'autonomie du système automatisé allant de faible à élevé. Dans la deuxième dimension, ces niveaux d'automatisation sont considérés selon quatre fonctions qui correspondent aux quatre fonctions simplifiées du traitement de l'information humaine et inspirées du précédent modèle de Endsley et Kaber (1999). Un système automatisé peut être automatisé à différents niveaux (élevé ou faible) pour chacune des quatre fonctions. Ce principe est illustré par les systèmes A et B dans la figure 5.

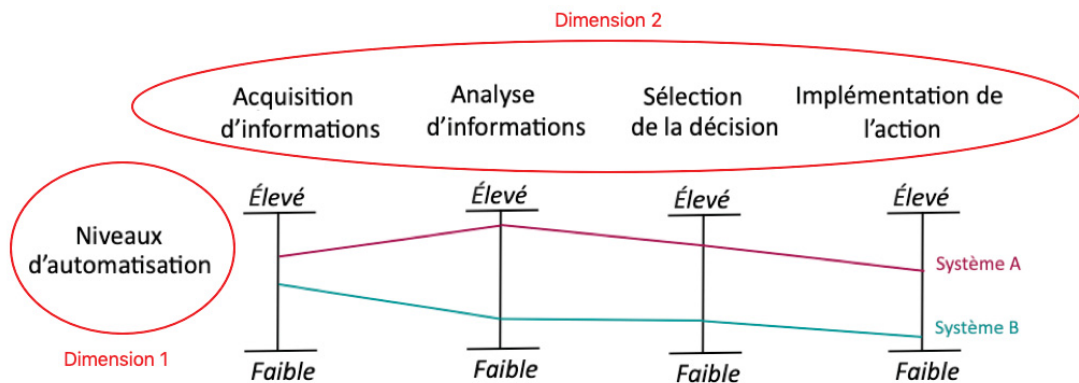


Figure 5. La taxonomie proposée par Parasuraman, Sheridan et Wickens (2000).

Plus précisément, la première fonction correspond à l'acquisition d'informations. Cette fonction implique la détection et l'enregistrement des données d'entrée. Ces opérations sont équivalentes à la première fonction de traitement de l'information humaine, soutenant les processus sensoriels humains. Par exemple, dans le trafic aérien, un niveau faible d'automatisation dans cette fonction peut correspondre à un balayage des données radar par des capteurs qui vont acquérir des informations (localisation, cap, vitesse, altitude) sur les avions présents dans une zone spécifique. Dans un niveau plus élevé de cette fonction, les informations entrantes peuvent être organisées selon certains critères tels que la priorité. Parasuraman, Sheridan et Wickens (2000) ont souligné que dans cette fonction, de faibles niveaux d'automatisation préservent la visibilité des informations

brutes (les données d'entrée) alors qu'avec des niveaux plus élevés, les informations brutes peuvent ne plus être transparentes (c'est-à-dire directement accessibles) pour les opérateurs humains suite au filtrage effectué. Par exemple, imaginons un balayage des données radar par des capteurs automatisés avec un niveau élevé. Ceux-ci priorisent les avions selon leur menace et affichent donc seulement dix avions qu'ils considèrent menaçant sur les trente qu'ils ont détecté. Ce filtrage ne rend plus transparentes les informations brutes puisque l'opérateur humain ne sait pas qu'il y a à la base 30 avions, le radar en mettant en lumière seulement une partie.

La deuxième fonction, l'analyse de l'information, implique la manipulation d'informations dans la mémoire de travail et des opérations cognitives telles que l'intégration, le diagnostic et l'inférence (Parasuraman & Wickens, 2008). Ainsi, de faibles niveaux d'automatisation peuvent correspondre à l'extrapolation d'une future trajectoire de vol dans le trafic aérien, ou encore à des niveaux plus élevés, de permettre l'aide à la visualisation de chemins qui se croisent ou convergent pour plusieurs avions.

La troisième fonction, la sélection de décision, se caractérise par la prise de décision. Les différents niveaux faibles ou élevés sont détaillés dans la taxonomie de Sheridan et Verplank (1978) et correspondent à différents niveaux d'augmentation ou de remplacement de la décision de l'opérateur humain. Dans des niveaux faibles, des propositions sont faites à l'opérateur humain qui peut par la suite décider laquelle choisir alors que dans des niveaux plus élevés, le système automatisé peut choisir automatiquement la décision si l'opérateur humain ne le fait pas (par exemple, les systèmes d'évitement des collisions au sol).

La quatrième fonction, l'implémentation de l'action, implique l'exécution de l'action. Les différents niveaux de prise en charge par le système automatisé seront en lien direct avec la quantité d'actions effectuées par l'opérateur humain pour de faibles niveaux jusqu'à l'absence d'actions effectuées par l'opérateur humain dans des niveaux plus élevés. Parasuraman, Sheridan et Wickens (2000) donne l'exemple du photocopieur où le tri manuel, le tri automatique, l'assemblage

automatique et l'agrafage automatique représentent différents niveaux d'automatisation de l'action qui peuvent être choisis par l'opérateur humain.

Au final, dans cette approche, le système automatisé peut être conçu pour prendre en charge certaines fonctions spécifiques (figure 5). Par exemple, Nysen et al. (2004) ont expliqué que dans le secteur hospitalier, certains systèmes pouvaient agréger des informations ou encore proposer des actions et donner des feedbacks. Un système peut avoir un niveau élevé dans les quatre fonctions d'une tâche, tandis qu'un autre système peut être très autonome dans l'acquisition d'informations mais utiliser un niveau moins élevé (impliquant plus d'intervention de l'opérateur) dans les trois autres fonctions (Calhoun, Ruff, Draper, & Wright, 2011).

De manière plus synthétique, comme le montre la figure 6, certains auteurs préfèrent simplifier les fonctions 1 et 2 à « automatisation de l'information », d'une part, et les fonctions 3 et 4 à « l'automatisation des décisions » (Navarro, 2017; Parasuraman & Wickens, 2008; Parasuraman & Wilson, 2008).

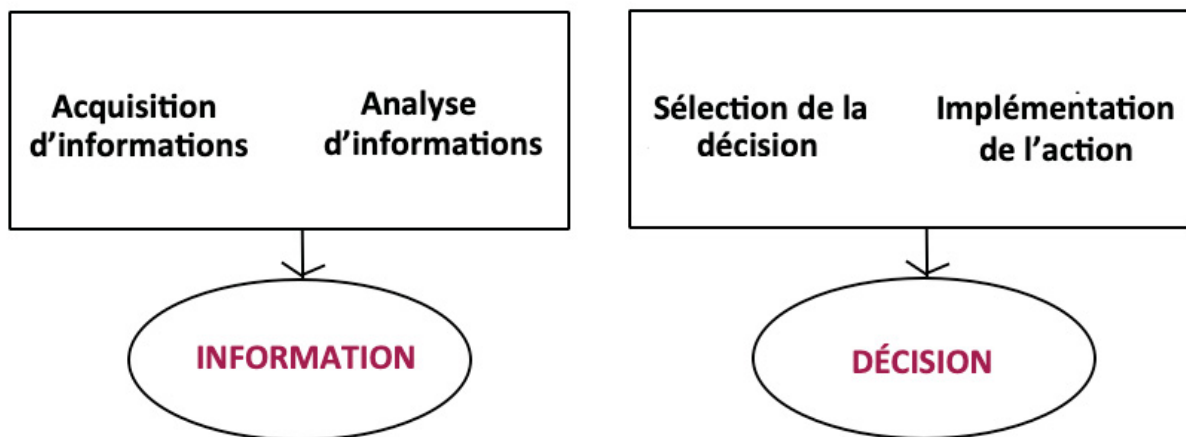


Figure 6. Synthèse des fonctions prises en charge par un système automatisé proposés par Parasuraman, Sheridan & Wickens (2000)

Cette simplification est proposée car les fonctions 1 et 2 d'information d'un côté et les fonctions 3 et 4 de l'autre sont étroitement liées. Parasuraman et Wickens (2008) donnent une explication neuroergonomique : les premières fonctions d'informations sont arbitrées par les

régions cérébrales postérieures (cortex occipito-tempo-pariétal) et les secondes par les régions antérieures (cortex frontal).

En lien avec leur approche des fonctions et niveaux d'automatisation, Parasuraman, Sheridan et Wickens (2000) ont également proposé une procédure schématique pour la conception de système automatisé (Figure 7)

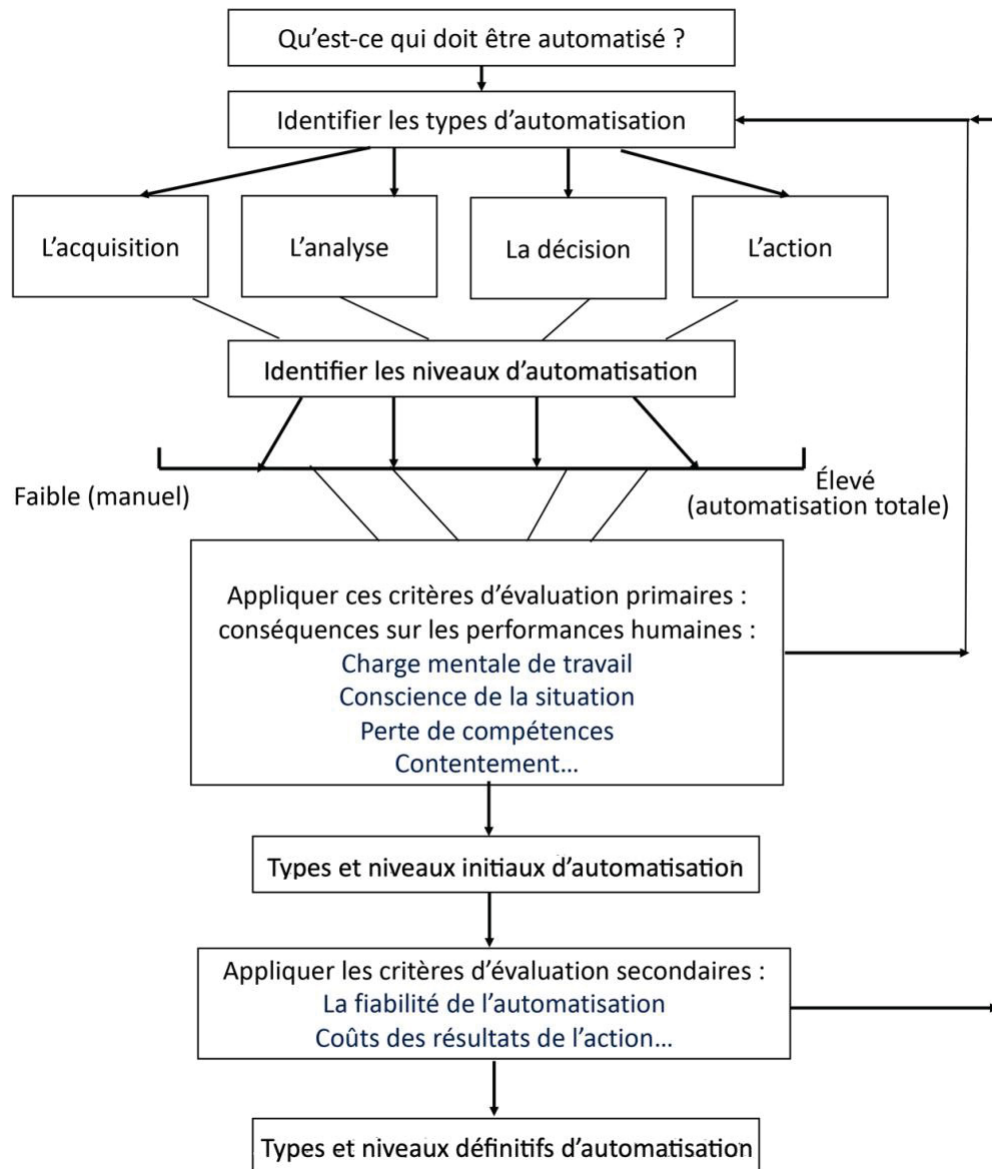


Figure 7. Organigramme montrant l'application du modèle proposé par Parasuraman, Sheridan et Wickens (2000) dans la conception de système automatisé.

Ceux-ci ont souligné les possibles coûts et bénéfices d'automatisation des fonctions et des niveaux d'automatisation. La première phase lors de la conception d'un système automatisé consiste à se poser la question suivante : qu'est-ce qui doit être automatisé : l'acquisition des informations ? L'analyse des informations ? La sélection de décision ou bien l'implémentation de l'action ? Est-ce que plusieurs de ces fonctions sont nécessaires en même temps ? A quel niveau d'automatisation est-il nécessaire d'automatiser chacune de ces fonctions ? Pour réfléchir à ces questions auxquelles il est peut-être difficile de donner une réponse, les auteurs proposent de se référer à deux critères d'évaluation. Les concepteurs doivent d'abord (1) évaluer les conséquences sur les performances humaines des différents niveaux d'automatisation selon les fonctions associées et (2) évaluer les conséquences de la fiabilité du système automatisé. En effet, si lors de la première évaluation, le système automatisé est considéré comme avantageux (par exemple pour réduire la charge mentale de travail de l'opérateur), il est probable que dans une seconde évaluation, si la fiabilité du système automatisé n'est pas assurée, ces avantages ne perdurent pas.

1.3.3. Hoc, Young & Blosseville (2009)

Cette approche intègre l'idée de coopération homme-machine qui est définie comme la gestion de l'interférence entre deux agents (opérateurs humains et/ou machines) dans le but de rendre plus facile les tâches individuelles et communes (Hoc, 2001). Trois niveaux de coopération sont proposés : la coopération dans l'action ; la coopération dans la planification et la méta-coopération.

La coopération dans l'action correspond aux activités coopératives des agents (opérateur et système automatisé) en temps réel pendant l'action à un niveau d'interférences locales et à court terme.

La coopération dans la planification correspond à un niveau d'abstraction plus élevé que l'action et implique des activités coopératives à moyen terme qui visent la création et à la mise un jour d'un cadre de référence commun. L'opérateur humain et le système automatisé partagent une représentation commune de l'environnement et de l'activité de l'équipe par le biais d'objectifs, de plans et de rôles à partager.

La méta-coopération correspond à un niveau élevé d'abstraction plus élevé que la planification. Elle vise l'optimisation de la coopération sur le long terme par le biais de la construction d'un modèle de soi-même ou du partenaire (Navarro, 2017). Hoc, Young, et Blosseville (2009) ont proposé de catégoriser quatre niveaux d'interférences appliquées à la conduite automobile. La figure 8 introduit ces modes : le mode perceptif, le mode de contrôle mutuel, le mode de délégations des fonctions et le mode complètement automatisé (figure 8).

<p>MODE PERCEPTIF</p>	<p>MODE DE CONTRÔLE MUTUEL</p> <p><i>Mode avertissement</i> <i>Mode suggestion d'action</i> <i>Mode limitatif</i></p>	<p>MODE DE DÉLÉGATIONS DES FONCTIONS</p>	<p>MODE COMPLÈTEMENT AUTOMATISÉ</p>
------------------------------	--	---	--

Figure 8. Modes de coopération du modèle de Hoc, Young, & Blosseville (2009)

Dans le mode perceptif, le système automatisé va agir comme une extension des capacités sensorielles de l'opérateur humain. Par exemple, pour le contrôle latéral en conduite automobile, l'amélioration de l'aspect visuel des bords de route à l'aide d'affichages tête haute aide le conducteur dans sa tâche de suivi des bords de route en vision nocturne (Navarro, 2017).

Dans le mode de contrôle mutuel, l'opérateur humain et la machine vont effectuer la tâche en parallèle et chacun peut contrôler ce que l'un ou l'autre effectue mutuellement. D'après (Payre, 2015), le mode de contrôle mutuel renvoie à plusieurs possibilités : un mode avertissement (par exemple une alarme sonore pour prévenir d'une collision), une mode de suggestion d'action (par exemple, un signal lumineux sur le tableau de bord pour conseiller à l'automobiliste de passer une vitesse), un mode limitatif (par exemple, la pédale d'accélération qui se rigidifie lorsque la vitesse du véhicule dépasse celle du limiteur de vitesse). Il peut arriver que les conducteurs soient distraits ou surpris et que la situation soit trop difficile à gérer par eux-mêmes (risque de collision imminente). Dans ce cas-là, une action du système automatisé est nécessaire (par exemple, la voiture freine automatiquement). Le système automatisé « critique » le comportement des conducteurs et par conséquent, les aide à améliorer la sécurité (Bellet, Hoc, Boverie, & Boy, 2011).

Dans le mode de délégation des fonctions, l'opérateur humain décide de confier une ou plusieurs parties des tâches au système automatisé.

Dans le mode complètement automatisé, toutes les tâches sont prises en charge par le système automatisé.

1.4. CONCLUSION INTERMEDIAIRE

La présentation des différentes approches permet de montrer que celles-ci se sont influencées les unes et les autres au cours des années. Lorsque les systèmes automatisés sont apparus, les premières approches d'attribution de tâches entre l'opérateur humain et le système automatisé se sont focalisées sur les capacités de l'un ou de l'autre ; ainsi que sur la quantité de décisions ou d'actions effectuées par ces derniers. Quelques années après, les conceptions ont évolué vers une prise en compte plus importante des interactions, allant même jusqu'à des modèles de coopération.

Parasuraman, Sheridan et Wickens (2000) ont proposé un organigramme des fonctions et des niveaux d'interaction humaine avec le système automatisé afin de guider les concepteurs de systèmes automatisés pour qu'ils choisissent les fonctions qui doivent être automatisées et dans quelle mesure celles-ci doivent l'être. Durant la conception des systèmes automatisés, il est nécessaire d'évaluer les conséquences sur les performances et sur le comportement des opérateurs humains afin d'éviter de créer davantage de difficultés dans l'interaction homme-machine que de bénéfices. Parasuraman, Sheridan et Wickens (2000) ont ainsi proposé une évaluation sur deux critères lors de la conception des systèmes automatisés : évaluer les fonctions et les niveaux prise en charge par le système automatisé sur les performances humaines, et également évaluer la fiabilité de ces systèmes automatisés qui peut agir sur l'interaction et sur les performances. Le chapitre suivant propose de revenir sur la notion de fiabilité.

CHAPITRE 2 : SYSTEME AUTOMATISE

IMPARFAIT

Un système automatisé peut être imparfait lorsque sa fiabilité n'est pas parfaite, le système automatisé n'accomplissant pas sa tâche parfaitement, en faisant des erreurs. La fiabilité du système automatisé est souvent impliquée dans les difficultés suscitées par l'interaction homme-machine. Cette fiabilité est même considérée comme un critère important d'évaluation par Parasuraman, Sheridan et Wickens (2000) dans leur procédure de conception et d'évaluation des systèmes automatisés. Après avoir clarifié les différentes approches de conception proposées au cours des dernières décennies, ce deuxième chapitre propose de clarifier le concept de fiabilité.

2.1. LA FIABILITE D'UN SYSTEME AUTOMATISE:

2.1.1. Définitions

La fiabilité peut être définie comme une « aptitude d'un dispositif à accomplir une fonction requise, dans des conditions données, pendant une durée donnée. Elle est donc fonction du temps et peut être quantifiée par une probabilité, un pourcentage de succès, mais surtout par un pourcentage de défaillance ou d'erreur. » (Boukerma, 1998). Dans la littérature sur l'interaction homme-machine, la fiabilité est communément évaluée par le prorata du nombre d'erreurs du système automatisé. C'est-à-dire, le nombre d'opérations correctes divisé par le nombre total d'opérations dans lesquelles une tâche est automatisée (Parasuraman, Molloy, & Singh, 1993).

C'est cette vision que l'on retrouve dans les études impliquant les systèmes automatisés imparfaits étudiés dans des situations contrôlées en laboratoire telles que des simulations ou des micromondes⁴ simplifiés (Bagheri & Jamieson, 2004 ; Onnasch, Ruff, & Manzey, 2014 ; Parasuraman, Molloy & Singh, 1993 ; Rovira, McGarry, & Parasuraman, 2007). De cette façon, la fiabilité est plus facile à manipuler en modulant le nombre d'actions correctement effectuées par le système automatisé. Il est souvent reproché à ces études d'être éloignées de la réalité en proposant

⁴ Les micro-mondes sont des simulations informatisées de tâche réelles pour étudier les processus mentaux des opérateurs dans cette tâche de manière plus contrôlée (Rigas, Carling, & Brehmer, 2002).

une « concentration » d'erreurs du système dans un laps de temps réduit. Cette concentration nuit à la validité écologique (Hoc, 2001) ; en effet, dans ces secteurs, le nombre d'erreurs du système automatisé est plutôt faible et elles sont peu fréquentes. De plus, il n'est pas toujours évident de les objectiver puisque les décisions sont multicritères (Mouloua & Hancock, 2019; Parasuraman & Manzey, 2010). Cependant, en situation réelle, il peut être compliqué d'observer ce type de situation puisqu'elles sont rares et que nous ne pouvons pas les provoquer, d'où l'intérêt de les reproduire en simulation afin d'en étudier expérimentalement les effets.

Dans cette thèse, nous nous intéressons plus particulièrement à la fonction information prise en charge par les systèmes automatisés. Il est à noter qu'une information peut être de plusieurs types. Un système automatisé peut donner **une information par un signe**, par exemple une alarme visuelle qui s'allume tel qu'un feu de circulation qui passe au rouge pour dire qu'un véhicule doit s'arrêter ; mais aussi par **une information textuelle (symbole)**, par exemple, l'information présente sur un système de navigation « Vous arriverez à Toulouse dans 20 minutes ». Cette différenciation des termes est proposée par Rasmussen (1983) dans le cadre du modèle *Skill Rule Knowledge* (SRK). Le modèle SRK aborde le contrôle du comportement cognitif et le traitement de l'information selon trois niveaux :

Le premier niveau basé sur les compétences (*Skills*) implique des réactions automatiques à des événements. En ce qui concerne le niveau perceptif, les informations sont perçues sous formes de signaux. Selon Vaudry (2003), les signaux sont des données sensorielles représentant des variables spatio-temporelles provenant d'une configuration spatiale dynamique de l'environnement.

Le second niveau basé sur les règles (*Rules*) implique l'exécution de règles ou procédures en fonction d'un événement spécifique. Cegarra et al. (2017) décrivent ce niveau comme impliquant l'exécution de règles conscientes et de procédures connues. De ce fait, les informations sont perçues sous forme de signes. Un contrôle par anticipation est utilisé pour réponse à l'événement puisque la règle ou la procédure est acquise avant l'événement par l'expérience ou la formation.

Le troisième niveau basé sur le raisonnement (*Knowledge*) implique la mobilisation de connaissances face à des situations nouvelles ou imprévues. Dans ce niveau, l'opérateur humain va devoir construire des raisonnements pour répondre aux objectifs en se basant sur ses connaissances et sans règles préétablies à suivre. L'information est traitée sous forme de symbole et va nécessiter de faire appel à ses connaissances.

La fiabilité du système automatisé « feu de circulation routière » peut être calculée par le nombre de fois où le « feu rouge » ne s'allume pas pendant 1h. Si le feu rouge ne s'allume pas 5 fois sur 10, je peux dire qu'il est fiable à 50%.

La fiabilité du système automatisé « système de navigation » peut être calculée par le nombre d'informations présentées par le système qui sont erronées. Par exemple, imaginons l'affichage de deux informations : vous arrivez à Toulouse dans 20 min ; Vous arrivez dans 10 Km. Si en réalité une des informations est fautive, et qu'en fait j'arrive dans 30 minutes à Toulouse en conduisant dans une route sans interférences dans la conduite, alors mon système de navigation est fiable à 50% car une information sur deux est erronée. Cependant, un opérateur humain peut estimer différemment la fiabilité de l'information : si 10 min sont considérées dans cette situation comme acceptable/ négligeable alors l'information selon laquelle j'arrive dans 20 min est considérée comme juste. La fiabilité perçue du système pour cet opérateur est alors de 100%. Cette fiabilité est perçue différemment selon le jugement de chacun et va être évaluée au regard des actions. Ce jugement se crée sur un certain nombre de principes (la propre perception du temps qu'il faut pour aller à Toulouse selon le vécu passé, la confiance envers les calculs du système de navigation ; l'écart de temps acceptable pour l'opérateur...) propres à chaque personne.

Dans cette thèse, nous allons étudier la fiabilité d'un système automatisé prenant en charge la fonction information. Nous nous sommes intéressées à la fiabilité de deux types d'informations : celles sous forme de signe et celles sous forme de symboles et ceci, dans deux secteurs d'application différents (figure 9).

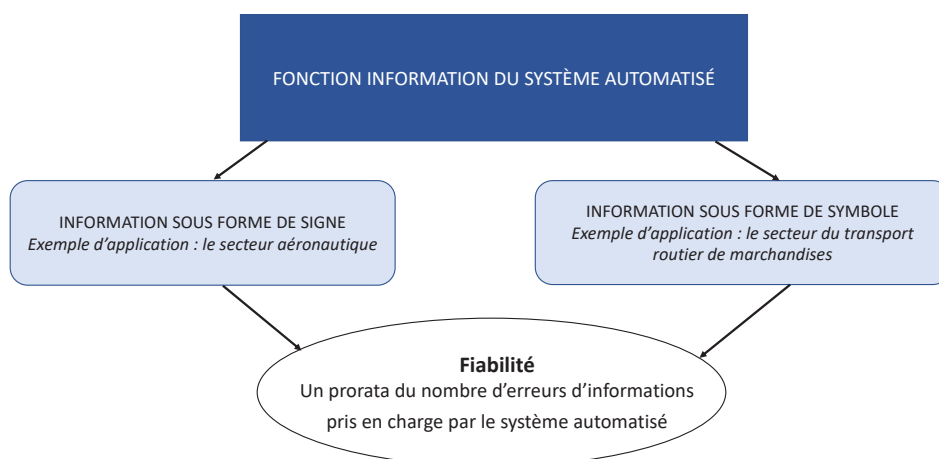


Figure 9. Les types d'informations pouvant être pris en charge par le système automatisé et la fiabilité.

2.1.2. Exemple d'application : le secteur aéronautique

L'avancée des systèmes automatisés et les possibles conséquences en cas de défaillances sont des sujets qui ont beaucoup interrogé les spécialistes des facteurs humains, particulièrement dans le secteur aéronautique (Parasuraman & Wickens, 2008). Il est important de se rappeler que l'automatisation d'un système ne signifie pas retirer l'humain de l'équation (Payre, 2015). Dès 1980, Wiener et Curry (1980) interrogeaient les conséquences de l'introduction des systèmes automatisés dans le secteur aéronautique sur les opérateurs humains. Cette mise en œuvre de l'automatisation dans les systèmes a certes amélioré le fonctionnement du système (cf. avec des progrès techniques tels que le pilotage automatique), mais les opérateurs ont commis des erreurs liées à l'introduction de ces systèmes automatisés qui ont pu conduire à des accidents (Strauch, 2017). En effet, les systèmes d'aviation sont des systèmes complexes qui engendrent des coûts inattendus de l'automatisation : une augmentation involontaire de la charge de travail (Sarter, Woods, & Billings, 1997), l'altération des performances de surveillance (Parasuraman & Manzey, 2010), la perte d'expertise (Hoc, 2000) ou la réduction de la conscience de la situation pour l'opérateur humain (Endsley, 1996). De multiples accidents sont les témoins de ses difficultés dans le monde réel

(Wiener, 1981). De nombreux rapports techniques d'accidents évoquent les difficultés de surveillance des systèmes automatisés (*monitoring inefficiency*) ou de perte de conscience de la situation (*loss of situation awareness*). Ces rapports permettent d'appréhender et de faire des hypothèses de recherche sur ces phénomènes. Cependant, étudier et décrire empiriquement ces phénomènes est devenu une nécessité pour les spécialistes des facteurs humains afin de les comprendre et de les anticiper, notamment en vue de la conception de futurs systèmes automatisés plus adaptés (Hoc, 1993; Parasuraman & Manzey, 2010).

Ainsi, dans leur étude, Parasuraman, Molloy, et Singh (1993) font varier la fiabilité d'un système automatisé par le nombre d'échecs de ce système automatisé à prendre en charge la détection de pannes. Ils ont souhaité évaluer le phénomène de contentement et approfondir les connaissances sur ce dernier, connu comme une insuffisance de surveillance et évoqué, suite à des accidents, dans le monde aéronautique (Billings, Lauber, Funkhouser, Lyman, & Huff, 1976; Parasuraman & Manzey, 2010). Pour ce faire, ils ont utilisé un micromonde de l'aéronautique, le MATB (*Multi Attribute Task Battery*, figure 10) dans lequel une tâche sur trois était automatisée (Santiago-Espada, Myer, Latorella, & Comstock, 2011). Dans cette tâche automatisée, les participants devaient débloquer le curseur en appuyant sur des touches spécifiques de leur clavier lorsque le système se bloquait. En fonctionnement normal, le système automatisé présent dans cette tâche débloquait automatiquement les curseurs, rendant l'opérateur superviseur du système. Le niveau de fiabilité était opérationnalisée par le nombre de gestion de la mauvaise position du curseur par le système automatisé. Ainsi, sur un total de 16 positions « anormales » du curseur, si le système en débloquait 14 alors il était fiable à 87.5% ; de même que s'il en débloquait 9, il était fiable à 57.25%. A travers cette opérationnalisation de la fiabilité, les auteurs ont confirmé expérimentalement l'existence du phénomène de contentement et d'une conséquence de ce phénomène inattendu : une insuffisance de surveillance.

Le Multi Attribute Task Battery (MATB)

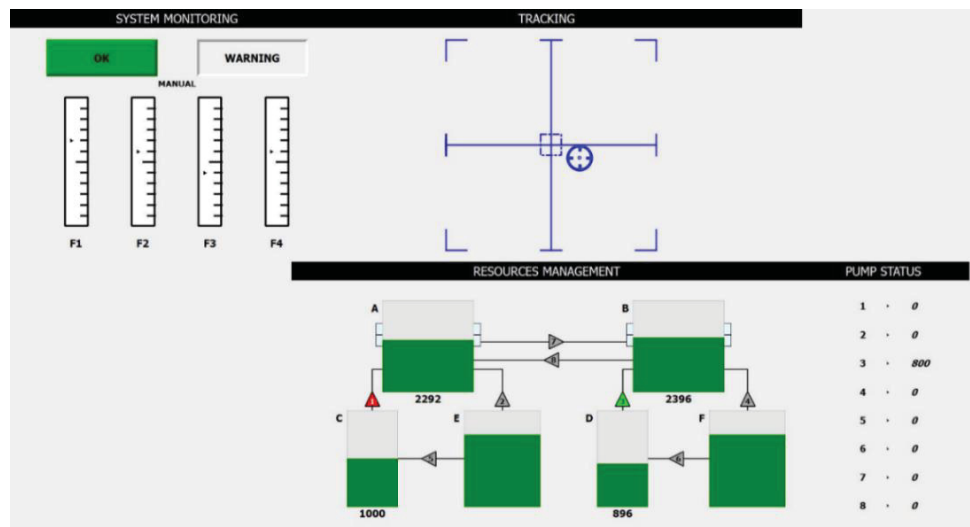


Figure 10. Illustration du OpenMATB proposé par Cegarra, Valéry, Avril, Calmettes et Navarro (2020), une version open-source du MATB.

Composé de trois tâches simultanées, le MATB est un micro-monde consistant en trois tâches représentatives de celles effectuées lors du pilotage d'avion. Sa mise en œuvre ne requiert pas forcément d'expertise approfondie ce qui rend son utilisation possible avec tous types de participants. La première tâche (en haut à gauche) de surveillance (system monitoring task) est construite autour de quatre colonnes dans chacune desquelles un curseur fluctue de haut en bas. Lorsqu'un des curseurs se bloque, les deux lumières d'alarmes changent et le curseur fluctue anormalement au centre de la colonne. Si cette tâche est automatisée, le système automatisé va débloquent automatiquement le curseur. Cependant, si le système automatisé est défaillant, les participants doivent débloquent eux-mêmes le curseur en appuyant sur les touches correspondantes du clavier. La seconde tâche (en haut au centre) de cible (tracking task) consiste à garder un curseur qui bouge dans une zone cible à l'aide d'un joystick. La troisième (en bas au centre) de gestion des ressources (resources management task) consiste en la gestion de plusieurs réservoirs principaux qui se vident continuellement. Le participant doit suivre un certain nombre de règles pour activer des pompes donnant accès au liquide des réservoirs de secours afin de renflouer les réservoirs principaux et de maintenir un niveau acceptable de liquide.

2.1.3. Exemple d'application : le secteur du transport routier de marchandises

Dans le secteur du transport routier de marchandises, les systèmes automatisés peuvent prendre la forme de systèmes de support à la décision (Cegarra & Van Wezel, 2012). A travers ces systèmes, la prise en charge de la fonction information est présente par le biais de signe et aussi de manière textuelle (symbole). Cette information textuelle peut également avoir un sens pour l'opérateur humain.

Afin de mieux comprendre la fiabilité dans le contexte de la planification dans le secteur du transport routier de marchandises (TRM), une première sous-partie revient d'abord sur l'activité d'exploitant⁵ (figure 11) puis en second lieu présente les systèmes automatisés et leur représentation dans ce secteur.

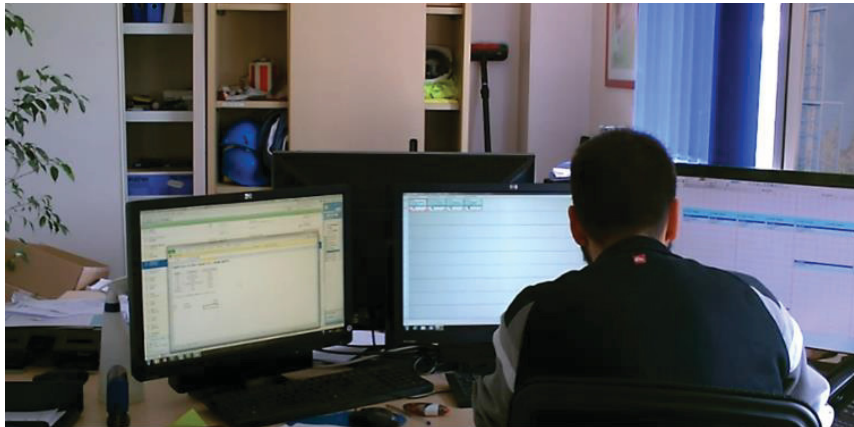


Figure 11. Un exploitant à son poste de travail.

2.1.3.1. *L'activité de planification dans le secteur du transport routier de marchandises*

Aujourd'hui, le secteur du TRM est un secteur important de l'économie nationale et mondiale puisqu'environ 90% des marchandises circulent, à un moment donné ou un autre, par voie terrestre (Tatineni & Demetsky, 2005). Ce secteur se caractérise par des organisations en juste

⁵ L'exploitant peut aussi être nommé planificateur puisque la planification est pris en charge par ce dernier.

à temps, une réglementation très dense, une forte compétitivité, une pression des clients et des fournisseurs, des délais serrés, une flexibilité exacerbée. À ces constats s'ajoute un accroissement des facteurs de compétitivité et de flexibilité, lié notamment à l'explosion du e-commerce (progression des ventes sur internet de 675% en 10 ans) et au développement de l'économie collaborative pour la livraison urbaine (par exemple l'ubérisation du transport). Ce secteur est confronté à plusieurs enjeux de types différents : économiques, environnementaux ou encore santé/sécurité (ce secteur concerne 400000 salariés qui, en termes de sinistralité, sont 2 à 3 fois plus souvent accidentés que ceux des autres secteurs d'activités.). Dans ce contexte particulier, l'exploitant a une position stratégique (Khademi, 2016). Ainsi, au carrefour des échanges entre la direction, les conducteurs, et les clients, son activité s'articule principalement autour :

- de la planification et du suivi des tournées de transport. Ainsi, l'exploitant est en charge de choisir le véhicule selon la nature de la marchandise, de réaliser les plannings en affectant les conducteurs et de fixer les heures de départ et d'arrivée. Il fixe également les fenêtres temporelles des livraisons et l'ordre de passage lorsque plusieurs destinataires sont concernés par la même tournée tout en respectant différents aspects (réglementation du travail, réglementation du transport, réglementation de la route). Pendant un transport, l'exploitant devient également un superviseur en restant en contact avec les chauffeurs pour apporter du soutien, de nouvelles consignes, assurer le suivi de commande aux clients tout en participant à la gestion des imprévus tels que les pannes, les bouchons liés à des accidents de la circulation, les retards chez les clients, etc. (Wioland & Govaere, 2014). Ces imprévus sont une des principales raisons des incertitudes présentes dans ce métier. Trois grandes sources d'incertitudes sont identifiables : (1) la demande du client qui peut être modifiée (2) l'instabilité du processus de transport avec des imprévus tels que cités ci-dessus (3) l'incertitude liée au chargement et déchargement tels que les ressources disponibles sur place, la durée du chargement, etc. (Khademi, 2016). Celui qui va exécuter ce que

L'exploitant a planifié est le conducteur et les conditions dans lesquelles il va la mettre en œuvre est en étroite interaction avec l'activité du planificateur (Wioland, 2013)

- d'une activité de « relation clientèle ». Celle-ci vise à constituer un stock de commandes de client. Ce stock comprend pour chaque commande des informations telles que le lieu et la date (heure) d'enlèvement et de livraison de la marchandise, le volume et les spécificités de la marchandise (type de produits, conditionnement...), les demandes spécifiques (opérations à réaliser, équipement de manutention nécessaire ou disponible, traçabilité à mettre en œuvre...). Pour produire ce stock, les exploitants vont collecter les demandes régulières des clients, c'est-à-dire celles dont la fréquence est fixée par contrat, et renseigner les volumes à transporter. Les exploitants ont également une activité de type commerciale pour rechercher des commandes supplémentaires et/ou de nouveaux clients ;

Les exploitants ne gèrent pas un unique flux de marchandises et d'informations mais plusieurs. Ces derniers sont hétérogènes à l'entrée de l'entreprise (demandes diversifiées de clients multiples, outils spécifiques) et/ou en sortie (services spécifiques effectués pour chacun des clients), ce qui accentue la complexité de leur traitement (Govaere & Wioland, 2018a ; Govaere & Wioland, 2018b ; Govaere, Wioland, Cegarra, Gourc, & Clement, 2018).

Au final, même si la planification n'est pas la seule activité de l'exploitant, elle représente une majeure partie de son temps. Par exemple, lors de la synthèse de plusieurs études de cas, Khademi, Cegarra, Govaere et Wioland (2015) ont montré que les traitements cognitifs pendant la planification de tournées sont l'élaboration d'informations (55% de son activité) ; les activités de diagnostics (13%) ainsi que les décisions d'intervention (32%) (figure 12).

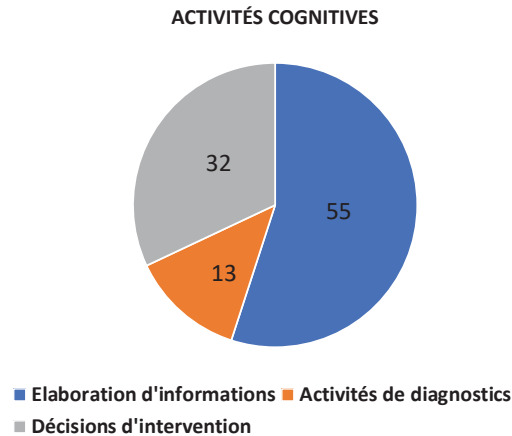


Figure 12. Répartition des activités cognitives pendant l'activité de planification chez un planificateur d'après Khademi (2016).

Lors de l'élaboration de cette planification, l'opérateur doit jongler entre de multiples contraintes et effectuer des compromis. Govaere et Wioland (2018b) ont hiérarchisé et catégorisé un ensemble de contraintes suite à des entretiens et expérimentations avec des exploitants du secteur de transport de marchandises. Plus précisément, elles ont souligné que cette gestion des contraintes était particulièrement liée à la situation et qu'elle n'était pas forcément la même selon la connaissance -et donc l'incertitude- de l'exploitant sur la situation. Leurs travaux ont également permis de montrer que la validité d'une tournée de transport était construite en considérant deux types de critères de performance : (1) une performance interne à l'entreprise : c'est-à-dire, une utilisation adaptée financièrement des ressources matérielles et humaines et du temps de livraison permettant une marge économique positive ; (2) une performance externe liée à la qualité du service client (répondre aux exigences des clients : livrer dans les temps, la marchandise adéquate, etc.). Cette validité est cependant contextuelle, une tournée valide étant dépendante du contexte de l'entreprise à un moment T. Une tournée peut être considérée valide un jour et moins valide le lendemain.

2.1.3.2. Les aides à la planification et leur représentation

Van Wezel et Jorna (2009) ont noté que les planificateurs se servent peu de logiciels sophistiqués. La plupart du temps, ils utilisent un traitement de texte ou une feuille de calcul. Même

si des algorithmes ou la génération automatique de plans sont disponibles dans les suites logicielles de transport, les planificateurs ont tendance à ne pas utiliser ces aides automatisées (Van Wezel, Cegarra, & Hoc, 2010). L'une des raisons pourrait être le manque d'une grande flexibilité pour le contrôle humain ou la difficulté liée directement au secteur : incertitude et instabilité.

La conception d'un système d'aide à la décision dans le secteur TRM semble complexe dans la mesure où il s'agit de problèmes non structurés. Cegarra et Van Wezel (2012) expliquent que la tâche de planification pouvait être mal documentée et les critères de solution pouvaient être mal définis. Il existe également de nombreuses contraintes non déclarées, des informations inexacts ou imprécises qui rendent complexe la conception de l'aide automatisée. Dans une étude, Cegarra (2008) montre que les planificateurs utilisent des processus de pensée différents des algorithmes généralement utilisés lorsqu'ils planifient. Par ailleurs, Kerr (1992) a mis en évidence les échecs de mise en place d'un système expert pour la planification ainsi que des explications à ces échecs : la complexité de la tâche, l'incertitude et les difficultés dans l'interaction entre système et opérateur humain.

Établir un critère de performance de la planification est aussi difficile en raison des priorités et objectifs différents que peuvent se donner chaque planificateur (Gary, Uzsoy, Smith, & Kempf, 1995) ou encore, qui évoluent selon le contexte de l'entreprise. Cegarra (2004) a expliqué que même si un système d'aide à la décision satisfait à toutes les contraintes, il est possible que l'expert humain rejette la solution car elle ne satisfait pas aux règles implicites. Par exemple, dans un atelier industriel, un expert peut constater que l'ordonnancement proposé par un système d'aide à la décision est trop sensible aux intempéries et rejeter la solution ; tandis que les autres experts peuvent accepter une solution qui ne satisfait pas à toutes les contraintes et la juger de bonne qualité (Jüngen & Kowalczyk, 1995).

Au final, bien que la fiabilité puisse être évaluée par un nombre d'erreurs du système automatisé (cf. le nombre d'informations erronées qu'il donne), la fiabilité perçue de ces informations fournies par le système est une notion importante à prendre en compte dans ce secteur. La fiabilité perçue

du système peut être considérée comme une réponse satisfaisante du système automatisé aux exigences de l'opérateur humain imposées par la situation et le contexte. Ce jugement de validité perçue est lié à la représentation de chaque exploitant. Cette représentation est personnelle, chaque exploitant ayant sa propre expérience et ses propres pondérations des critères (contraintes à prendre en compte, etc.) impliqués dans chaque performance.

2.2. CONCLUSION INTERMEDIAIRE

Ainsi, ces deux situations bien que différentes, montrent la supervision d'opérateurs avec des systèmes automatisés qui ne peuvent pas être considérés comme parfaitement fiable. La fiabilité est considérée comme un prorata d'un nombre d'erreurs du système automatisé. Lorsqu'elle est appliquée à la prise en charge de la fonction information d'un système elle est calculée par le nombre d'informations fiables fournies par le système automatisé. Cette information peut être fournie de deux façons différentes : elle peut prendre la forme d'un signe (visuelle, auditif...) ou elle peut prendre une forme textuelle (symbole).

La fiabilité d'un système automatisé influence l'interaction homme-machine. Parasuraman, Sheridan et Wickens (2000) l'avaient déjà évoqué dans leur organigramme comme critère à prendre en compte dans la conception des systèmes automatisés (figure 1). Plusieurs modèles ont aussi souligné l'importance de celle-ci dans l'interaction homme-machine tels que le modèle de Endsley (2016) ou le modèle proposé par Parasuraman et Manzey (2010). Tous ces modèles évoquent les possibles conséquences pouvant être engendrées par la fiabilité des systèmes automatisés et soulignent l'importance de bien penser la conception de ceux-ci. Le prochain chapitre propose de revenir plus précisément sur ces difficultés engendrées par l'interaction homme-machine, notamment en cas de système automatisé imparfait.

CHAPITRE 3 : INTERACTION HOMME- MACHINE

Après avoir clarifié le concept de fiabilité, l'objectif de ce deuxième chapitre est cette fois-ci de faire un point sur les difficultés engendrées par l'utilisation de ces systèmes automatisés, notamment en cas de systèmes automatisés imparfaits.

3.1. LES DIFFICULTES ENGENDREES PAR L'INTERACTION HOMME-MACHINE

L'introduction des systèmes automatisés est parfois défendue par des intentions telles que l'augmentation de la sécurité, la réduction de la charge mentale, le remplacement des gestes répétitifs... Toutefois, ces avantages peuvent être assombris par des conséquences imprévues qui viennent rajouter des difficultés qui n'étaient pas toujours présentes au départ et induites par cette interaction (Parasuraman & Riley, 1997). C'est ce que Bainbridge (1983) nomme les « ironies de l'automatisation », rappelant au passage que, bien que les concepteurs de systèmes reprochent aux opérateurs d'être parfois inefficaces et peu fiables, ce sont le plus souvent les erreurs de conception de ceux-ci qui sont la cause principale des problèmes rencontrés (Payre, 2015).

3.1.1. Les ironies d'automatisation

Hoc (2000) a identifié quatre conséquences négatives qui sont souvent influencées les unes par les autres et directement liées à l'introduction des systèmes automatisés :

3.1.1.1. Le mauvais calibrage de la confiance

La première conséquence correspond à un mauvais calibrage de la confiance. Parasuraman and Riley (1997) ont évoqué plusieurs façons inappropriées d'utiliser les systèmes automatisés qui expliquent pourquoi ces derniers ne permettent pas toujours d'apporter les bénéfices qu'ils devraient : leur mauvais usage arrive lorsque l'opérateur fait trop ou pas assez confiance au système automatisé. Les auteurs ont ainsi décrit un grand nombre d'accidents du fait que les opérateurs humains ne suivaient pas les signaux d'avertissement, les considérant souvent comme non fiables. Pour eux, ne pas suivre un système automatisé étant pourtant plus fiable que faire les opérations manuellement, amènent à une désuétude des opérateurs n'utilisant pas le système automatisé

comme ils le devraient (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003). Cette sur-confiance ou sous-confiance est influencée par de nombreux facteurs. Hoff et Bashir (2015) évoquent trois grandes sources de variabilités impliquées dans la confiance envers le système automatisé : la confiance dispositionnelle, la confiance situationnelle et la confiance apprise. La confiance dispositionnelle représente la tendance durable d'un individu à faire confiance au système automatisé. La confiance situationnelle dépend du contexte spécifique bien qu'il est important de noter que l'état mental d'un opérateur peut également altérer la confiance situationnelle. La confiance acquise est basée sur des expériences passées avec un système automatisé spécifique. Elle est étroitement liée à la confiance situationnelle dans la mesure où elle est guidée par l'expérience passée (Marsh & Dibben, 2003). Ces trois niveaux de confiance sont interdépendants.

En ce qui concerne la confiance dispositionnelle, plusieurs facteurs sont à prendre en compte : la culture ; l'âge ; le genre ou encore la personnalité des opérateurs humains.

Concernant la confiance situationnelle, les facteurs à prendre en compte sont : la variabilité externe, c'est-à-dire que la confiance envers le système automatisé dépend du système en lui-même, de sa complexité et de la difficulté de la tâche pour laquelle il est utilisé (Bailey & Scerbo, 2007; Rovira, McGarry, & Parasuraman, 2007; Madhavan, Wiegmann, & Lacson, 2006) ; la variabilité interne, c'est-à-dire les traits liés aux opérateurs humains. Par exemple, Lee et Moray (1994) ont montré que les opérateurs utilisaient les systèmes automatisés lorsque leur confiance en eux ne dépassaient pas la confiance qu'ils avaient en la machine (et vice versa). De même, l'expertise peut également altérer la confiance dans le système automatisé. Certaines études ont montré que les personnes possédant une plus grande expertise sont moins susceptibles de compter sur le système automatisé que les opérateurs novices (Sanchez, Rogers, Fisk, & Rovira, 2014).

En ce qui concerne la confiance apprise, les facteurs à prendre en compte sont les connaissances pré-existantes, c'est-à-dire l'avis que l'opérateur a déjà du système, avis souvent construit par la réputation du système (de Vries & Midden, 2008) mais aussi l'expérience acquise avec le système. Yuliver-gavish et Gopher (2011) ont montré que les participants avaient plus

confiance envers le système automatisé après avoir gagné de l'expérience en l'utilisant ; la transparence du système automatisé est également un facteur important (Seong & Bisantz, 2008). Par exemple, donner des feedbacks sur la fiabilité d'un système automatisé peut permettre d'améliorer la performance car l'opérateur peut plus facilement démontrer une confiance appropriée (Bagheri & Jamieson, 2004). Enfin, la performance du système automatisé et ses résultats jouent également un rôle dans la confiance, notamment selon la fiabilité et la validité des fonctions de ces systèmes (de Visser & Parasuraman, 2011 ; Bailey & Scerbo, 2007).

Cela souligne que la confiance envers le système automatisé est multifactorielle et que beaucoup de facteurs peuvent influencer celle-ci (la fiabilité, l'utilisation, la transparence du système...).

3.1.1.2. La perte d'adaptabilité

La deuxième conséquence est la perte d'adaptabilité qui est caractérisée par un manque de « rétroactions » sur la situation (figure 13).

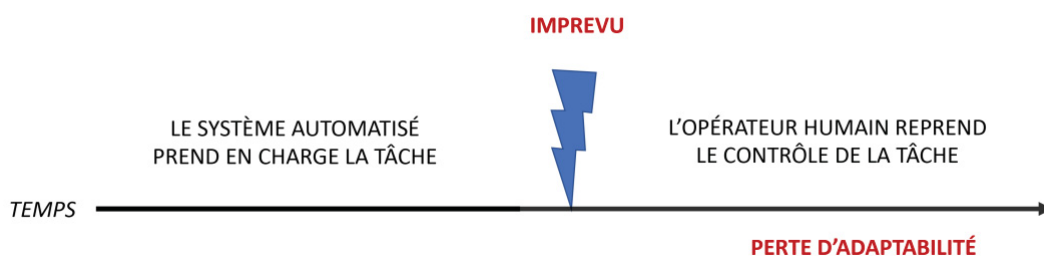


Figure 13. Illustration de la perte d'adaptabilité.

L'opérateur humain peut être amené à reprendre le contrôle du système en cas de défaillances du système automatisé et va se retrouver face à une situation où il n'est pas complètement maître de la situation, n'ayant pas eu la main sur tout ce qui précédait. Endsley et Kaber (1999) évoquent le syndrome de l'humain « hors de la boucle », c'est-à-dire qui n'est plus au fait de l'état de la situation, hors de la boucle de décisions (*loss of situation awareness*). Cette conscience de la situation est définie par Endsley (1996) comme «la perception des éléments de

l'environnement dans un volume de temps et d'espace, la compréhension de leur signification et la projection de leur statut dans un avenir proche » (Endsley, 1996, p. 165). Ce syndrome est présent dans de nombreux secteurs impliquant des systèmes complexes tels que le contrôle aérien (Falkland & Wiggins, 2019) ; le secteur militaire (Riley, Endsley, Bolstad, & Cuevas, 2006) ; la conduite (Skrypchuk, Langdon, Saywer, Mouzakitis, & Clarkson, 2019) ; le nucléaire (Burns et al., 2008). Il est au centre de nombreuses études qui soulignent l'importance de la conception des systèmes automatisés (niveaux et fonctions d'automatisation) dans le maintien de cette conscience de la situation (Kaber & Endsley, 1997 ; Lorenz, Di Nocera, Röttger, & Parasuraman, 2001 ; Jipp & Ackerman, 2016). Cette conscience de la situation est souvent liée à une perte de vigilance de l'opérateur et à l'augmentation du contentement (Endsley & Kiris, 1995 ; Kaber & Endsley, 2004) ou encore l'augmentation de la charge mentale de l'opérateur lors de la transition d'un mode automatique à manuel (Hainley, Duda, Oman, & Natapoff, 2013).

3.1.1.3 Le contentement

La troisième conséquence est le contentement. Le phénomène de contentement s'est vu attribuer plusieurs définitions au cours du temps. Ce terme, d'abord utilisé dans le secteur aéronautique, était défini d'abord comme un état psychologique se caractérisant par un faible taux de suspicion (Wieners, 1981) ; puis comme une autosatisfaction qui peut entraîner une non vigilance fondée sur une hypothèse non justifiée d'état satisfaisant du système (Billings, Lauber, Funkhouser, Lyman & Huff, 1976). Dans cette situation, les opérateurs vont être capables d'adopter les propositions des machines sans les interroger même s'ils sont conscients des limites de ces dernières. La notion de contentement a été étudiée par Layton, Smith et McCoy (1994) dans le secteur aéronautique. Ceux-ci ont repéré que les pilotes se contentaient de suivre la solution proposée par le système automatique, même si la solution n'était pas optimale et bien que le pilote soit assez expérimenté pour s'en rendre compte. Dans une situation de planification de vol de

croisière, le système proposait un plan qui n'était pas optimal. Ce plan était la plupart du temps suivi par les pilotes expérimentés. Ce phénomène a également été étudié dans des études empiriques et a souvent été décrit comme la cause d'une insuffisance de surveillance de l'état du système (Parasuraman & Manzey, 2010).

3.1.1.4. La perte d'expertise

La quatrième conséquence est la perte d'expertise de l'opérateur. Celle-ci correspond à la perte des conditions nécessaires permettant aux opérateurs humains d'exercer leurs responsabilités sur le système (Jones & Mitchell, 1994). Elle est très souvent liée à la perte de conscience de la situation puisqu'avec les systèmes automatisés, les opérateurs humains tendent à devenir passifs et non plus actifs dans leur activité de surveillance (Endsley, 1996). Cela se traduit par une perte de compétences qui se produit généralement en raison d'une exposition prolongée au système automatisé : par exemple lorsque ce système a effectué une tâche pendant une longue période de temps et que l'opérateur humain n'a pas la possibilité de maintenir ses compétences. Selon Jones (2007), une telle perte de compétences peut d'ailleurs augmenter la probabilité de coûts catastrophiques en cas de non-fiabilité du système automatisé, c'est-à-dire en cas de défaillances. C'est pourquoi, les compétences des opérateurs humains doivent être maintenues et prises en compte dans la conception d'un système homme-machine. Ce maintien de compétences est également important pour réduire la perte d'adaptabilité qui peut en résulter.

3.1.2. Les conséquences d'un système automatisé imparfait

La fiabilité des systèmes automatisés est une caractéristique définitivement importante et jouant un rôle déterminant sur la détérioration des performances humaines. Cette détérioration peut s'illustrer par une baisse importante de la surveillance du système automatisé, ayant des conséquences sur la performance et sur l'allocation d'attention des opérateurs humains.

3.1.2.1. Les effets de la fiabilité sur les performances humaines

L'insuffisance de surveillance est souvent évoquée comme étant le résultat d'un contentement des opérateurs humains vis-à-vis des systèmes automatisés (Parasuraman, Molloy, & Singh, 1993 ; Bagheri & Jamieson, 2004 ; Parasuraman & Manzey, 2010). Parasuraman et Manzey (2010) ont d'ailleurs proposé un modèle d'intégration des biais d'automatisation dans l'interaction homme-machine dont la fiabilité du système était un facteur pouvant jouer sur le potentiel de contentement et sur l'allocation de l'attention en plus du niveau d'automatisation (figure 14). Dans ce modèle, les propriétés du système automatisé (telles que la fiabilité ou la stabilité), et les propriétés de l'individu, (telles que les attitudes envers la technologie et les traits de personnalité), influencent toutes deux le potentiel de contentement. Ce potentiel de contentement influe sur l'allocation d'attention des participants et peut engendrer une perte de conscience de la situation de l'opérateur. Une perte de conscience de la situation par les participants en cas de défaillances du système peut engendrer des conséquences sur les performances humaines.

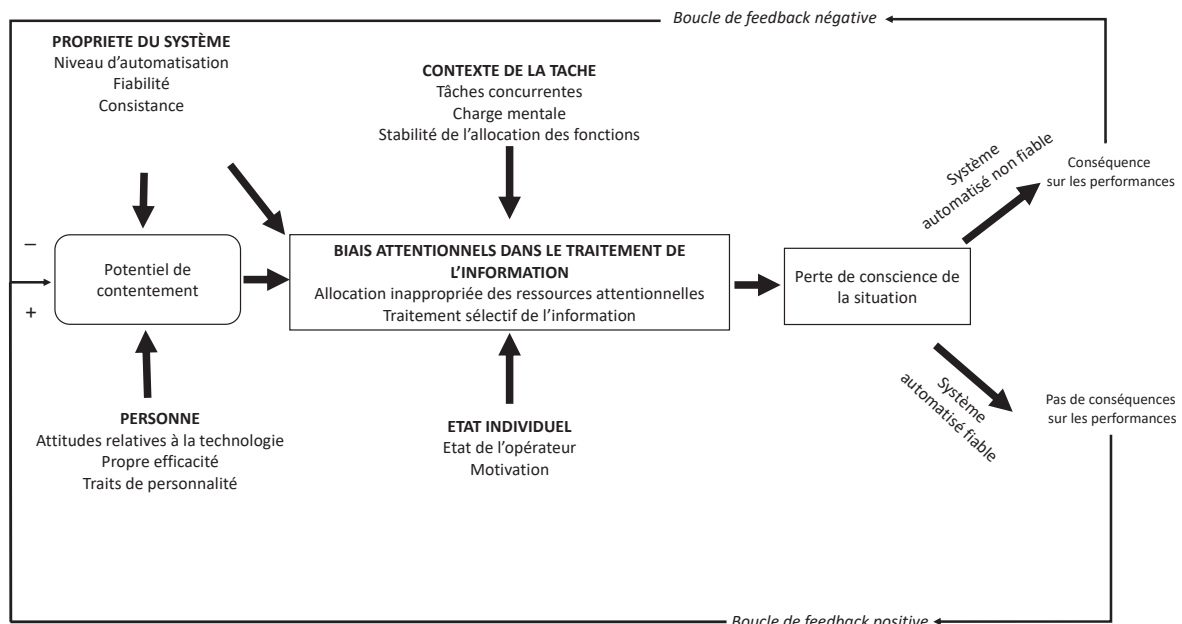


Figure 14. Le modèle d'intégration des biais d'automatisation et du contentement proposé par Parasuraman et Manzey (2010).

Dans les études sur l'insuffisance de surveillance, le point de vue défendu est qu'un système automatisé conduit à une dépendance qui est associée à une surveillance occasionnelle de son efficacité (Mouloua & Hancock, 2019). De fait, un système automatisé non fiable devrait engendrer une surveillance plus accrue de celui-ci. C'est notamment ce que montre l'étude fondamentale de Parasuraman, Molloy et Singh (1993). Dans cette étude, très largement citée dans les études sur les systèmes automatisés imparfaits, les participants accomplissent les tâches simultanées du MATB : la tâche de suivi, de gestion du carburant et la tâche de surveillance (cf. une présentation du MATB est présente dans l'encart du chapitre 2). Dans la tâche de surveillance, ils surveillent quatre jauges et s'assurent qu'aucune ne présente de comportement anormal. Bien que cette tâche soit automatisée, le système automatisé est calibré pour générer des dysfonctionnements, le système est imparfait. Les participants sont donc responsables de la détection et de la correction des dysfonctionnements. Par ailleurs, les auteurs font varier la fiabilité de la tâche de surveillance : elle est soit constante d'une session à l'autre, soit elle varie. Leurs résultats ont souligné qu'avec une fiabilité constante, les participants identifient moins de dysfonctionnements que lorsque la fiabilité varie. Cette étude montre également que la performance des participants est moins bonne lorsque la fiabilité est élevée (87.5%) que lorsqu'elle est moins élevée (56.25%). Ce résultat est considéré comme une insuffisance de surveillance liée à un contentement des participants envers le système automatisé hautement fiable et constant. Quelques années après, Bagheri et Jamieson (2004) ont répliqué l'étude en ajoutant de mesures oculométriques. Ils ont trouvé des résultats similaires et ont notamment montré que les participants allouaient moins d'attention visuelle lorsque la fiabilité était constante que lorsqu'elle était variable. Il est intéressant de noter qu'en plus de l'oculométrie, Bagheri et Jamieson (2004) ont ajouté des mesures de confiance envers le système automatisé. Cependant, leur étude n'a pas permis de montrer de lien entre la détection de comportement anormal du système automatisé et le niveau de confiance. Lee et See (2004) ont confirmé dans leur revue de la littérature qu'il n'y avait pas de lien établi convaincant dans la littérature entre une insuffisance de la surveillance et la confiance envers le système automatisé (Parasuraman & Manzey,

2010) ; Bailey et al. (2006) étant la seule étude à montrer cette tendance. A noter que bien que plusieurs études aient suggéré que les opérateurs humains étaient sensibles à différents niveaux de fiabilité (Parasuraman, Molloy, & Singh, 1993 ; Bagheri & Jamieson, 2004) d'autres ont affirmé l'inverse, interprétation pouvant expliquer le manque de résultats et d'études prouvant le lien entre la confiance et l'insuffisance de surveillance (Dzindolet, Pierce, Beck, Dawe & Anderson, 2001 ; Wiegmann, Rich & Zhang, 2001).

Lorsqu'on s'attarde sur les performances humaines, les différents niveaux de fiabilité semblent donc jouer un rôle important comme souligné dans l'étude exploratoire de Oakley, Mouloua et Hancock (2003). Les performances humaines, mesurées sur un taux de détection des dysfonctionnements, augmentent avec un système présentant une fiabilité entre 5% et 35%. A partir, de 35% un palier, les performances humaines diminuent jusqu'à devenir nulles avec un système fiable à 95%. Singh, Tiwari et Singh (2009) montre que les performances humaines de surveillance du système automatisé sont meilleures avec un faible niveau de fiabilité qu'avec, un système automatisé hautement fiable et dont les dysfonctionnements sont constants. C'est dans la continuité de ces études que Wickens et Dixon (2007) ont proposé un seuil de fiabilité du système automatisé (70%) en dessous duquel son utilité n'a pas de bénéfices par rapport à la réalisation de la tâche manuellement. Ce point de vue vient à nouveau corroborer d'autres points de vue insistant sur l'importance de considérer la fiabilité dans la conception des systèmes d'aides automatisés (Parasuraman & Manzey, 2010 ; Endsley, 2016).

3.1.2.2. L'implication des niveaux et fonctions d'automatisation sur les performances humaines

Dans une méta-analyse, Onnasch, Wickens, Li et Manzey (2014) ont souligné l'implication des fonctions et niveaux d'automatisation sur les performances humaines, notamment en cas de systèmes automatisés imparfaits. Selon ces auteurs, les performances sont améliorées avec de hauts

niveaux d'automatisation si ceux-ci sont fiables. Cependant, en cas de défaillance du système, le coût de reprise en main est plus élevé pour l'opérateur humain si les niveaux d'automatisation sont aussi élevés. En fin de compte, plus la fonction d'automatisation est élevée, plus le coût est élevé en cas de dysfonctionnement du système automatisé (Li, Wickens, Sarter & Sebok, 2014).

Le nombre d'études ayant croisé les fonctions d'automatisation et les niveaux de fiabilité sont limitées. Alors que certaines études montrent des conséquences plus préjudiciables en cas de défaillances du système automatisé prenant en charge les dernières fonctions de décision et d'action (Crocoll et Coury, 1990; Sarter et Schroeder, 2001; Rovira, Cross, Leitch, & Bonaceto, 2014) ; d'autres études soulignent les effets préjudiciables en cas de problèmes de fiabilité avec la prise en charge par le système automatisé des premières fonctions information (Galster, Bolia, & Parasuraman, 2002 ; Rovira, McGarry, and Parasuraman, 2007). Dans une étude récente, Onnasch (2015) a croisé deux niveaux de fiabilité (68.75% et 87.5%) et deux fonctions prises en charge par le système automatisé (acquisition d'informations et implémentation d'action). Bien qu'elle ait trouvé un effet de la fiabilité sur la performance et l'attention des participants, l'auteur n'a pas trouvé d'impacts des fonctions sur les performances humaines. L'interprétation donnée est que dans les études antérieures, les auteurs utilisent la fonction 3 de sélection de la décision alors que dans son étude, Onnasch (2015) utilise la fonction 4 d'implémentation de l'action. Cette dernière fonction change le rôle de l'opérateur humain qui devient superviseur de la tâche entière effectuée par le système automatisé ; alors que dans la fonction 3 de décision, l'opérateur est à moitié impliqué : en tant que superviseur à la fonction 4, l'opérateur humain est responsable de l'ensemble de la tâche. En revanche, la définition du rôle d'un opérateur dans la fonction 3 n'est pas aussi claire. Dans la fonction 3, l'opérateur humain est tenu à l'écart des processus cognitifs d'une tâche, mais il est toujours responsable de la mise en œuvre d'une action qui était proposée par le système automatisé. Par conséquent, un opérateur qui est toujours en charge de la mise en œuvre d'une action peut définir son rôle plus ou moins limité à cette sous-tâche au lieu de se sentir responsable de la tâche entière comme dans la fonction 4. Ainsi, il semble plausible que les opérateurs soutenus

par l'automatisation de la fonction 3 restent principalement concentrés sur l'exécution de ce que le système automatisé recommande. Cela pourrait réduire la probabilité qu'ils détectent un dysfonctionnement du système automatisé et les rendre plus vulnérables aux effets négatifs sur les performances résultant du manque de fiabilité de ce système.

Ces études soulignent la nécessité d'en apprendre plus sur les effets de l'automatisation des niveaux d'informations.

3.2. REPENSER LA CONCEPTION DES SYSTEMES AUTOMATISES POUR REDUIRE LES DIFFICULTES

Afin de garder l'opérateur humain dans la boucle de décisions et de réduire les possibles conséquences négatives, le système automatisé adaptatif s'est développé et a été envisagé comme une solution aux difficultés d'interaction homme-machine. Avec le système automatisé adaptatif, le niveau ou la fonction d'automatisation n'est pas fixe mais flexible et peut donc changer selon les exigences de la situation (Parasuraman, Sheridan, & Wickens, 2000). Cette nouvelle conception du système automatisé a été pensée pour tenir compte des capacités, des forces et limites humaines (Jones, 2007). Selon, Parasuraman et Wickens (2008), ce changement peut être décidé par le système automatisé (adaptatif) ou par l'opérateur humain lui-même (adaptable).

Comme expliqué dans la partie précédente, les opérateurs humains peuvent rencontrer des difficultés à surveiller les comportements d'un système automatisé de manière passive pendant une période étendue (Parasuraman, Molloy, & Singh, 1993 ; Bagheri & Jamieson, 2004). C'est aussi une des raisons qui les poussent à privilégier les tâches manuelles plutôt que laisser le système automatisé effectuer automatiquement les tâches (Navarro & Osuirak, 2015 ; Navarro, Heuveline, Avril, & Cegarra, 2018). Dès lors, trouver la balance équilibrée entre les niveaux et fonctions d'automatisation est un enjeu important pour les facteurs humains. Si, dans de nombreuses situations, un système automatisé infaillible ne peut pas être garanti, il est important de savoir si cette prise en charge par le système automatisé est toujours bénéfique pour l'opérateur. Alors que

certains auteurs ont montré les bénéfices d'un changement de niveau de contrôle sur les performances de surveillance en cas de systèmes automatisés imparfaits (Hilburn, Molloy, Wong, & Parasuraman, 1993 ; Parasuraman, Mouloua, Molloy, & Hilburn, 1993 ; Parasuraman, Mouloua, & Molloy, 1996 ; De visser & Parasuraman, 2011), d'autres auteurs ont alerté sur l'imprédictibilité de ce genre de systèmes pour l'utilisateur (Billings & Woods, 1994).

Parasuraman, Mouloua et Molloy (1996) ont étudié deux types de systèmes automatisés adaptatifs utilisant l'environnement MATB. Ceux-ci ont varié entre les sessions prises en charge par le système automatisé et les sessions manuelles. Les participants ont commencé par une session de 10 minutes au cours de laquelle toutes les tâches MATB ont été effectuées manuellement. Après celle-ci, ils ont effectué trois séances de 30 minutes au cours desquelles la tâche de surveillance a été automatisée (à 56,25% de fiabilité) tandis que les deux autres tâches sont restées sous contrôle manuel. Avec le premier système automatisé adaptatif dit « basé sur modèle », le contrôle revenait toujours au participant (session manuelle) après 40 minutes sous prise en charge du système automatisé. Le deuxième système automatisé adaptatif était basé sur les performances. Les participants ont reçu un contrôle manuel aux moments prévus si la performance était inférieure à un taux spécifique (55%). Les résultats ont montré une augmentation similaire des performances (c'est-à-dire un taux de détection de dysfonctionnement) pour les deux systèmes automatisés adaptatifs).

Ainsi faire varier les niveaux d'automatisation semble être un bon compromis puisque cela permet d'améliorer les performances, notamment en cas de systèmes automatisés imparfaits. Il reste tout de même nécessaire de continuer les études sur les fonctions et les niveaux des systèmes automatisés en cas de systèmes imparfaits. Notamment, puisque la plupart des études que l'on peut trouver dans la littérature croisant ces aspects se focalisent surtout sur des niveaux de fiabilité extrême (très fiable ou pas du tout fiable), mais ne s'attardent pas sur des niveaux plus intermédiaires (Onnasch, 2015).

3.3. CONCLUSION INTERMEDIAIRE

L'introduction des systèmes automatisés dans nos vies à certes eu des impacts positifs mais a également contribué à l'apparition de nouvelles difficultés. Bien concevoir les systèmes automatisés est donc un sujet majeur dans le domaine des facteurs humains. Un système automatisé ne pouvant pas être garanti comme totalement fiable, la prise en considération de cette fiabilité dans la conception des systèmes automatisés est nécessaire pour réduire les impacts potentiels de celle-ci. Étudier ces impacts permet également d'amener des pistes de solution dans la conception comme en témoigne les études sur les systèmes automatisés adaptatifs.

Un intérêt particulier est à porter sur la présence de systèmes automatisés imparfaits avec une prise en charge de différentes fonctions (Parasuraman, Sheridan, & Wickens, 2000), puisque les études croisant ces deux aspects sont limitées et qu'il manque un consensus sur les impacts des problèmes de fiabilité selon les différentes fonctions.

Que ce soit dans l'aéronautique ou dans le transport routier de marchandises, la situation reste la même : un opérateur humain supervise un système automatisé qui peut être imparfait et automatisé selon différentes fonctions. La partie expérimentale de ce manuscrit vise à répondre à la problématique suivante :

Quelles sont les conséquences d'un système automatisé imparfait prenant en charge la fonction information sur le comportement humain ?

Pour y répondre, nous nous intéresserons dans un premier temps à la prise en charge de la fonction information sous forme de signe dans le secteur aéronautique, puis dans un second temps, sous forme d'informations textuelle (symbole) dans le secteur du transport routier de marchandises.

Comme évoqué par Onnasch (2015), évaluer les effets des systèmes automatisés imparfaits, et particulièrement, d'une plus grande variété de fiabilité selon les fonctions pris en charge par ces systèmes est une nécessité pour les prochaines études.

Dans le premier article, les effets d'un système automatisé imparfait seront évalués par le biais de quatre niveaux de fiabilité sur les deux fonctions « information et décision » proposées par le modèle de Parasuraman, Sheridan et Wickens (2000). En effet, un système automatisé peut mettre en place simultanément les deux fonctions et n'est pas limité qu'à la prise en charge d'une seule fonction à la fois. Dans la continuité des travaux cités dans le cadre théorique, la batterie de tâches MATB, (Santiago-Espada, Myer, Latorella, & Comstock, 2011) sera utilisée afin d'évaluer les effets de multiples niveaux de fiabilité sur le comportement humain. Les mesures de performance, les mesures subjectives sur la charge mentale et les mesures d'oculométrie permettront d'évaluer les potentielles conséquences négatives telles que l'insuffisance de surveillance de la tâche automatisée.

Dans le second article, les effets d'un système automatisé imparfait seront évalués par le biais de quatre niveaux de fiabilité. Cette fois-ci, un changement dans les niveaux d'automatisation sera testé afin de mieux comprendre les conséquences d'un système automatisé de type adaptatif

sur le comportement humain. La batterie de tâche MATB sera également utilisée. Des mesures subjectives sur la confiance viendront s'ajouter aux mesures de performance et aux mesures d'oculométrie et permettront d'évaluer les potentielles conséquences négatives telles que l'insuffisance de surveillance de la tâche automatisée, mais aussi les possibles avantages d'un système automatisé adaptatif sur ces mêmes conséquences.

Dans le troisième article, les effets d'un système automatisé imparfait seront étudiés à travers la fonction information prise en charge par le système automatisé dans l'activité spécifique de la planification (cf. le modèle de Parasuraman, Sheridan, & Wickens, 2000). La fiabilité du système automatisé sera investiguée dans des scénarios représentant des tournées avec des informations erronées, de façon à créer un écart avec la représentation que peuvent avoir les planificateurs de tournées valides. L'étude sera construite autour de la volonté de tester l'introduction d'un nouveau prototype de système d'aide à la planification appelé Smart-planning. Cette étude est une des premières à donner une vision d'ergonomie prospective⁶ dans le secteur du transport routier de marchandise, tout en questionnant la notion de fiabilité du système automatisé prenant en charge la fonction information. Elle anticipe les futures transformations de l'activité avec des systèmes d'aide automatisés. Des mesures subjectives sur la qualité des propositions créées par l'aide automatisée seront recueillies afin d'obtenir la position de chaque planificateur face aux différents scénarios.

⁶ L'ergonomie prospective se focalise sur la prospection de nouveaux besoins, la prédiction des conduites futures des usagers et « la conception de futures choses » (Brangier & Robert, 2010).

PARTIE EXPERIMENTALE

ARTICLE 1.

**EFFECT OF IMPERFECT INFORMATION AND ACTION
AUTOMATION ON ATTENTIONAL ALLOCATION**

Avril Eugénie, Benoît Valéry, Jordan Navarro, Liên Wioland, and Julien Cegarra

*Avril Eugénie, Benoît Valéry and Julien Cegarra, Département de Psychologie, Université de
Toulouse, INU Champollion, Albi, France.*

*Jordan Navarro, Département de Psychologie, Université Lumière Lyon 2, Bron, France. Institut
Universitaire de France, Paris.*

Liên Wioland, Département de l'Homme au travail, INRS, Paris, France.

Research article

Effect of imperfect information and action automation on attentional allocation

The study aimed to assess the contributions of imperfect information and action automation on attentional allocation and performance in a multitasking environment. Numerous studies have stressed the attentional and behavioral effects of imperfect automation as a function of automation stages. When information and action automation are reliable, both can support performance. However, with unreliable aids, the literature suggests that action automation tends to be detrimental. Participants (n=96) completed three Multi-Attribute Task Battery (MATB) tasks. The monitoring task was automated with two types of automation (action or information) of four reliabilities each (0%; 56.25%; 87.5%; 100%). Ocular behaviors and performance were assessed. The reliability of information automation influenced visual resource allocation. When information automation was the most reliable, participants spent the least amount of time sampling the monitoring task. Finally, the reliability of action automation triggered no effect on performance or cumulative dwell times. Our results suggest that in complex multitasking situations where information and action automation occur simultaneously, participants allocated fewer visual resources to automated task with increased automation reliability, but only with information automation. Similarly, their performance was better only with increased information automation reliability.

Keywords: automation, reliability, multitasking, attentional resources

Introduction

Automated systems are being developed rapidly and have become widespread in professional settings, such as manufacturing scheduling (e.g., Cegarra & Hoc, 2008), pipeline supervision (e.g., Wesley & Dau, 2017), car driving (e.g., Navarro, 2019; Navarro et al., 2016; Navarro et al., 2019; Stanton et al., 2001), space missions (e.g., Wickens et al., 2015), health care (e.g., Asan & Carayon, 2017; Zybak et al., 2016), or drilling tasks (e.g., Gressgard et al., 2013). Although automation has contributed to the security of modern socio-technical systems, some authors have expressed concern over the possible paradoxes inherent in human-automation interaction, such as task disengagement, reduced system awareness, mistrust, complacency, or cognitive skill reduction (Bainbridge, 1983; Hoc, 2000; Parasuraman & Riley, 1997; Sarter et al., 1997; Wiener & Curry, 1980).

In their seminal study, Parasuraman et al. (1993) investigated the role of automation reliability in the presence of complacency. Participants completed the Multi-Attribute Task Battery in a multitasking environment (MATB, designed by Comstock & Arnegard, 1992). They had to perform three tasks, tracking, fuel management, and monitoring, simultaneously. In the monitoring task, they had to monitor four gauges and make sure that none showed any abnormal behavior. Although this task was automated, the automation was imperfect, so participants were responsible for detecting and correcting malfunctions. Moreover, the reliability of the monitoring task could be constant or vary from one session to the next. Their results had stressed that when the reliability was constant, participants identified less abnormal behavior than when reliability varied.

However, as Sarter and Shroeder (2001) noted, it is surprising that in Parasuraman et al.'s (1993) study, researchers have been more interested in the effect of automation reliability in the later stages of action and implementation (Bagheri & Jamieson, 2004a; Manzey et al., 2009; Singh et al., 1993; Singh et al., 2000). Various automation stages have been considered as possible candidates for levels of automation. These stages include (1) information acquisition, (2)

information analysis, (3) action selection, and (4) action implementation (Parasuraman et al., 2000). Moreover, some studies have reported that each stage is likely to drive attentional allocation strategies (Bagheri & Jamieson, 2004b; Rovira et al., 2002; Sarter & Schroeder, 2001). Indeed, operators tended to reallocate their attentional resources when interacting with an imperfect automation aid in any of these stages. In this way, and based on the analysis of 20 studies, Wickens and Dixon (2007) proposed that at a reliability threshold of 70%, automation was probably inferior to a manual mode.

Few researchers have compared the consequences of information and action automation aids on performance. For instance, Sarter and Schroeder (2001) compared the effects of the reliability of two types of automation on inflight icing. The first one informed about the location of ice accretion (i.e., information analysis) while the second one recommended action (e.g., power settings or pitch attitude) according to the icing conditions (i.e., action selection). When providing reliable information, both automation types improved performance. However, an unreliable action recommendation was more detrimental to overall performance on the location of ice accretion than was unreliable information (see also Rovira et al., 2002). In general, such a result has been interpreted as evidence that reliability issues are more detrimental to situational awareness when they concern later automation stages (Rovira et al., 2014)

However, other empirical evidence about reliability is not fully compatible with such a late-stage interpretation. Rovira et al. (2007) compared the consequences of reliabilities of information automation and three levels of action automation in a simulated military task. At below 80% reliabilities, imperfect action automation worsened the performance compared to imperfect information automation. According to these authors, operators relying on information automation value many possible paths of action and implement alternative ones. This is not possible when action automation itself imposes the values of the different possibilities of action. However, when reliability is only around 60%, participants' performance decreased with both information and action automation.

Moreover, Lorenz et al. (2002) noted that reliance on the highest automation stage does not necessarily lead to poorer performance. They tested its interaction with an automated aid supervising life-support system in space missions. The highest automation stage focused on diagnosis and recovery; the earliest stage detected faults. They noted that the highest stage, though unreliable, did not have the most detrimental effect on performance because it freed up more resources so that operations could be more engaged in better fault comprehension.

Moreover, it should be noted that in most studies, the authors did not assess the combination of several types of automation directly but only compared each type of automation and their consequences in terms of performance and/or visual resources allocation separately (e.g., Rovira et al., 2014; Sarter & Schroeder, 2001). However, evaluating the reliability effects of two simultaneous automation stages on the same task is the only way to understand the arbitration conducted by the operators when facing multiple automation types, each with different levels of reliability. In other words, such manipulation would allow more accurate identification of the role of the automation stage in the operator's behavior.

Therefore, we empirically studied the differential effects of the automation stages (information vs. action) using four possible values of reliability on the attentional allocation policy and performance of the operators. The study utilized an open-source version of the Multi-Attribute Task Battery (OpenMATB; Cegarra et al., 2019), modulating four levels of reliability, 0%, 56.25%, 87.5%, 100%, of two simultaneous automation in the monitoring task. These two automation types consisted of information automation (warning system) and action automation (automated solving), i.e., early (information acquisition) and late (action implementation) automation stages. Beyond performance measures, we measured the attentional allocation of the participants by measuring their gaze behavior with a non-invasive eye-tracking device.

According to Wickens and Dixon (2007), imperfect automation with reliability greater than 70% is beneficial for human performance. These authors specified that this statement is relative only to stage 1 and 2 (information). However, concerning other stages, as Bailey and Scerbo (2007)

explained, operator trust increased and monitoring performance decreased as a function of increased system reliability. Moray and Inagaki (2000) confirmed this finding and added that reliable variables tend to be sampled less and less frequently. Accordingly, we tested the following hypotheses:

Hypothesis 1a: Increasing the reliability of information automation decreases participants' attention allocated to the automated task.

Hypothesis 1b: Increasing the reliability of action automation decreases participants' attention allocated to the automated task.

Hypothesis 2a: Increasing the reliability of information automation increases the performance of participants when automation fails.

Hypothesis 2b: Increasing the reliability of action automation decreases the performance of participants when automation fails.

Method

Participants

Ninety-six participants (69 women and 27 men) aged 18 to 27 years ($M = 20.25$) participated in the experiment. All participants were students at the University of Albi. All had a normal or corrected visual acuity and no prior experience with the paradigm. Before the experiment, all participants read and signed a consent form. There were no rewards or compensation for their participation. We certify that we complied with the APA ethical principles regarding research with human participants in the conduct of the research presented in this manuscript (APA, 2017).

Apparatus

The experiment took place in a quiet and isolated experimental room. All the tasks were run on a computer attached to a 17" monitor (1440×900 pixels). Participants sat approximately 70 cm from the screen. Throughout the experiment, eye gazes were recorded at 250 Hz using a

RED 500 system (SensoMotoric Instruments). Responses were recorded utilizing a traditional keyboard.

An open version of the revised Multi-Attribute Task Battery (MATB-II, Santiago-Espada et al., 2011), the OpenMATB (Cegarra et al., 2019), was used. The MATB task was used because one of its advantages is that the nature of the task allows for the combination of two automation types. The MATB is also often used to assess the consequences of imperfect automation. The simultaneous tasks allow us to better understand different visual allocations with multiple tasks. The MATB is a flight environment simulation where monitoring, tracking, resources management, communication, and scheduling must be handled simultaneously (Fig. 1). This study focused only on monitoring, tracking, and resources management tasks.

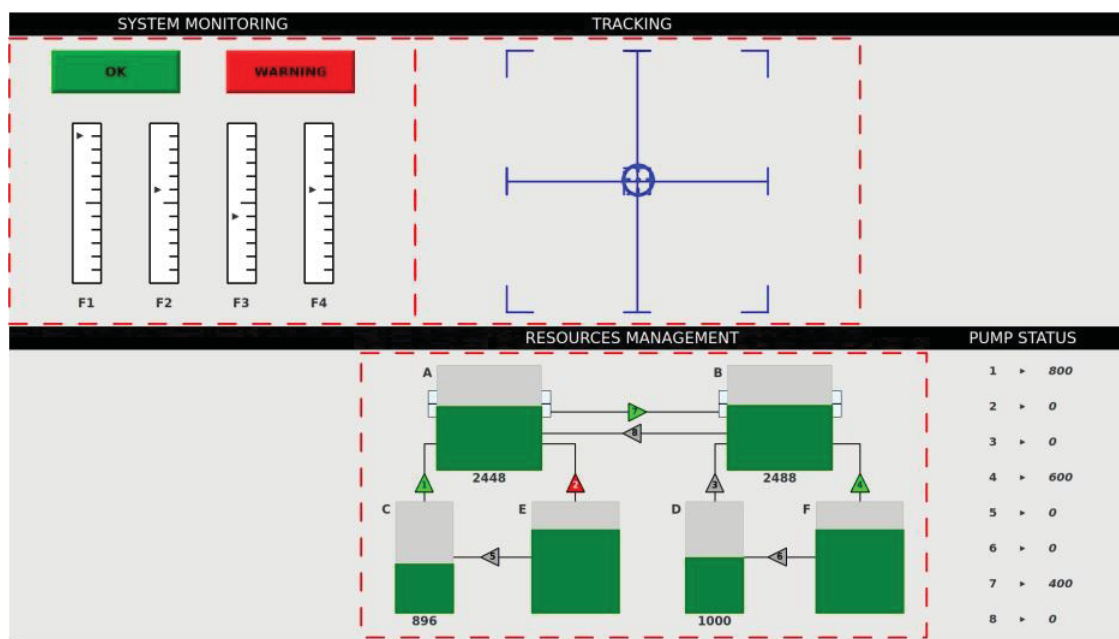


Figure 1. Screen capture of the MATB environment. The system monitoring task is at the top left, the tracking task is at the top center, the resources management is at the bottom center, and the pump flow values are displayed at the bottom right. Red dashed rectangles represent areas of interest used for visual attention analyses for each task and are only displayed here for illustrative purposes. In this example, the system monitoring task shows a classical failure event, where the cursor (triangle) of the first gauge deviates from the center, and the “Warning” button is red. Note that pump flows are only displayed for activated pumps (in this example, pumps 1, 7, and 4).

The tracking task

This task replicates continuous compensatory actions of the aircraft. Participants must use a joystick to compensate for cursor deviations, keeping it in a central rectangular area. The cursor movements follow a forcing function corresponding to a sum of sinusoidal waves for the X and Y axes with a cut-off frequency of 0.06 hertz. As a result, the cursor path is comparable to a Lissajous curve, appearing at random but strictly identical for all participants, regardless of the compensations.

The resources management task

This task is a diagram representation of an aircraft fuel system. It comprises six rectangular tanks, two of which (A and B) are depleting at a constant rate (800/min). Between the tanks, eight numbered pumps allow transferring fuel from one tank to another at a fixed rate. These pumps can be turned on and off using the corresponding numpad key (1 to 8). Participants must maintain tanks A and B as close as possible to their initial level (2500) by turning the eight available pumps on and off using the corresponding numpad key (from 1 to 8).

Pump flows are displayed on the right of the task and only appear for activated pumps. When a tank is full (or empty), all incoming (or outgoing) pumps are automatically disabled. Pumps are color-coded for their status: grey means it is disabled; green indicates an active transfer; red means the pump is failing and cannot be activated (for 30 seconds). The tank and pump status is updated every 2 seconds. Flow and tank values used in the present study are detailed in Table 1.

Table 1. Default tank and pump properties used in the present study.

Pump	Flow rate	Tanks	Tank Leak rate	Leak rate	Initial level
1	800	C → A	A	800	2500
2	600	E → A	B	800	2500
3	800	D → B	C	0	1000
4	600	F → B	D	0	1000

5	600	$E \rightarrow C$	E	0	NA
6	600	$F \rightarrow D$	F	0	NA
7	400	$A \rightarrow B$			
8	400	$B \rightarrow A$			

Note. NA = Not available because these tanks were unlimited.

The system monitoring task

For this task, participants had to watch four graduated gauges, labeled “F1,” “F2,” “F3,” and “F4,” whose cursor shifted around the center. If a cursor moves outside the central zone, participants must hit the corresponding response key (F1, F2, F3, and F4). Two kinds of visual feedback were then displayed: (1) the cursor stopped at the center of the scale, and (2) a yellow rectangle appeared at the bottom of the gauge for 1.5 seconds. Participants had only 10 seconds to detect any abnormality, after which the situation returned to normal. Missing an abnormality was considered a human error.

The system monitoring automation

In this study, the monitoring task was automated during the test session. Two automation were present, which handled the automatic resolution of a failure and/or its signaling. These two systems had varied reliabilities (0, 56.25, 87.5, or 100%). When the information automation was functioning, a gauge failure resulted in the “Warning” button turning red until the failure was resolved. When information automation was not functioning, failure events were simply not signaled (the “Warning” button did not light up). When the action automation was functioning, an incoming gauge failure was automatically resolved after a 1.5-second delay. The positive feedback described above was then provided. These two dimensions were independent so that throughout the session, the “Warning” signal (information) was not contingent on the action automation. This independence was achieved through appropriate event contingencies, as illustrated in Table 2.

Table 2. An example of automation contingencies for 87.5% information reliability and 56.25% action reliability.

		Information		
		Yes	No	Margin (%)
Action	Yes	8	1	56.25
	No	6	1	43.75
	Margin (%)	87.5	12.5	100

Procedure

After the participants had read and signed informal consent, the MATB tasks were explained to them. To ensure their understanding, they had to describe the three tasks to the experimenter. If necessary, the tasks were explained again. At that point, automation was not mentioned. The participants sat in front of the computer before the eye-tracker calibration was performed. A 10-minute training session started, during which they had to perform the three tasks simultaneously. After this phase, participants were informed that during the test phase, the system monitoring task would be automated either through information or action. The two automation systems were presented to the participant, as described in the Apparatus section. Notably, participants were not informed about the automation reliability; they were told only that these two automation systems would not be completely reliable. They had to compensate for an automation dysfunction manually. Subsequently, they completed four consecutive 10-minute experimental sessions, separated from each other by a pause of 5 minutes. Throughout the procedure, the three tasks were presented as equally important. Participants were instructed to use their right hand to perform the tracking task (joystick) and the resources management task (numeric pad on the keyboard) and their left hand to perform the monitoring task (F1 to F4 keys).

In each 10-minute session (training and test), the system monitoring task underwent 16 gauge failures, separated from each other by delaying the values of which were randomly chosen, ranging from 13 to 72 seconds. In the meantime, the resources management task had six pump failures, separated by a random delay of 10 to 120 seconds. A complete session lasted about 70 minutes.

Measures and design

Measures

Performance. For each session, performance on the tracking task was assessed by the amount of time that the cursor spent inside the target area. Regarding the resources management task, the mean distance to the target was computed for the two depleting tanks (A and B), and the two distances were averaged. For the monitoring task, the detection rate (i.e., the proportion of corrected malfunctions) and mean reaction time (i.e., mean delay between the malfunction onset and its potential correction) were computed. The participant detection rate corresponded to the detection when the failure was not corrected by automation. All measures were normalized and reversed when relevant (i.e., for the resources management task), so that 0 expressed the worst and 1 represented the best possible performance, except for the reaction time.

Visual resources allocation. Concerning ocular behavior, three areas of interest (AOI) were defined, corresponding to the virtual rectangles surrounding the three MATB tasks on the screen (Figure 1). For each AOI, the cumulative dwell time was computed, corresponding to the sum of the recorded fixation durations. One fixation was characterized as a set of gaze samples separated from each other by up to 25 pixels, representing a time window of at least 50 milliseconds.

Design

A full between-subject was not achievable because it would require 384 participants (24 participants x 16 conditions). Similarly, if information automation reliability were also arranged as within-subject factors, the total experiment time would last 40*4 minutes, which was too long.

A 4 x 4 analysis of variance (ANOVA) was used to analyze the data, with the action reliability (AR) being a within-subject factor (0, 56.25, 87.5, 100) and the information reliability (IR) being a between-subject factor (0, 56.25, 87.5, 100). We tested the same intermediate levels of reliability used by Parasuraman et al. (1993) and Bagheri and Jamieson (2004a, 2004b), which showed effects on the attentional behavior of the participants. These levels were determined based on the number of events in a 10-minutes session and the support provided by the automation: a 56.25% reliability level of automation corresponded to nine out of sixteen events successfully detected by automation; an 87.5% reliability level of automation corresponded to fourteen out of sixteen events successfully detected by automation; 0 and 100 corresponded to none or sixteen out of sixteen events successfully detected by automation. We added two levels of reliability (0 and 100) to expand these levels of reliability. This allowed us to compare one perfect level of reliability and multiple levels of imperfect reliability, as was done in other studies with imperfect automation (Chavaillaz et al., 2016; Moray & Inagaki, 2000).

As a consequence, 24 participants were randomly assigned to each IR level. The order of presentation of the four AR levels was counterbalanced across participants. For each 10-minute session, trials (like in Table 2) were ordered randomly. When no interaction effect was triggered, the analysis was completed with two one-way ANOVAs. Post-hoc multiple pairwise comparisons were always corrected with the Bonferroni method. Given a large number of statistical tests, the significance level was set at 0.01. All p-values and degrees of freedom were always corrected for non-sphericity with the Greenhouse-Geisser method.

Results

Analysis of the automated task

Cumulative dwell time

The 4 x 4 ANOVA revealed no interaction effect of IR and AR levels over the cumulative dwell time on the monitoring task, $F(3,9) = 0.56, p = .83$. Additionally, cumulative dwell time was not influenced by the level of AR, $F(3,281) = 0.78, p = .50$, in top right of Figure 2. Nonetheless,

we obtained a main effect of the IR level over the cumulative dwell time on the monitoring task, $F(3,92) = 5.45, p < .01, \eta_p^2 = .15$. More precisely, post-hoc indicated a significant difference between the level 0% and 100% ($p < .01$), with 0% of IR triggering much more cumulative dwell time ($M = 70.07, SD = 42.41$) than 100% of IR ($M = 39.63, SD = 33.42$; Figure 2, top left).

Performance

For detection rate and reaction time measures, conditions with 100% of AR were excluded because they did not trigger any malfunction to correct manually.

Detection rate. There was no interaction effect of IR and AR levels over the detection rate at the monitoring task, $F(2,6) = 0.76, p = .59$. We did not find any main effect of the AR level, $F(2,175) = 1.57, p = .21$. However, as for the cumulative dwell time, there was a main effect of the IR level over the detection rate, $F(3,92) = 9.90, p < .001, \eta_p^2 = .24$. More precisely, post-hoc showed that 0% of IR led to a drastic decrease in detection rate (0% with $M=0.65, SD=0.32$), when compared to the three other levels (56.25% with $M=0.81, SD=0.16$; 87.5 with $M=0.89; SD=0.12$ and 100 with $M=0.94; SD=0.12$; fig. 3, middle left for IR and right for AR).

Reaction time. In terms of the main and interaction effects on the detection rate, an interaction effect of IR and AR levels, $F(2,4) = 0.64, p = .69$, and the main effect of AR, $F(2,161) = 4.63, p = .013$, over the mean reaction time in the monitoring task were non-significant. However, the main effect of the IR level over the reaction time was significant, $F(3,90) = 13.0, p < .001, \eta_p^2 = .30$. Post-hoc tests showed that this effect was due to longer reaction times in the condition with 0% level of IR ($M=4.76s, SD=1.2$), as compared with the other three IR levels (56.25% with $M=3.37, SD=1.04$; 87.5% with $M=3.23, SD=0.89$; and 100% with $M=3.11, SD=0.87$; the bottom of Fig. 3).

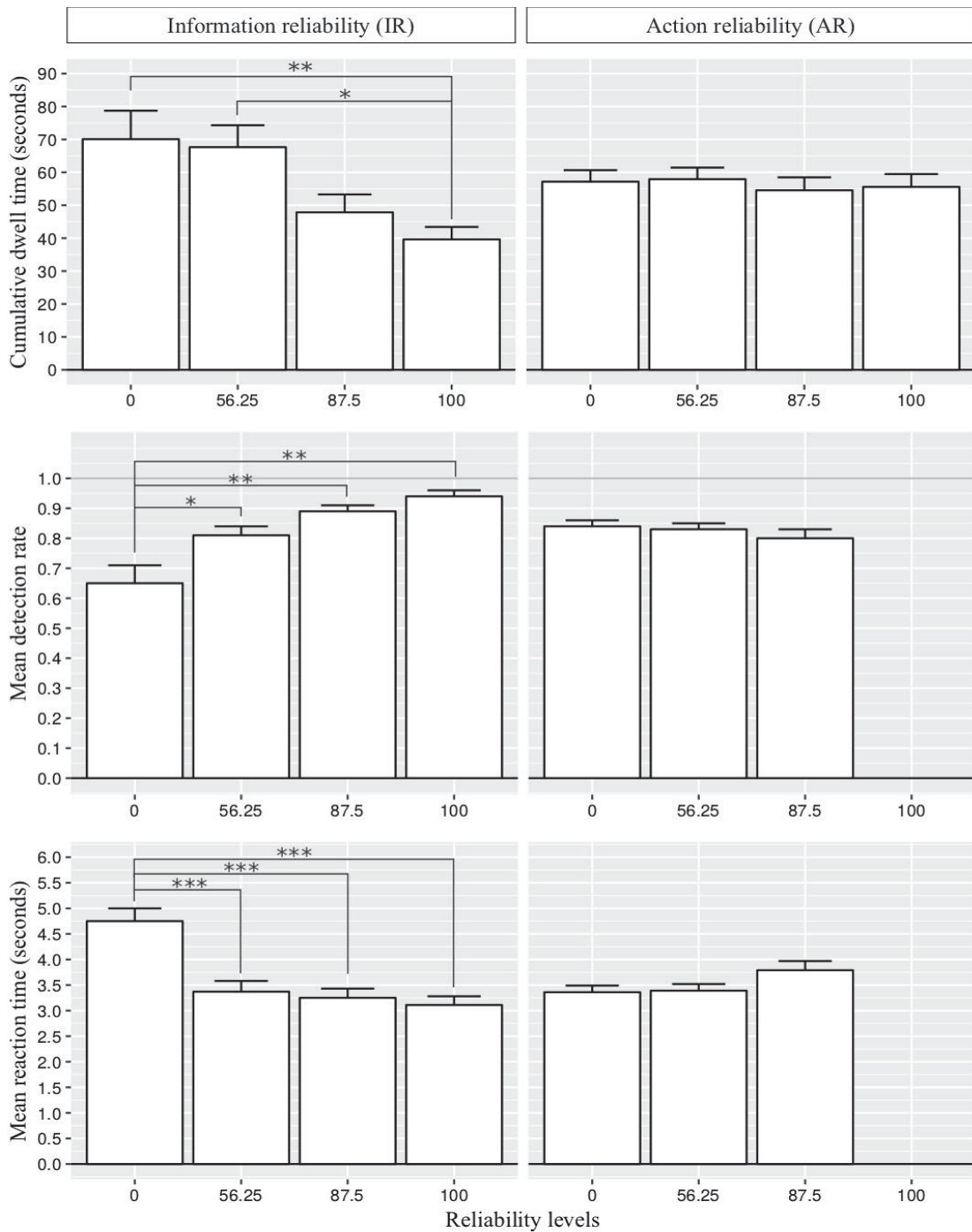


Figure 2. Mean cumulative dwell time (top), detection rate (middle), and reaction time (bottom) on the monitoring task as a function of the information reliability (IR) levels (left) and the action reliability (AR) levels (right). Only the IR levels triggered significant effects. Error bars represent 95% confidence intervals (* < .05, ** < .01, *** < .001). Error bars are standard error.

Considering these preliminary performance results, one might question whether the decrease in performance in the 0% IR conditions (Fig. 3) could be attributable to the evolution of the

signal/noise ratio across the various conditions. The higher the IR level, the higher the proportion of signaled failures. In this case, the performance output would not depend on the IR level but rather on the ease of detecting a failure, assuming this latter was signaled.

Effect of the signal presence on performance. We conducted multiple student's t-test comparisons of the detection rate and the reaction time across the various IR conditions, distinguishing between whether failures were signaled or not (Table 3).

Table 3. Mean detection rates and reaction times as a function of information reliability (IR) levels, distinguishing cases where failures were signaled or not. † In 0% IR conditions, failure was not signaled. In 100% IR conditions, failures were always signaled.

		Information reliability (IR) levels				
Dependent variable	Signaled failure	0	56.25	87.5	100	All
Detection rate	Yes			0,93 ±	0,95 ±	0,95 ±
		†	0,97 ± 0,07	0,10	0,10	0,09
	No	0,65 ±		0,61 ±		0,63 ±
		0,32	0,63 ± 0,31	0,35	†	0,32
Reaction time (seconds)	Yes			3,11 ±	3,11 ±	3,04 ±
		†	2,90 ± 1,15	0,92	0,88	0,98
	No	4,75 ±		4,49 ±		4,66 ±
		1,19	4,54 ± 1,32	1,88	†	1,48

Detection rate. Regardless of the IR level, participants manually corrected more failures when these failures were signaled ($M = 0.95$, $SD = 0.09$) than when there were not ($M = 0.63$, $SD = 0.32$). This difference was significant, $t(142) = -8.03$, $p < .001$. Moreover, multiple pairwise comparisons led to the conclusion that the global proportion of signaled failures (IR level) did not influence the probability to detect a failure when signaled ($p \geq .199$, for 3 pairwise comparisons) and not signaled ($p \geq .702$, for three pairwise comparisons; Table 3).

Reaction time. Regardless of IR level, reaction times were significantly shorter when failures were signaled ($M = 3.04$, $SD = 0.98$) than when they were not ($M = 4.66$, $SD = 1.48$), $t(138) = 7.68$, p

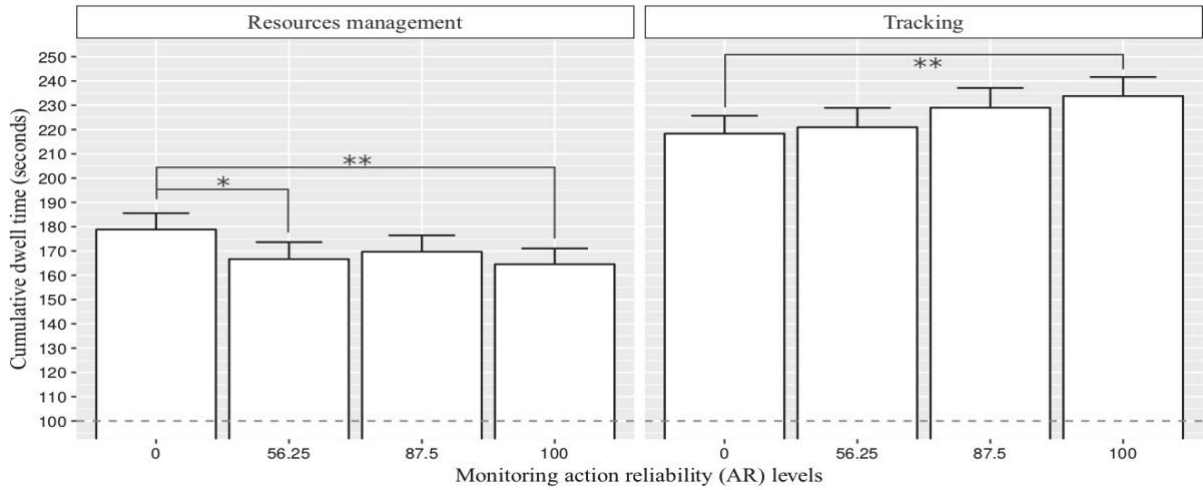
< .001. Multiple pairwise comparisons showed that the IR level did not influence the reaction time when the failure was signaled ($p \geq .476$, for 3 pairwise comparisons) and was not signaled ($p \geq .590$, for 3 pairwise comparisons; Table 3).

Analysis of the other tasks (tracking and resources management)

Cumulative dwell time

No interaction effect of IR and AR levels was found for both the tracking, $F(3,8) = 0.81, p = .61$, and the resources management tasks, $F(3,8) = 0.42, p = .42$. IR levels triggered no main effect on the cumulative dwell time for both the tracking, $F(3,92) = 1.37, p = .27$, and the resource management tasks, $F(3,92) = 0.26, p = .86$. Conversely, AR levels had a significant effect on the cumulative dwell time for the tracking, $F(3,274) = 3.963, p < .01, \eta^2 = 0.04$ (0% with $M=218.33s, SD=72.13$; 56.25% with $M=220.94s, SD=78.49$; 87.5 with $M=229s, SD=79.69$; and 100% with $M=233.79s, SD=77.10$). Post-hoc showed a significant difference between the AR 0% level and 100% level, $p < .01$. AR levels had also a significant effect on dwell time for the resource management tasks, $F(3,262) = 4.08, p < .01, \eta^2 = 0.04$ (0% with $M=178.84s, SD=66.03$; 56.25% with $M=166.64s, SD=68.60$; 87.5 with $M=169.7s, SD=66.25$; and 100% with $M=164.54, SD=63.45$; Figure 3). Post-hoc showed a difference between the AR 0% level and the 56.25% level ($p < .05$) and between the 0% level and 100% level ($p < .01$).

Figure 3. Mean cumulative dwell time for the tracking (right) and the resource management (left) tasks as a function of the action reliability (AR) level. Error bars represent 95% confidence intervals (* < .05, ** < .01, *** < .001). Error bars are standard error.



Performance

The normalized performance indices were not affected by the main or the interaction effect of the IR and AR levels. Exhaustive inferential and descriptive results are given in Table 4.

Table 4. Descriptive (top) and inferential (bottom) results for the performance indices at both the tracking and the resource management tasks.

Task	Reliability	0	56.25	87.5	100
Tracking	Information	0,38 ± 0,10	0,39 ± 0,12	0,38 ± 0,12	0,4 ± 0,14
	Action	0,37 ± 0,12	0,38 ± 0,12	0,39 ± 0,13	0,39 ± 0,13
Resources management	Information	0,9 ± 0,15	0,9 ± 0,11	0,91 ± 0,09	0,9 ± 0,13
	Action	0,9 ± 0,14	0,91 ± 0,11	0,89 ± 0,14	0,9 ± 0,13

Task	Effect	F(df)	P
Tracking	Action reliability (AR)	(3,264) = 3,46	0,02
	Information reliability (IR)	(3,92) = 0,092	0,97
	AR:IR	(3,8) = 0,70	0,7
Resources management	AR	(2,224) = 1,73	0,17
	IR	(3,92) = 0,089	0,97
	AR:IR	(2,7) = 0,74	0,64

Discussion

The purpose of this study was to evaluate the role of the two main automation stages located at both ends of the automation stages (i.e., information versus action automation) and their reliability in the performance and the resource allocation policy of the operators in a high-demanding multitasking situation. We believed that an increase in information automation reliability would decrease the cumulative dwell time allocated to the monitoring task (Hypothesis 1a).

Our results confirmed this hypothesis. More reliable information automation was associated with fewer visual resources allocated to the automated task. Participants spent less time sampling the monitoring task under high levels of IR than under low levels in that 0% of IR triggered much more cumulative dwell time compared to 100% of IR. Concerning the performance, we supposed that increasing the reliability of information automation would increase the performance of participants when automation fails (Hypothesis 2a). Our results confirmed this hypothesis. Participants had a lower detection at 0% information automation compared to other levels. Hence, under-sampling was associated with better detection. From the point of view of the visual resources, this corroborates previous evidence that diminution in attentional allocation is a major consequence of reliable automation in a multitasking situation (Molloy & Parasuraman, 1996). Performance results should be discussed concerning the notion of complacency. In most studies, the reduction of attentional allocation led to poorer performance. Moray and Inagaki (2000) warned against an interpretation of complacent or skeptical behavior towards automation simply based on performance changes. They stressed the importance of determining an optimal sampling behavior and attributed complacency to under-sampling. In our experiment, the improvements in performance were attributable only to better signal visibility and not to a behavioral change related to the global IR level. When we distinguished whether failures were signaled or not across the various IR conditions, participants manually corrected more failures

when failures were signaled ($M=0.95$, $SD=0.09$) than when they were not ($M=0.63$, $SD=0.32$). That is to say, participants showed a strong automation bias (Mosier & Skitka, 1996; Parasuraman & Manzey, 2010) towards information automation, regardless of its reliability.

We hypothesized that an increase in action automation reliability would decrease the cumulative dwell time allocated to the monitoring task (Hypothesis 1b). This pattern was, however, not observed across the various AR levels, $F(3,281)=0.78$, $p=.50$. Instead, the cumulative dwell time was similar for all levels (0% with $M=57.16s$, $SD=34.55$; 56.25% with $M=57.93s$, $SD=34.55$; 87.5, $M=54.55s$, $SD=38.56$, and 100% with $M=55.57s$, $SD=38.36$). This hypothesis was rejected. We obtained equivalent results in terms of performance (Hypothesis 2b). We did not find any main effect of the AR level, $F(2,175)=1.57$, $p=.21$. Indeed, the detection rate was similar for all levels (0% with $M=0.84$, $SD=0.23$; 56.25% with $M=0.83$, $SD=0.24$ and 87.5,% with $M=0.80$, $SD=0.31$). We hypothesized that increasing the reliability of action automation decreases the performance of participants when automation fails. This hypothesis was rejected.

Although no interaction effect was found between action and information automation on performance or cumulative dwell time, the result showing that IR level but not action automation (AR) influenced visual resource allocation raises questions about the influences of the automation stage on the attention resources allocation and the associated performance.

The vast majority of studies have observed the influence of the automation stage on the operator's ability to detect automatism malfunctions, suggesting that late-stage automation appears to be more detrimental to performance compared to earlier-stage automation (Crocoll & Coury, 1990; Sarter & Shroeder, 2001). Onnasch et al. (2014) reviewed a set of 18 studies and concluded that the most negative consequences of imperfect automation seem to emerge when switching from stage 2 (information analysis) to stage 3 (action selection) of automation. Accordingly, authors have recommended the use of “medium” levels of automation (Comings & Mitchell, 2007), which are focused on information automation and not on the action itself (see Onnasch, 2015). One study showed that when operators utilize several automation stages in support of a task, they

tend to rely more on these “medium” stages (Navarro et al., 2018). We found that information automation reliability (IR) mainly drives the participants’ attentional behavior, which implies that IR should influence their behavior. We demonstrated that when facing two types (information vs. action) of potentially unreliable automation, operators tend to manage their attentional behavior as a function of the earliest stage of automation. More precisely, we showed that the reliability of information automation but not action automation determined their compliance/reliance behaviors (Meyer, 2004). This strengthens the importance of studying the effects of automation reliability and its associated effects on attentional resource allocation.

Limitations

Further studies should consider measuring participants’ perceptions of reliability based on subjective evaluations.

First, in this study, participants learned to judge automation reliability during 10-minute trials, but their judgments might change across time. Therefore, it would be interesting to evaluate how they perceive the reliability over time. This was not an objective in this study, but future experiments with much longer trials would be highly valuable. Indeed, adopting complacent behavior towards automation takes time. For example, in Parasuraman, Molloy, and Singh’s (1993) experiment, complacent behavior was notable only after 20 minutes of exposure to automation.

Second, it would be worthwhile to consider the role of the training phase. In this study, we favored a training with an IR level of 100% (i.e., all failures were signaled) and an AR level of 0% (i.e., all the failures had to be corrected manually). Replicating the study with, for instance, a null-reliability training session for the two automation types would also clarify the specific role of training in attentional allocation.

Practical implications

Our results have strong consequences for future studies on automation. Indeed, information is a crucial aspect of automation supervision, and there are multiple types of information. Even the correct human monitoring of late-stage automation (e.g., action

implementation) ultimately relies on its ability to provide accurate signals (feedback). In other words, studying human-automation interaction in the context of potential unreliability issues involves assessing various types of provided information (e.g., early signals, feedback). For instance, some studies have suggested that providing operators with context-related information about automation reliability could help them adapt their monitoring behavior (Albayram et al., 2019; Bagheri & Jamieson, 2004b; Moray & Inagaki, 2000; Wickens & Dixon, 2007). In this way, Bagheri and Jamieson (2004b) replicated Parasuraman et al.'s (1993) study by adding a condition in which participants were more explicitly informed about the ongoing reliability (e.g., "automation reliability will be slightly more than 50%"). The latter condition led to better information sampling behavior and a higher failure detection rate. This should draw designers' attention to the issue of information dominance, especially considering that multiple stages of automation could involve the same task and compete for the operator's limited resources.

Acknowledgments

This research was supported by grants from ANR (« Smart Planning » – ANR-16-CE26-0017).

References

- American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct* (2002, amended effective June 1, 2010, and January 1, 2017). <https://www.apa.org/ethics/code/index.aspx>
- Bagheri, N., & Jamieson, G. A. (2004a). Considering subjective trust and monitoring behavior in assessing automation-induced “complacency.” *Human performance, situation awareness, and automation: Current research and trends*, 54–59.
- Bagheri, N., & Jamieson, G. A. (2004b). The impact of context-related reliability on automation failure detection and scanning behaviour. *IEEE Transactions on Systems, Man and Cybernetics*, 1, 212–217.

- Bainbridge, L. (1983). Ironies of automation. In *Analysis, design and evaluation of man-machine systems* (Vol. 19, pp. 129–135). London: Elsevier.
- Cegarra, J., & Hoc, J.-M. (2008). The role of algorithm and result comprehensibility of automated scheduling on complacency. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *18*(6), 603–620.
- Cegarra, J., Valéry, B., Avril E., Calmettes, C., & Navarro, J. (2019). *OpenMATB: A Multi-Attribute Task Battery promoting task customization, software extendability and experiment replicability*. Manuscript submitted for publication.
- Chavaillaz, A., Wastell, D., & Sauer, J. (2016). System reliability, performance and trust in adaptable automation. *Applied Ergonomics*, *52*, 333-342.
- Comstock, J. R. & Amegard, R. J. (1992). *The Multi-Attribute Task Battery for human operator workload and strategic behavior research* (Tech. Memorandum 104174). Hampton, VA: NASA Langley Research Center.
- Crocoll, W. M., & Coury, B. G. (1990). Status or recommendation: Selecting the type of information for decision aiding. In *Proceedings of the Human Factors society annual meeting*, (Vol. 34, No. 19, pp. 1524-1528). Los Angeles, CA: SAGE Publications.
- Cummings, M. L., & Mitchell, P. J. (2007). Operator scheduling strategies in supervisory control of multiple UAVs. *Aerospace Science and Technology*, *11*, 339–348.
- Gressgard, L. J., Hansen, K., & Iversen, F. (2013). Automation systems and work process safety: Assessing the significance of human and organizational factors in offshore drilling automation. *Journal of Information Technology Management*, *24*(2), 47.
- Hoc, J.-M. (2000). La relation homme-machine en situation dynamique. *Revue d'intelligence artificielle*, *14*(1-2), 55–71.
- Lorenz, B., Di Nocera, F., Röttger, S., & Parasuraman, R. (2002). Automated fault management in a simulated space flight microworld. *Aviation, Space, and Environmental Medicine*, *73*, 886–897.

- Manzey, D., Reichenbach, J., & Onnasch, L. (2009). Human performance consequences of automated decision aids in states of fatigue. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 53, No. 4, pp. 329-333)*. Los Angeles, CA: Sage Publications.
- Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human Factors, 46*(2), 196-204.
- Molloy, R., & Parasuraman, R. (1996). Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors, 38*(2), 311–322.
- Moray, N., & Inagaki, T. (2000). Attention and complacency. *Theoretical Issues in Ergonomics Science, 1*(4), 354–365.
- Mosier, K. L., & Skitka, L. J. (1996). Human decision makers and automated decision aids: Made for each other? In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and application* (pp. 201–220). Mahwah, NJ: Erlbaum
- Navarro, J. (2019). A state of science on highly automated driving, *Theoretical Issues in Ergonomics Science, 20*:3, 366-396
- Navarro, J., François, M., & Mars, F. (2016). Obstacle avoidance under automated steering: Impact on driving and gaze behaviours. *Transportation Research Part F: Traffic Psychology and Behaviour, 43*, 315–324.
- Navarro, J., Heuveline, L., Avril, E., & Cegarra, J. (2018). *Influence of Human-Machine Interactions and Task Demand on Automation Selection and Use. Ergonomics, 1–39.*
- Onnasch, L. (2015). Crossing the boundaries of automation—Function allocation and reliability. *International Journal of Human-Computer Studies, 76*, 12-21.
- Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human Factors, 56*(3), 476–488.
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors, 52*(3), 381–410.

- Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse, *39*(2), 230–253.
- 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., 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.
- Rovira, E., Cross, A., Leitch, E., & Bonaceto, C. (2014). Displaying contextual information reduces the costs of imperfect decision automation in rapid retasking of ISR assets. *Human Factors*, *56*(6), 1036-1049.
- Rovira, E., McGarry, K., & Parasuraman, R. (2007). Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors*, *49*(1), 76-87.
- Rovira, E., Zinni, M., & Parasuraman, R. (2002). Effects of information and decision automation on multi-task performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, (Vol. 46, No. 3, pp. 327-331). Sage CA: Los Angeles, CA: SAGE Publications.
- 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*. Hampton, VA: National Aeronautics and Space Administration, Langley Research Center.
- Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 1926–1943). New York: Wiley.
- Sarter, N. B., & Schroeder, B. (2001). Supporting decision making and action selection under time pressure and uncertainty: The case of in-flight icing. *Human Factors*, *43*(4), 573-583.
- Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Individual differences in monitoring failures of automation. *The Journal of General Psychology*, *120*(3), 357–373.

- Singh, I. L., Sharma, H. O., & Parasuraman, R. (2000). Effects of training and automation reliability on monitoring performance in a flight simulation task. In *Proceedings of the Human factors and Ergonomics Society Annual Meeting*, (Vol. 44, No. 13, pp. 53-56). Los Angeles, CA: SAGE Publications.
- 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.
- Wesley, D., & Dau, L. A. (2017). Complacency and automation bias in the enbridge pipeline disaster. *Ergonomics in Design*, 25(1), 17–22.
- Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science*, 8(3), 201–212.
- Wickens, C. D., Sebok, A., Li, H., Sarter, N., & Gacy, A. M. (2015). Using modeling and simulation to predict operator performance and automation-induced complacency with robotic automation: A case study and empirical validation. *Human Factors*, 57(6), 959-975.
- Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation: Promises and problems. *Ergonomics*, 23(10), 995-1011.
- Zybak, S., Scerbo, M. W., & Ashdown, A. (2016). System Reliability, Trust, and Complacency in Fetal Heart Rate Monitoring. *Proceedings of the Human factors and Ergonomics Society Annual Meeting*, 60, 1250-1254.

ARTICLE 2.

AUTOMATION TYPE AND RELIABILITY IMPACT ON VISUAL AUTOMATION MONITORING AND HUMAN PERFORMANCES

Avril Eugénie¹, Julien Cegarra¹, Liên Wioland², and Jordan Navarro³

¹Avril Eugénie and Julien Cegarra, Département de Psychologie, Université de Toulouse, INU Champollion, Albi, France. Eugenie.avril@univ-jfc.fr; Julien.cegarra@univ-jfc.fr

²Liên Wioland, Département de l'Homme au travail, INRS, Paris, France. Lien.wioland@inrs.fr

³Jordan Navarro, Département de Psychologie, Université Lumière Lyon 2, Bron, France. Institut Universitaire de France, Paris. Jordan.navarro@univ-lyon2.fr

Research article

Abstract

We compared automation monitoring evolution of static or adaptive automation of four different reliability levels over 90 minutes. Previous studies have demonstrated degraded human performances when monitoring automation and that it is possible to mitigate this monitoring performance drop thanks to adaptive automation. We used the Open Multi-Attribute Task Battery to manipulate two type of automation (static automation without manual take-over session and adaptive automation with planned take-over sessions) and four levels of reliability. Participants performed three simultaneous tasks in which one task was automated. Our results suggests that a perfectly reliable or a totally unreliable automation lead to different strategies of the participants in terms of visual allocation policy. Under static automation the time spent looking at the automated task in the 0% reliability level increased over the duration of the experiment, the opposite was observed for 100% reliability level. Although similar, this pattern of results was largely mitigated under adaptive automation. For static automation, the reported data also showed a direct link between trust in automation and visual scanning strategies. The more trust increased the less the automated task was looked.

Keywords: automation, task allocation, manual, reliability, monitoring

Introduction

Automated systems became an important support for human operators thanks to their multiple benefits (e.g., load reduction, safety enhancement, time saving). However, these systems are not always perfectly reliable which can lead operators to follow incorrect advice or make inappropriate decisions that can result in major problems (e.g., safety issues, Chavaillaz, Wastell, & Sauer, 2016). The advent of autonomous automation has changed ever-more the role of the human operator from active controller to automation supervisor system; who has to react to the potential failures of the automated system (Metzger & Parasuraman, 2001). For this reason, imperfect automation (i.e., not fully reliable) became an important issue for human factors specialists, especially to better understand the potential effects of reliability and type of automation on human behavior and suggest solutions to countermeasure these effects (Wickens & Dixon, 2007).

Monitoring an automated system during an extended period of time in order to correct potential malfunctions is a challenging situation for humans (Navarro, 2019a; 2019b) that can lead to inefficient monitoring referred as the *complacency* phenomenon (Parasuraman, Molloy, & Singh, 1993; Parasuraman & Manzey, 2010). This is even more an issue when function allocation between the automated system and the operator stays constant over time. Parasuraman, Molloy, and Singh (1993) used a multitasking environment, the Multi-Attribute Task Battery (Santiago-Espada, Myer, Latorella, & Comstock, 2011) composed of three simultaneous tasks (tracking, monitoring and resources management tasks) to compare constant and variable automation, with low (56.25%) and high (87.5%) levels of reliability. With variable automation, reliability alternates between high and low reliability over time, whereas with constant automation, reliability remains fixed over time. Participants performance was assessed by the number of detections of system malfunctions. The results

showed that under constant automation, the detection rate was poorer (33%) compared to variable reliability condition (82%).

This result suggests that humans need a certain amount of stimulation to remain alert and attentive. Within the automation context, humans tend to favor manual task completion over automation completion in order to avoid boredom (Navarro & Osiurak, 2015). So as to keep the human in the decision loop, adaptive automation (i.e. automation takes the decision to change automation level) has been introduced (Scerbo, 2006). Adaptive automation was found to improve situation awareness (Kaber, Riley, Tan, & Endsley, 2001), to enhance performance (Parasuraman, Mouloua, & Molloy, 1996) and to reduce workload (Bailey et al., 2006). Parasuraman, Mouloua, and Molloy (1996) investigated two types of adaptive automation using the MATB environment. These adaptive automations altered between sessions under automation and manual sessions. Participants started with 10-minute sessions in which all MATB tasks were performed manually. After this session, they performed three sessions of 30 minutes in which the monitoring task was automated (at 56.25% of reliability) while the two other tasks remained under manual control. In the first adaptive automation called “model-based”, the control always returned to the participant (manual session) after 40 minutes under automation. The second adaptive automation was based on performance. Participants were given manual control at planned moments if the performance was under a specific rate (55%). Results showed a similar increase of performance (i.e. malfunction detection rate) for the two adaptive automations.

Although the implementation of manual sessions during automation monitoring helps the human operator avoid monitoring inefficiency, the negative effects of imperfect automation must be considered. Oakley, Mouloua, and Hancock (2003) highlighted that static automation reliability level was directly impacting performance. These authors reported that above 35% of unreliability, the detection of system malfunctions by participants decreased.

Similarly, based on a meta-analysis, Wickens and Dixon (2007) reported that automation with a reliability level under 70% may be more prejudicial than no automation.

In the reported work, the impact of automation reliability (from fully reliable to fully unreliable, with two intermediate reliability levels) was manipulated in combination with static and adaptive automation (similar to the model-based automation used by Parasuraman et al., 1996) over a prolonged period of time. The main objective was to investigate how adaptive automation reliability impact human performances, as compared to static automation, in case of malfunction. On top of performances investigations, visual behaviors were also under the scope of investigations. Indeed, if monitoring performances can be assessed directly through performances, visual scanning behaviors of the automated task were undertaken to shed light on automation monitoring strategies. Gaze behaviors are of particular relevance in the current context because directly linked to the allocation of visual attention (Moray, 2003). Bagheri and Jamieson (2004a) reported an increase of the time between two gazes at the automated task to be monitored, for static automation with a reliability level of 87.5% as compared to static automation with a reliability level of 56.25%. Suggesting that highly reliable automation trigger visual attention reallocation compared to less reliable automation. Another objective of the reported experiment is to extend investigations on visual scanning to both static and adaptive automation fully reliable and fully unreliable.

It has been argued that trust in automation and self-confidence may influence automation induced complacency (Lee & See, 2004, Moray & Inagaki, 2000; Moray, Inagaki, & Itoh, 2000). In their meta-analysis, Wickens and Dixon (2007) also noticed that trust in automation was able to modulate the 70% automation reliability threshold. Several models of trust in human-automation interaction have been proposed (e.g. Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003; Lee & See, 2004; Hoff & Bashir, 2015). Some studies suggested that operators were insensitive to automation reliability (Dzindolet, Pierce, Beck, Dawe, & Anderson, 2001), while others showed the opposite (Parasuraman et al., 1993;

Wiegmann, Rich, & Zhang, 2001) and that operators modulate their trust depending on the automation reliability (Lee & Moray, 1992). Regarding self-confidence, Lee and Moray (1994) found that operators preferred manual control over automatic control when their self-confidence was greater than their trust in automation. Conversely, when their self-confidence was lower than their trust in automation, they preferred automated control even if the reliability was low (Moray, Inagaki, & Itoh, 2000).

In the current experiment, four levels of automation reliability (0%; 56.25%; 87.5% or 100%) were used with static automation (hereafter referred as No Take-Over condition - NTO) and with an adaptive automation (hereafter referred as Take-Over condition -TO-). The focus was specifically set on visual scanning behaviors, human performances in case of automation malfunction, trust and self-confidence changes over the duration of the experiment. We hypothesize that an increase of the reliability level would lead to monitoring inefficiency with static automation (NTO), with less time spent looking at the task to be monitored (reliability levels of 87.5% and 100%) and a performance drop in case of malfunction (reliability level of 87.5%). Monitoring inefficiency is expected to be mitigated by adaptive automation (TO). Concerning trust in automation, we hypothesized that participants would trust automation according to automation reliability. The better the reliability, the higher trust in automation will be. We also assume that self-confidence would be negatively correlated to trust in automation.

Method

Participants

Eighty postgraduate students (62 women and 18 men) at the University of Albi aged from 18 to 28 years ($M=19.9$ years, $SD=2.2$) participated in the experiment. All had a normal or corrected-to-normal vision and had no prior experience with the task at hand. They were randomly assigned to one of the eight conditions. This research complied with the American

Psychological Association Code of Ethics (1992). Informed consent was obtained from each volunteer.

Apparatus

The Multi-Attribute Task Battery

All tasks were run on a computer with a 23" monitor (1920 × 1080 pixels). A flight environment simulation, the Multi-Attribute Task Battery (MATB II, Santiago-Espada et al., 2011) was performed through the use of an open version, named the OpenMATB (Figure 1, Cegarra, Valéry, Avril, Calmettes, & Navarro, 2020) and composed of three of the five original MATB-II tasks. Participants were instructed to perform as accurately and as fast as possible the three tasks. During the whole experiment, eye movements were recorded at 250 Hz, using a RED 500 system (SensoMotoric Instruments).

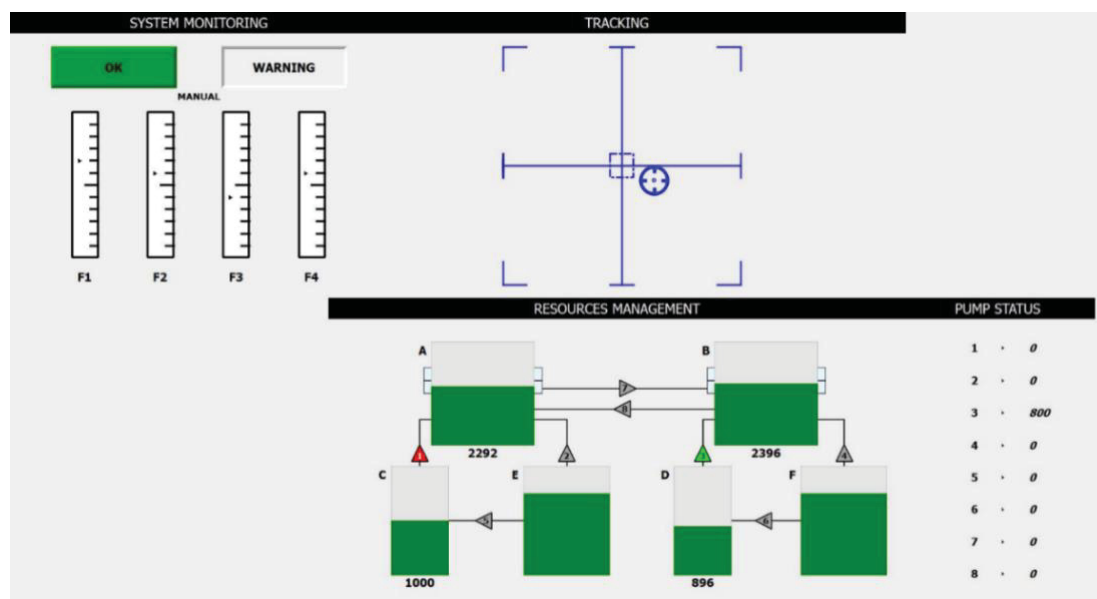


Figure 1. Screenshot of the OpenMATB used in the experiment: monitoring task (top left), tracking task (top center) and resources management task (bottom).

In the *monitoring task*, participants have to monitor four gauges (named F1 to F4) in which a cursor oscillates around the center. In case of a system malfunction, one of the four cursors freeze to the top or the bottom of a gauge. Participants then have up to 10s to press the

corresponding key (F1 to F4 on the keyboard) to unfreeze it. If they failed to do so, the cursor was unfrozen by the program and a missed detection counted for the human operator.

In the *tracking task*, participants must keep a cursor (i.e. red circle, Figure 1) in a central area (i.e. blue square, Figure 1) using a joystick. The cursor continuously drifts away from the central area following a forcing function determined by a sum of sinusoidal waves, for the X and Y axes, with a cut-off frequency of 0.06 hertz. As a result, the cursor path is comparable to a Lissajous curve, appearing random but strictly identical for all participants, regardless of the compensations operated.

In the *resources management task*, participants are required to maintain a given level (2500 L) in the two main tanks (A and B) that are continuously depleting at a constant rate (800 L/min). To compensate for the liquid depletion in both tanks, participants have to activate pumps to allow a fuel transfer from one tank to another. To do so participants had to press keys 1-8 on the keyboard, each key number was associated to a given pump (pumps number were continuously displayed inside the grey triangle representing a pump on the interface, see Figure 1). Keys 1-8 were able to activate and deactivate pumps as many times as needed. The eight pumps allow connections between the six tanks. All eight pumps were directional (e.g. pump 4 allowed fluid transfer from tank F to tank B exclusively, see Figure 1). The flow rate of each pump was displayed on the right side of the resources management task. When a pump is activated, the color grey (inactive) change to green (activate); red (pump 1 on Figure 1) means the pump is out of order and can't be used for 30 seconds.

Monitoring task automation

In this study, only the monitoring task was automated. Resources management and tracking tasks were to be performed without any automation.

In case of a system malfunction in the monitoring task (i.e. cursor freeze at the top or bottom of one of the four gauges), automation simultaneously signaled and corrected the malfunction. First the automation signaled the malfunction through a gauge failure led light “the warning light” that turn red until the malfunction was corrected. Secondly, automation corrected the malfunction with a delay of 1.5 seconds after the red signal to allow time of participants to notice the intervention of the automation.

In case of automation failure, system malfunctions were not signaled by the warning red light and were not corrected automatically. In that case, participants had to detect and correct the malfunction manually by pressing the key corresponding to the gauge where the cursor freeze (F1 to F4).

Procedure

After informed consent signature, participants were presented with the three tasks. Then the eye-tracker was calibrated. Next, participants were trained to perform manually each of the three MATB tasks individually for two minutes (2 min x 3 tasks). Participants were then informed that during the experiment, the monitoring task would be automated but was subject to failures. The functioning of automation was explained to participants using screenshots of the MATB. No further details about automation reliability were given. To finish the training phase, participants were exposed to the three MATB-II tasks simultaneously for four minutes. During that final training session, the monitoring task was automated, out of 16 system malfunctions, 8 were signaled and corrected by automation.

Then, the experimental sessions, in which participants performed the three tasks simultaneously, started and lasted 90 minutes (9 consecutive sessions of 10 minutes), following by questionnaires completion time. Before the fourth and seventh session, participants completed the NASA-TLX and two questions in order to evaluate (a) the level of self-

confidence to perform the monitoring task and (b) the level of their trust in automation. According to the experimental condition, participants had to take-over control during 20 min out of the 90 min of the experimental task (i.e. Take-Over condition) or not (i.e. No Take-Over condition) and the monitoring task was automated with one level of reliability.

Experimental Design

Eight groups of 10 participants were composed to carry out a between-participant design [4 (automation reliability level: 0%; 56.25%; 87.5% or 100%) x 2 (type of automation: TO, NTO)], see Figure 2.

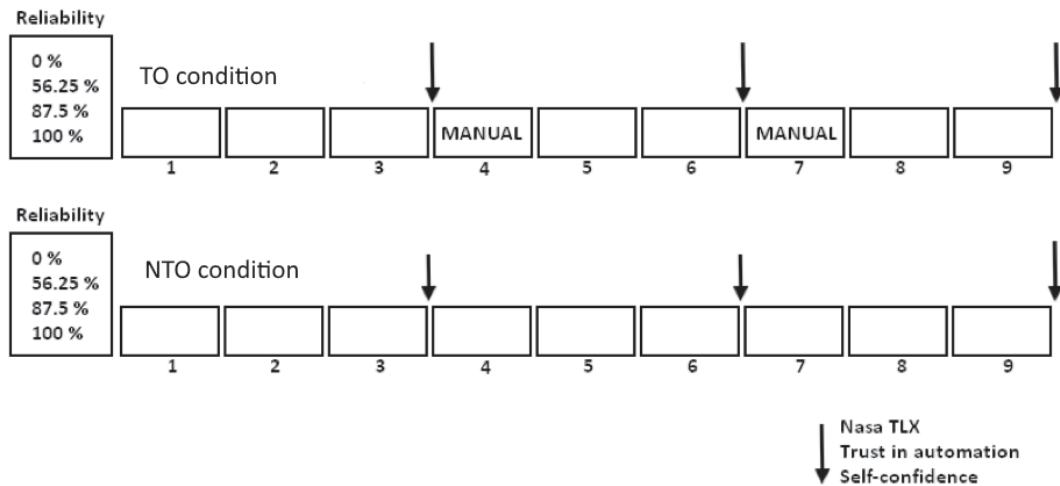


Figure 2. Experimental conditions and experimental sessions. All participants performed nine consecutive sessions of 10 minutes. Half of the participants were confronted to the No Take-Over condition -NTO- (with automation available for all 9 sessions), the other half was confronted to Take-Over condition -TO- (with automation available for all sessions but session 4 and 7). Both NTO and TO conditions, were composed of four sub-groups of participants facing a different automation reliability level (0; 56.25; 87.5 or 100%).

Each of the nine 10-minutes sessions included sixteen systems malfunctions, separated from each other by a delay that was randomly chosen ($M=39s$; $SD=16,3s$). Table 1 presents automation hit and misses for an automated session depending on the level of reliability. Under TO, session 4 and 7 had to be carried out manually (representing about 22% of the

experimental time). Participants were informed with the word “MANUAL” placed above the monitoring task during all sessions 4 and 7 duration, when automation was available again “AUTO ON” was displayed for 20s at the beginning of sessions 5 and 8.

Table 1. Number of automations hits, misses and total number of system malfunction per session with automation depending on automation level of reliability.

Level of reliability	Automation hits	Automation misses	Total system malfunctions
0 %	0	16	16
56,25%	9	7	16
87,5 %	14	2	16
100 %	16	0	16

Measures

Relative fixation time. Concerning ocular behavior, three Areas Of Interest (AOI) corresponding to the three MATB tasks were defined. For each AOI, the relative fixation time, defined as the sum of the recorded fixation durations, was computed (see Onnasch et al., 2015). One fixation was characterized as a set of gaze samples, separated from each other by up to 25 pixels, and representing a time window of at least 50 milliseconds.

Subjective measures. At the end of the third and the sixth session, the participants had to complete a computerized version of the NASA-TLX (Hart & Staveland, 1988). Moreover, the participants were asked to evaluate their self-confidence and trust into automation by responding to both questions “How high was your self-confidence to perform the monitoring task?” and “How much did you trust the automation of the monitoring task to detect the system malfunction?” (Lee & Moray, 1994) on a scale from 0 (completely disagree) to 10 (completely agree).

Performances. For the monitoring task, the detection rate (number of malfunctions corrected by participants) and the mean reaction time (mean delay between the system malfunction appearance and its correction) were analyzed. In order to allow direct comparison between the different levels of automation reliability, detection rates were expressed as a ratio (number of malfunctions corrected by participants / number of malfunctions that could have been corrected by participants). Resulting in detection rates performances expressed between 0 (no malfunctions corrected) and 1 (all malfunctions corrected). No data were collected for the automation reliability 100% as automation successfully detected and corrected all malfunctions.

For the tracking task, the performance was assessed through the proportion of time that the cursor spent inside the target area. Regarding the resources management task, the mean distance to the target level was computed for the two depleting tanks (A and B), and the two distances were averaged.

Results

No Take-Over (NTO) condition

Monitoring task

Relative fixation time. We conducted t-test comparisons on the relative fixation time to investigate the evolution of the gaze time spent at the monitoring task between session 1 and session 9 for each level of reliability.

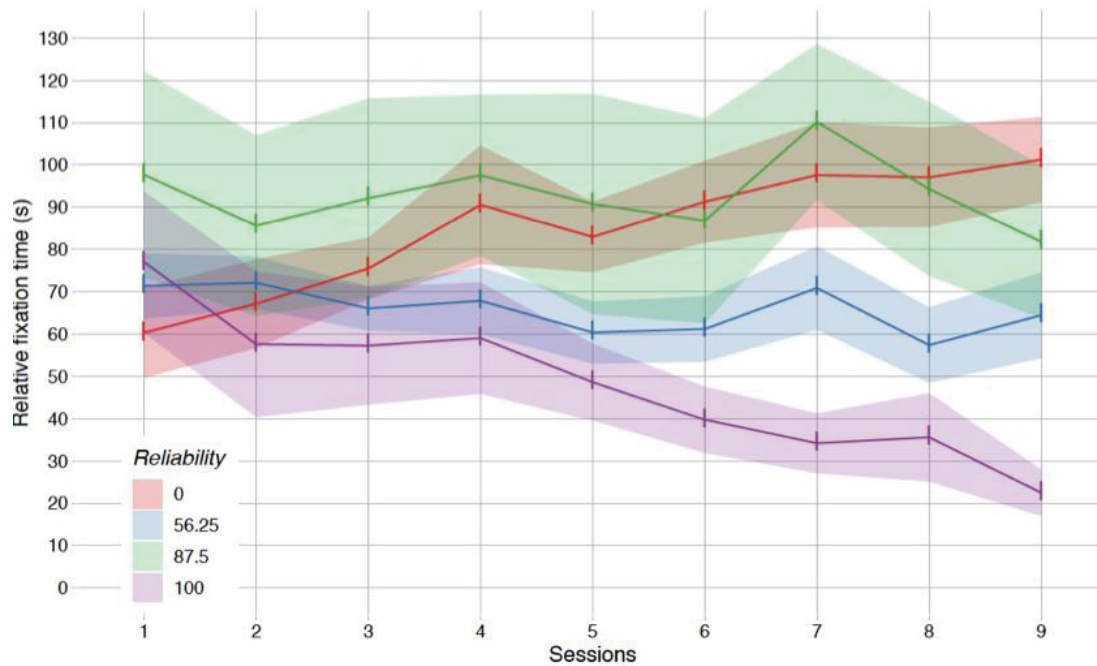


Figure 3. Evolution of relative fixation time (s) over the nine 10-min sessions in the automation group without take over, depending on the reliability level. Error envelop represent standard errors.

With NTO and 0% of reliability, participants looked at the monitoring task significantly less during session 1 than during session 9 ($t(9) = -2.669, p < .05, r = .44$, see figure 3). For reliability levels of 56.25% and 87.5% no significant differences between session 1 and 9 were recorded ($t(9) = .778, p = .45$ and $t(9) = 1.210, p = .25$ respectively). With the reliability level of 100%, participants spent significantly more time looking at the monitoring task during session 1 than during session 9 ($t(9) = 3.667, p < .01, r = .60$). See Table 2 for detailed descriptive data.

Table 2. Mean relative fixation times \pm standard error and difference between the session 1 and the session 9 for NTO condition along with the four levels of reliability.

Reliability	Session 1	Session 9	Difference (s)	$t(9)$	p
-------------	-----------	-----------	----------------	--------	-----

0%	60.31 ± 10.61	101.26 ± 10.07	40.95	-2.669	0.026*
56.25%	71.36 ± 24.18	64.45 ± 32.07	-6.91	0.778	0.45
87.5%	97.70 ± 24.35	81.82 ± 18.09	-15.88	1.210	0.25
100%	77.02 ± 16.76	22.48 ± 5.46	-54.54	3.667	0.005**

Performance.

Detection rate. We conducted *t*-tests on the detection rate to investigate the evolution of the performance on the monitoring task between session 1 and session 9 for each level of reliability (see figure 4).

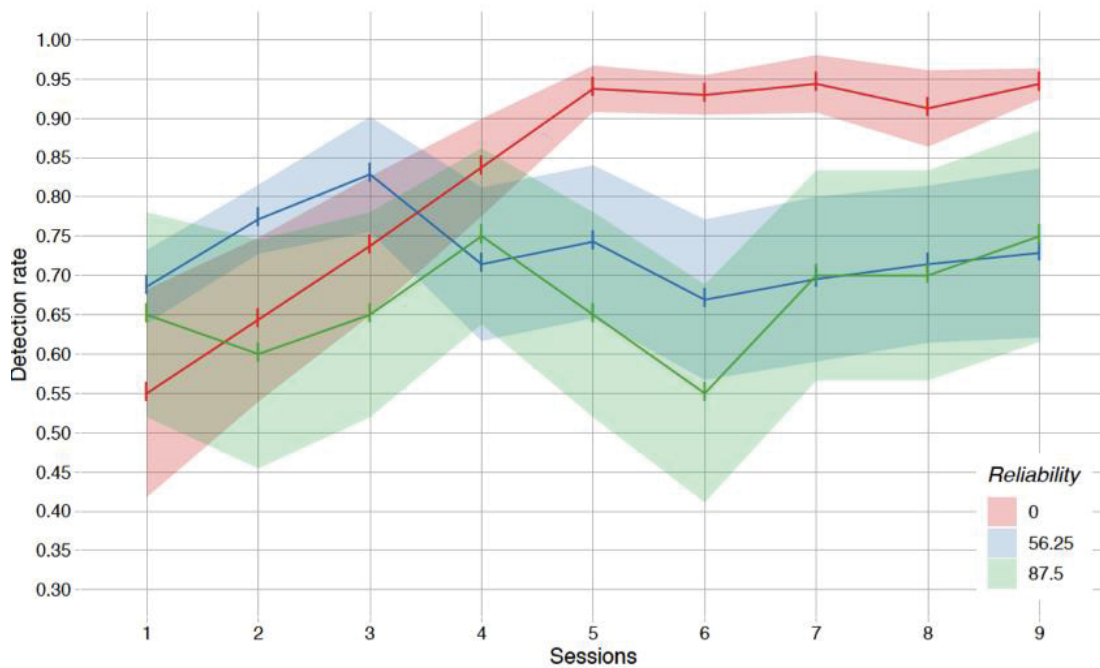


Figure 4. Evolution of detection rate over the nine 10-min sessions in NTO condition depending on automation reliability. Error envelop represent standard errors.

With NTO automation and 0% of reliability, the detection rate increased in session 9 compared to session 1, ($t(9) = -3.124, p < .05, r = .52$). For reliability levels of 56.25% and 87.5

% no significant differences between session 1 and 9 were recorded ($t(9) = -0.439, p = .67$ and $t(9) = -1.0, p = .34$ respectively). See Table 3 for detailed descriptive data.

Table 3. Mean detection rate \pm standard errors and difference between the session 1 and the session 9 for NTO condition along with the four levels of reliability.

Reliability	Session 1	Session 9	Difference	$t(9)$	p
0%	0.55 \pm 0.13	0.94 \pm 0.06	0.39	-3.124	0.012*
56.25%	0.68 \pm 0.05	0.73 \pm 0.11	0.05	-0.439	0.671
87.5%	0.65 \pm 0.13	0.75 \pm 0.13	0.10	-1.0	0.343

Responses time. We analyzed response time with t -tests comparisons. In the NTO condition and 0% of reliability, participants were faster to correct system malfunctions during session 1 than during session 9 ($t(9) = 3.192, p < 0.015$). For reliability levels of 56.25% and 87.5% no significant differences between session 1 and 9 were found ($t(9) = 0.364, p = .726$ and $t(9) = 0.916, p = .390$), see Table 4.

Table 4. Mean response time \pm standard errors and difference between session 1 and the session 9 for NTO condition along with the four levels of reliability.

Reliability	Session 1	Session 9	Difference	t	p
0%	4.50 \pm 0.27	3.42 \pm 0.35	-1.08	$t(9) = 3.192$	0.015*
56.25%	4.75 \pm 0.53	4.50 \pm 0.58	-0.25	$t(9) = 0.364$	0.726
87.5%	5.56 \pm 0.79	5.51 \pm 0.60	-0.05	$t(9) = 0.916$	0.390

Tracking and resource management tasks

We conducted t -tests comparisons on detection rate for the un-automated tracking and resource management tasks. In the tracking task, participants had a better performance in

session 9 than in session 1 for 0% and 56.25%. For reliability levels of 87.5% and 100% no significant differences between session 1 and session 9 were found. For the resource management task, *t*-tests comparisons ran on performance did not revealed any significant differences between session 1 and session 9 (see Table 5 for details).

Table 5. Mean performance \pm standard errors and difference between the session 1 and the session 9 for NTO condition along with the four levels of reliability for the tracking and resource management tasks.

	Reliability	Mean session 1 to 9	Session 1	Session 9	Difference	<i>t</i> (9)	<i>p</i>
Tracking task	0%	0.34 \pm 0.05	0.30 \pm 0.05	0.37 \pm 0.06	0.07	-2.920	0.017*
	56.25%	0.30 \pm 0.05	0.25 \pm 0.04	0.35 \pm 0.05	0.01	-4.585	0.001**
	87.5%	0.26 \pm 0.04	0.24 \pm 0.04	0.27 \pm 0.04	0.03	-1.095	0.302
	100%	0.21 \pm 0.03	0.17 \pm 0.03	0.22 \pm 0.03	0.05	-2.029	0.073
Resource management task	0%	0.92 \pm 0.05	0.88 \pm 0.07	0.93 \pm 0.05	0.05	-0.779	0.456
	56.25%	0.88 \pm 0.01	0.89 \pm 0.09	0.85 \pm 0.09	-0.04	1.613	0.141
	87.5%	0.80 \pm 0.09	0.83 \pm 0.06	0.77 \pm 0.12	-0.06	0.708	0.497
	100%	0.74 \pm 0.13	0.75 \pm 0.11	0.79 \pm 0.13	0.04	0.540	0.602

Take-Over (TO) condition

Monitoring task

Relative fixation time. Figure 5 presents the relative fixation time for each reliability level in TO condition over the nine sessions of the experiment.

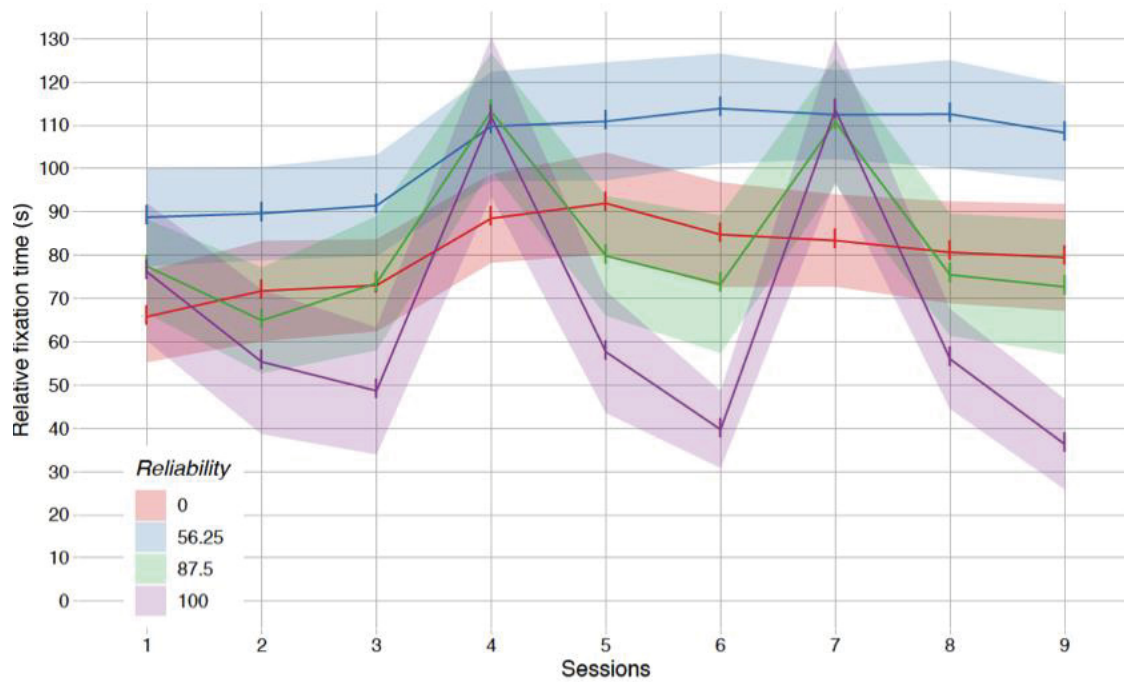


Figure 5. Evolution of relative fixation time (s) over the nine 10-min sessions in the TO condition along with the four levels of reliability. Session 4 and 7 are manual. Error envelopes represent standard errors.

The time spent looking at the monitoring task did not differ significantly between session 1 and session 9 for reliability level of 0% ($t(9) = -1.433, p = 0.186$), 56.25% ($t(9) = -1.706, p = 0.122$) and 87.5% ($t(9) = 0.365, p = .725$). However, for reliability level of 100%, participants gazed significantly more at the monitoring task during session 1 than during session 9 ($t(9) = 4.063, p < .01, r = .64$). See Table 6 for detailed descriptive data.

Table 6. Mean relative fixation times \pm standard error and difference between the session 1 and the session 9 in the TO condition along with the four levels of reliability.

Reliability	Session 1	Session 9	Difference	$t(9)$	p
0%	65.80 \pm 10.50	79.50 \pm 12.36	13.70	-1.433	0.186
56.25%	88.78 \pm 11.34	108.27 \pm 11.13	19.49	1.706	0.122

87.5%	77.40 ± 10.85	72.70 ± 15.50	-4.70	0.365	0.725
100%	76.12 ± 15.87	36.32 ± 10.43	-39.80	4.063	0.003**

We also have conducted *t*-test comparisons between each manual session and (a) the preceding automated session (session 3 VS session 4 and session 6 VS session 7) and (b) the subsequent automated session (session 4 VS session 5 and session 7 VS session 8).

Table 7. Mean relative fixation times and standard errors (SE) in the TO condition along with the four levels of reliability for manual sessions VS preceding automated sessions (in grey) and manual sessions VS subsequent automated sessions (in white).

		Reliability (%)			
		0	56.25	87.5	100
Session 3 vs 4	Mean	73.04 vs 88.50	91.46 vs 109.76	73.54 vs 113.20	48.67 vs 112.00
	SE	10.64 vs 10.25	11.66 vs 12.58	15.47 vs 13.46	14.65 vs 18.37
	<i>t</i> (9)	-3.940	-3.416	-2.076	-4.193
	<i>p</i>	0.003**	0.008**	0.068	0.002**
Session 4 vs 5	Mean	88.50 vs 91.99	109.76 vs 110.91	113.20 vs 79.84	112.00 vs 57.65
	SE	10.25 vs 11.72	12.58 vs 13.62	13.46 vs 13.81	18.37 vs 14.07
	<i>t</i> (9)	-0.606	-0.162	1.674	3.668
	<i>p</i>	0.559	0.875	0.129	0.005**
Session 6 vs 7	Mean	84.77 vs 83.39	113.88 vs 112.45	73.31 vs 110.80	39.74 vs 113.54
	SE	12.06 vs 10.65	12.69 vs 10.24	15.88 vs 14.52	8.86 vs 16.42
	<i>t</i> (9)	0.255	0.147	-1.945	-5.401
	<i>p</i>	0.804	0.886	0.084	0.000***
Session 7 vs 8	Mean	83.39 vs 80.66	112.45 vs 112.59	110.80 vs 75.47	113.54 vs 56.03
	SE	10.65 vs 11.77	10.24 vs 12.47	14.52 vs 14.10	16.42 vs 11.51
	<i>t</i> (9)	0.708	-0.020	1.758	4.429
	<i>p</i>	0.497	0.985	0.113	0.002**

The results showed a significant difference between session 3 (automated) and 4 (manual) for the all reliability level except 87.5% that failed to reach the significance level (see Table 7). For all reliability levels, participants spent more time looking at the monitoring task during the first manual session (session 4) than during the preceding automated session (session 3). The impact of the manual session last in the subsequent automated session (i.e. session 5) with no significant decrease in relative fixation time between session 4 and 5 for all reliability levels but 100%. If a similar pattern of results is observed for the second manual session (i.e. session 7), significant differences are only observed for the reliability level of 100%.

Moreover, the time spent looking at the monitoring task is extremely similar for sessions 4 and 7, for reliability levels of 56.25%, 87.5% and 100% (respectively 109.76 s, 113.20 s and 112 s for session 4 and 112.45 s, 110.80 s and 113.54 s for session 7).

Performance.

Detection rate. Detection rate did not differ significantly between session 1 and session 9 for all three reliability levels (see figure 6): 0% ($t(9) = -2.144, p = .061$), 56.25% ($t(9) = -1.673, p = .129$) and 87.5% ($t(9) = -0.429, p = .678$). Table 8 also presents the detailed results.

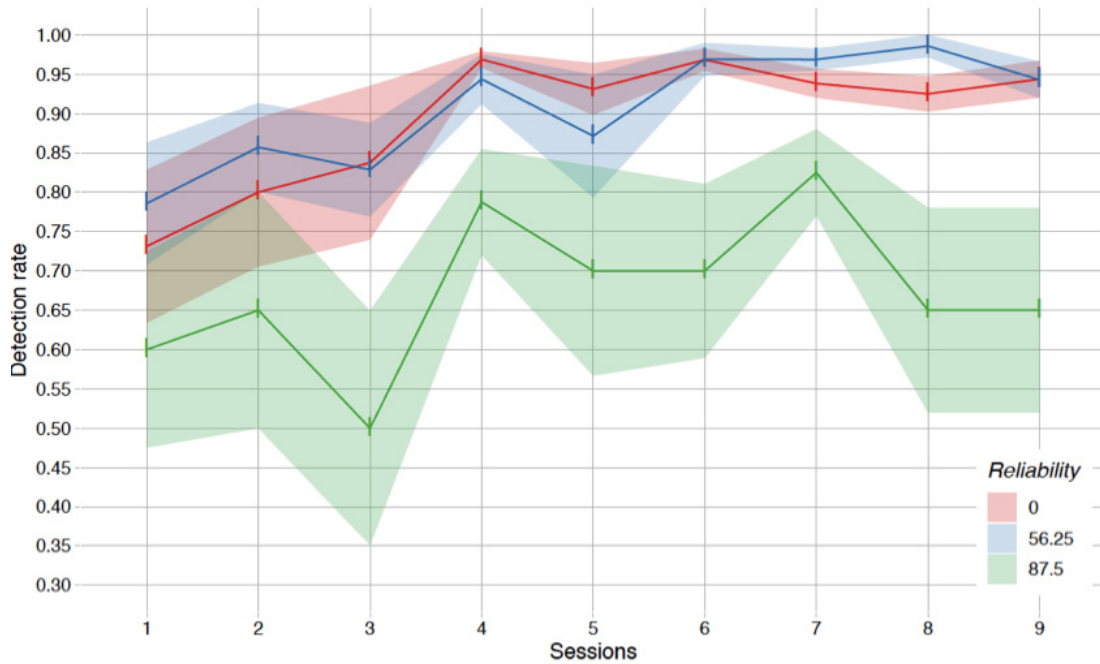


Figure 6. Evolution of the detection rate over the nine 10-min sessions depending on the reliability levels in TO condition. Session 4 and 7 are manual. Error envelop represent standard errors.

Table 8. Mean detection rate \pm standard errors and difference between the session 1 and the session 9 in the TO condition along with the four levels of reliability.

Reliability	Session 1	Session 9	Difference	$t(9)$	p
0%	0.73 \pm 0.10	0.94 \pm 0.02	0.21	-2.144	0.061
56.25%	0.78 \pm 0.08	0.94 \pm 0.02	0.16	-1.673	0.129
87.5%	0.60 \pm 0.12	0.65 \pm 0.13	0.05	-0.429	0.678

We also conducted t -test comparisons between each manual session and (a) the preceding session (session 3 VS session 4 and session 6 VS session 7) and (b) the subsequent session (session 4 VS session 5 and session 7 VS session 8). In 0% of reliability level, participants had a lower detection rate in manual session 7 than in session 6 ($t(9)=2.983$,

$p < .05$). In 56.25% and 87.5% of reliability, participants had a better performance in manual session 4 than in session 3, ($t(9) = -2.634$, $p < .05$ and $t(9) = -2.769$, $p < .05$ respectively). All other comparisons were not significant. See Table 9 for detailed descriptive data.

Table 9. Mean detection rate times and standard errors (SE) in the TO condition along with the four levels of reliability for manual sessions VS preceding automated sessions (in grey) and manual sessions VS subsequent automated sessions (in white).

		Reliability		
		0	56.25	87.5
Session 3 vs 4	Mean	0.84 vs 0.97	0.83 vs 0.94	0.50 vs 0.79
	SE	0.10 vs 0.01	0.06 vs 0.03	0.15 vs 0.07
	$t(9)$	-1.406	-2.634	-2.769
	p	0.193	0.027*	0.022*
Session 4 vs 5	Mean	0.97 vs 0.93	0.94 vs 0.87	0.79 vs 0.70
	SE	0.01 vs 0.03	0.03 vs 0.08	0.07 vs 0.13
	$t(9)$	1.327	1.491	0.751
	p	0.217	0.170	0.472
Session 6 vs 7	Mean	0.97 vs 0.94	0.97 vs 0.97	0.70 vs 0.82
	SE	0.01 vs 0.02	0.02 vs 0.01	0.11 vs 0.05
	$t(9)$	2.983	0.020	-1.519
	p	0.015*	0.984	0.163
Session 7 vs 8	Mean	0.94 vs 0.92	0.97 vs 0.99	0.82 vs 0.65
	SE	0.02 vs 0.02	0.01 vs 0.01	0.05 vs 0.13
	$t(9)$	0.605	-0.760	1.472
	p	0.532	0.467	0.175

Response time. With 0% of automation reliability, participants take less time to correct system malfunctions during session 1 than during session 9 ($t(8) = 5.625$, $p < 0.001$). For reliability levels of 56.25% and 87.5% no significant differences between session 1 and 9 were recorded ($t(9) = 0.220$, $p = .831$ and $t(5) = 0.217$, $p = .837$). See Table 10 for detailed descriptive data.

Table 10. Mean response time \pm standard errors and difference between session 1 and the session 9 in the TO condition along with the four levels of reliability.

Reliability	Session 1	Session 9	Difference	<i>t</i>	<i>p</i>
0%	4.10 ± 0.31	3.43 ± 0.35	-0.67	t(8)=5.625	0.000***
56.25%	3.76 ± 0.37	3.65 ± 0.28	-0.11	t(9)=0.220	0.831
87.5%	3.97 ± 0.62	4.17 ± 0.50	0.20	t(5)= - 0.217	0.837

Tracking and resource management tasks

We conducted *t*-tests comparisons on detection rate for the un-automated tracking and resource management tasks. No significant differences between session 1 and 9 were recorded for both tasks except for 100% reliability level in the tracking task (see Table 11 for details).

Table 11. Mean performance ± standard errors and difference between the session 1 and the session 9 for TO condition along with the four levels of reliability for the tracking and resource management tasks.

	Reliability	Mean session 1 to 9	Session 1	Session 9	Difference	<i>t</i> (9)	<i>p</i>
Tracking task	0%	0.30 ± 0.04	0.27 ± 0.04	0.32 ± 0.04	0.05	-1.330	0.216
	56.25%	0.30 ± 0.04	0.28 ± 0.04	0.29 ± 0.05	0.01	-0.546	0.598
	87.5%	0.25 ± 0.02	0.24 ± 0.03	0.26 ± 0.02	0.02	-0.807	0.440
	100%	0.32 ± 0.04	0.28 ± 0.04	0.35 ± 0.04	0.07	-3.541	0.006**
Ressources management task	0%	0.95 ± 0.02	0.98 ± 0.10	0.97 ± 0.01	-0.01	0.513	0.620

	56.25%	0.82 ± 0.02	0.86 ± 0.06	0.90 ± 0.07	0.04	-0.413	0.689
Ressources management task	87.5%	0.71 ± 0.12	0.72 ± 0.11	0.68 ± 0.14	-0.04	0.696	0.504
	100%	0.80 ± 0.13	0.84 ± 0.10	0.79 ± 0.13	-0.05	1.066	0.314

Subjective data

A mixed ANOVA with sessions (i.e. 3, 6 & 9, see Figure 2) as within-participants factor and automation reliability (0, 56,25, 87,5 and 100%) and type of automation (NTO and TO) as between-participants factors was used to investigate self-confidence, trust in automation and NASA-TLX measures.

Self-confidence. Regarding self-confidence the ANOVA revealed a main effect of the session factor ($F(2,135)=7.532, p<.001$) and an interaction between the session factor and automation reliability ($F(6, 136)= 2.494, p<.05, \eta_p^2=0.094$). Post-hoc analyses with Bonferroni comparisons indicated significant differences for the reliability level of 100%. Self-confidence in session 3 ($M=4.2, SE=0.494$) was smaller than in session 6 ($M=5.900, SE=0.499$) and 9 ($M=6.0, SE=0.520$).

Trust in automation. Regarding trust in automation, and contrary to self-confidence data, no main effect of the session factor ($F(2,136)=1.347, p=0.263$) nor interaction between the session factor and automation reliability ($F(2, 136)= 1.556, p=0.169$) were found. However, the ANOVA revealed a main effect of automation reliability ($F(3,72)=1,437, p<.01, \eta_p^2=.192$) and an interaction between the session factor and automation type ($F(2, 136)= 3.549, p<.05, \eta_p^2=0.047$).

Concerning the main effect of automation reliability, post-hoc showed a significant difference between the level 0% ($M=4.28, SE=0.53$) and 87.5% ($M=6.53, SE=0.43$) as well as between the level 0% and 100% ($M=6.48, SE=0.58$). Concerning the interaction between the

session factor and automation type, no significant differences were revealed by post-hoc analysis.

To further investigate the link between trust in automation evolution and the time spent gazing at the monitoring task (i.e. relative fixation time) over time, the difference between the measures made at session 9 and at session 3 was made for both variables. A positive difference between session 9 and 3 indicates an increase of trust in automation/relative fixation time whereas a negative difference indicates a decrease of trust in automation/relative fixation time. These differences between session 9 and 3 for trust in automation and relative fixation time were tested with a Pearson correlation test.

A significant negative correlation between trust in automation and the time spent gazing at the monitoring task was found ($r(78)=-0.253, p<.05$). The more trust in automation decreases, the more the relative fixation time on the monitoring task increases. However, a closer look at the data revealed that this correlation was significant only in the NTO condition, $r(38)=-0.469, p<.01$ and that no significant correlation is present in the TO condition, $r(38)=0.106, p=.258$ (see Figure 7).

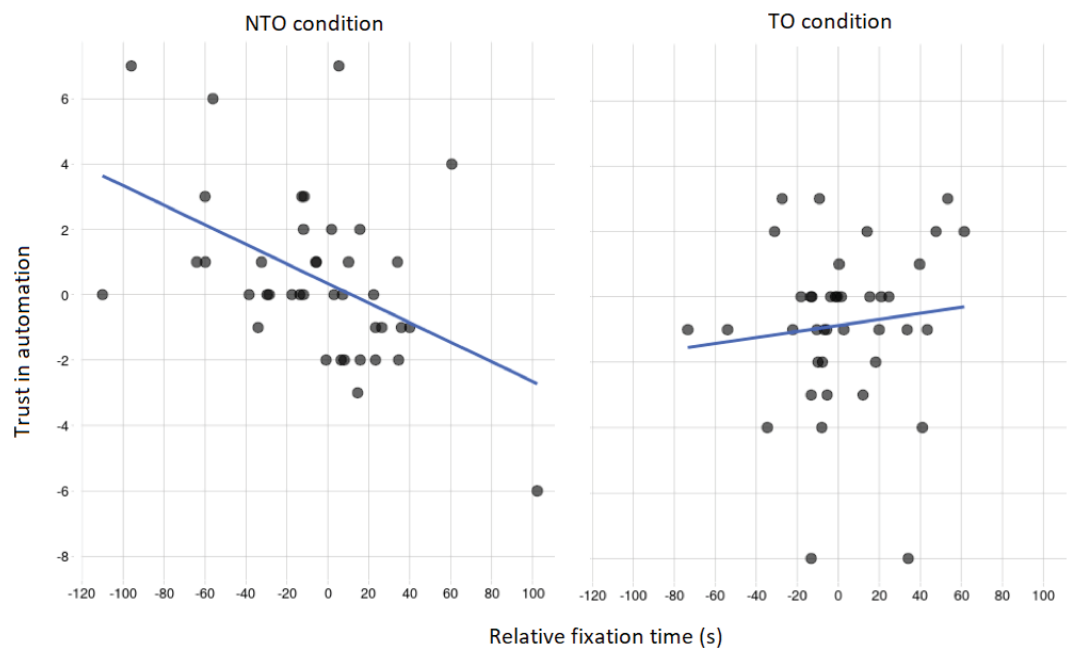


Figure 7. Correlation between the trust in automation evolution over time and relative fixation time at the monitoring task over time in NTO (left) and TO conditions (right).

NASA TLX. The ANOVA ran on subjective workload did not revealed any significant differences.

Discussion

The purpose of this study was to investigate the effects of two types of automation (i.e., without -NTO- or with planned take-over -TO-) along with four levels of reliability (i.e., 0%, 100% with two intermediate levels: 56.25 % and 87.5 %) over the 90 minutes of MATB use, on performances, trust and visual allocation policy of the participants.

The results showed that participants confronted with a perfectly reliable static automation without take-over during 90 minutes spent less and less time looking at the automated monitoring task (see Figure 3) with a decrease of about 71% of the time spent on monitoring task between session 1 (first 10 minutes) and session 9 (last 10 minutes). On the contrary, with an automation fully unreliable, participants spent more and more time looking at the automated monitoring task with about 68% more time in session 9 compared to session 1. The amount of time spent looking at the automated monitoring task remained constant over the 90 minutes spent under automation with 56.25 or 87.5 % of reliability.

Concerning adaptive automation with planned take-overs, the time spent looking at the automated monitoring task between session 9 and 1 decreased of about 50% for the 100% perfectly reliable automation whereas a no significant increase of about 21% for 0% reliable automation was observed. Similar to NTO condition, the time spent looking at the automated monitoring task remained constant between session 1 and 9 under TO condition with 56.25% and 87.5 % level of reliability. Return of automation after the 10-minutes manual session was not followed by a significant difference of the time spent looking at the monitoring task in

subsequent automated sessions (except for the 100% reliability level). This result indicates that manual sessions had an impact on visual behaviors on the subsequent automated session, in line with previous observations with automated sessions following manual session, benefit of the performances improvement gained during manual session (Mouloua, Parasuraman, & Molloy 1993; Parasuraman et al., 1996).

Moreover, the relative fixation time measured in the two manual sessions of the TO condition are similar for the reliability levels of 56.25, 87.5 and 100% ($M=112s$, $SD=12.02$). This time seems to correspond to the time needed to perform manually the monitoring task that was previously automated, regardless of the level of reliability of automation. Also, this time is much longer (26s) than the time spent looking at the monitoring task in the 0% reliability level ($M=86s$, $SE=10.17$). This difference can correspond to the visual cost of returning to manual control for the levels of reliability 87.5 and 100%. However, for the 56.25% level, the time spent looking at the monitoring task stayed high after the first manual session in the following automated sessions. This indicates that with 56.25% of reliability, monitoring the automated task required the same visual involvement than manual performance. Under a 70% level of reliability automation is not beneficial compared to performing the task manually (Oakley et al., 2003; Onnasch, Wickens, Li, & Manzey, 2014; Wickens & Dixon, 2007). This is reinforced by these results on visual sampling.

Regarding subjective measures, no significant difference was found on subjective workload. However, an increase of trust in automation was measured depending on automation reliability level. Trust was rated as significantly lower for 0% reliable automation ($M=4.28$, $SE=0.53$) compared to both 87.5% ($M=6.53$, $SE=0.43$) and 100% ($M=6.48$, $SE=0.58$) reliability levels. Contrary to Bagheri and Jamieson (2004a) observations, our data confirmed that automation reliability is a factor of the trust in automation, in line with previous studies where trust was found associated to automation reliability (e.g. Moray, Inagaki & Itoh,

2000; Ross, Szalma, Hancock, Barnette, & Taylor, 2008; Wiegmann et al., 2001). However, if our participants' reported trust was sensitive to automation reliability, trust and automation reliability were not perfectly correlated. Participants did not give a score of 0/10 of trust in automation for the level 0% of reliability although automation was totally unreliable. Similarly, they didn't give a score of 10/10 for the 100% level of reliability although automation was perfectly reliable. This can be interpreted as a difficulty for the participants to estimate accurately automation reliability (Bailey & Scerbo, 2007; Wiegmann et al., 2001; Wiegmann, 2002). For example, in the reported data, the score of trust in automation was very close for the 87.5 ($M=6.53$, $SE=0.43$) and 100% ($M=6.48$, $SE=0.58$) condition of reliability. The lack of feedback is possibly a factor that interfered with a good level of reliability (Bagheri & Jamieson, 2004b; Seong & Bisantz, 2008; Wang, Jamieson, & Hollands, 2009). In future studies, a feedback about automation reliability would enhance our understanding of trust in automation depending on automation reliability level.

For the 100% level of reliability, participants reported an increase of self-confidence over the time of the experiment (self-confidence after sessions 6 and 9 was greater than after session 3). The perfect automation functioning leads to an increase of participants self-confidence, suggesting that participants consider automation performance as their own performance. This result argues in favor of a symbiotic relationship between individuals and automation (Brangier, Dufresne, & Hammes-adelès, 2009; Hancock, 2009), with participants considering automation performances as an extension of their own performance.

Moreover, in the NTO condition, a correlation was found between trust in automation change between sessions 3 and 9, and the time spent looking at the monitoring task. The more trust in automation decreases, the more the relative fixation time on the monitoring task increases. This result suggests that participants regulated their relative fixation time on the monitoring task according to their trust in automation. Bagheri and Jamieson (2004b) found

similar results and showed that the more participants trusted automation, the smaller the mean time between fixations at the monitoring task. However, this result was not found in the TO condition. This might indicate that adaptive automation is mitigating the link between trust in automation and visual allocation strategies.

For both NTO and TO conditions performances at the monitoring task between session 1 and session 9 were improved only with automation reliability level 0%. This suggests that participants increased their level of performance at the task during the experiment thanks to the repetitions of manual task completion. At the reverse with an automated task completion that leave room for manual interventions (56,25% and 87% of reliability), participants did not improve their performances over the duration of the experiment. If further investigations are required in order to control the frequency of manual actions, a possible explanation would be that the learning curve of a given task is flattened by automation.

For concurrent tasks, performance did not change significantly for resource management task between session 1 and 9, and only a few improvements have been measured between the session 1 and 9 in the tracking task. Progress in the tracking task could be explained by the nature of that task that requires fine motor control with the use of a joystick. A visuo-motor coordination that could be improved within minutes (Sailer et al., 2005). Whereas the resources management task is a much more complex task requiring problem solving and would have required more time to elicit performance improvements. Importantly, automation of the system monitoring task (including type and reliability level) did not translated into major performance changes in the concurrent tasks.

Conclusion

This study suggests that a perfectly reliable or a totally unreliable automation lead to different strategies of the participants in terms of visual allocation. In a 0% reliability condition, participants spend more time looking on the monitoring task over the time, while in the 100% reliability condition, they sampled less the automated task. This study also highlighted the difficulty for the participants to calibrate accurately their trust in automation, a central tendency in the assessments being recorded. Nevertheless, trust in automation evolution during the experiment was found negatively correlated with evolution of the time spent looking at the automated task for static automation. Further researches should be undertaken to investigate how the different levels of automation might impact on adaptive automation performances (Navarro et al., 2018). Experiments over longer periods of time should also be undertaken so as to explore longer term effects of adaptive automation on human behaviors.

Acknowledgments

This research was supported by grants from ANR (« Smart Planning » – ANR-16-CE26-0017).

The Authors declares that there is no conflict of interest.

References

- Bagheri, N., & Jamieson, G. A. (2004a). Considering subjective trust and monitoring behavior in assessing automation-induced “complacency.” *Human performance, situation awareness, and automation: Current research and trends*, 54–59.
- Bagheri, N., & Jamieson, G. A. (2004b). The impact of context-related reliability on automation failure detection and scanning behaviour. *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, 1, 212–217.
- Bailey, N. R., & Scerbo, M. W. (2007). Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science*, 8(4), 321-348. <https://doi.org/10.1080/14639220500535301>
- Bailey, N. R., Scerbo, M. W., Freeman, F. G., Mikulka, P. J., & Scott, L. A. (2006). Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. *Human factors*, 48(4), 693–709.
- Brangier, E., Dufresne, A., & Hammes-Adel , S. (2009). Approche symbiotique de la relation

- humain-technologie : perspectives pour l'ergonomie informatique. *Le Travail Humain*, 72(4), 333. <https://doi.org/10.3917/th.724.0333>
- Cegarra, J., Valéry, B., Avril, E., Calmettes, C., & Navarro, J. (2020). OpenMATB: A Multi-Attribute Task Battery promoting task customization, software extensibility and experiment replicability. *Behavior Research Methods*, 1-11.
- Chavaillaz, A., Wastell, D., & Sauer, J. (2016). System reliability, performance and trust in adaptable automation. *Applied Ergonomics*, 52, 333–342.
- Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International journal of human-computer studies*, 58(6), 697–718.
- Dzindolet, M. T., Pierce, L. G., Beck, H. P., Dawe, L. A., & Anderson, B. W. (2001). Predicting misuse and disuse of combat identification systems. *Military Psychology*, 13(3), 147–164.
- Hancock, P. A. (2009). Mind, Machine and Morality. In *Mind, Machine and Morality: Toward a Philosophy of Human-Technology Symbiosis*. Retrieved from <https://www.taylorfrancis.com/books/9781315248905>
- Hoff, K. A., & Bashir, M. (2015). Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors*, 57(3), 407-434. <https://doi.org/10.1177/0018720814547570>
- Kaber, D. B., Riley, J. M., Tan, K.-W., & Endsley, M. R. (2001). On the design of adaptive automation for complex systems. *International Journal of Cognitive Ergonomics*, 5(1), 37–57.
- 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., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10), 1243–1270.
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors*, 46(1), 50–80.
- Metzger, U., & Parasuraman, R. (2001). The Role of the Air Traffic Controller in Future Air Traffic Management: An Empirical Study of Active Control versus Passive Monitoring. *Human Factors*, 43(4), 519-528. <https://doi.org/10.1518/001872001775870421>
- Moray, N. (2003). Monitoring, complacency, scepticism and eutactic behaviour. *International Journal of Industrial Ergonomics*, 31(3), 175–178.
- Moray, Neville, & Inagaki, T. (2000). Attention and complacency. *Theoretical Issues in Ergonomics*

- Science*, 1(4), 354–365. <https://doi.org/10.1080/14639220052399159>
- Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology. Applied*, 6(1), 44–58. <https://doi.org/10.1037//0278-7393.6.1.44>
- Mouloua, M., Parasuraman, R., & Molloy, R. (1993). Monitoring automation failures: Effects of single and multi-adaptive function allocation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 37, 1–5.
- Navarro, J. (2019a). A state of science on highly automated driving. *Theoretical Issues in Ergonomics Science*, 20(3), 366–396. <https://doi.org/10.1080/1463922X.2018.1439544>
- Navarro, J. (2019b). Are highly automated vehicles as useful as dishwashers? *Cogent Psychology*, 6(1), 1–9. <https://doi.org/10.1080/23311908.2019.1575655>
- Navarro, J., & Osiurak, F. (2015). When do we use automatic tools rather than doing a task manually? Influence of automatic tool speed. *The American journal of psychology*, 128(1), 77–88.
- Navarro, Jordan, Heuveline, L., Avril, E., & Cegarra, J. (2018). Influence of human-machine interactions and task demand on automation selection and use. *Ergonomics*, 61(12), 1601–1612. <https://doi.org/10.1080/00140139.2018.1501517>
- Oakley, B., Mouloua, M., & Hancock, P. (2003). Effects of automation reliability on human monitoring performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 47, 188–190.
- Onnasch, L. (2015). Crossing the boundaries of automation—Function allocation and reliability. *International Journal of Human-Computer Studies*, 76, 12-21.
- Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human factors*, 56(3), 476–488.
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human factors*, 52(3), 381–410.
- 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*, 38(4), 665–679.
- Ross, J. M., Szalma, J. L., Hancock, P. A., Barnett, J. S., & Taylor, G. (2008). The Effect of Automation Reliability on User Automation Trust and Reliance in a Search-and-

- Rescue Scenario. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 52(19), 1340-1344. <https://doi.org/10.1177/154193120805201908>
- Sailer, U., Flanagan, J. R., & Johansson, R. S. (2005). Eye-hand coordination during learning of a novel visuomotor task. *Journal of Neuroscience*, 25(39), 8833–8842. <https://doi.org/10.1523/JNEUROSCI.2658-05.2005>
- Santiago-Espada, Y., Myer, R. R., Latorella, K. A., & Comstock Jr, J. R. (2011). *The multi-attribute task battery ii (matb-ii) software for human performance and workload research: A user's guide*.
- Scerbo, M. W. (2006). "Dynamic Function Allocation." In *International Encyclopedia of Ergonomics and Human Factors*, edited by Karwowski Waldemar, 2nd ed., 1080–1082. London: CRC/Taylor & Francis.
- Seong, Y., & Bisantz, A. M. (2008). The impact of cognitive feedback on judgment performance and trust with decision aids. *International Journal of Industrial Ergonomics*, 38(7-8), 608–625.
- Wang, L., Jamieson, G. A., & Hollands, J. G. (2009). Trust and reliance on an automated combat identification system. *Human factors*, 51(3), 281–291.
- Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science*, 8(3), 201–212.
- Wiegmann, D. A. (2002). Agreeing with automated diagnostic aids: A study of users' concurrence strategies. *Human Factors*, 44(1), 44–50.
- Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomics Science*, 2(4), 352–367.

ARTICLE 3.

DECISION SUPPORT SYSTEM RELIABILITY AND TRANSPARENCY: A QUALITATIVE ANALYSIS OF A PLANNING TASK IN ROAD FREIGHT TRANSPORT SECTOR

Avril Eugénie¹, Jordan Navarro², Liên Wioland³, Virginie Govaere³
and Julien Cegarra¹

¹*Avril Eugénie and Julien Cegarra, Département de Psychologie, Université de Toulouse, INU Champollion, Albi, France.*

²*Jordan Navarro, Département de Psychologie, Université Lumière Lyon 2, Bron, France. Institut Universitaire de France, Paris.*

³*Liên Wioland and Virginie Govaere, Département de l'Homme au travail, INRS, Paris, France.*

Research article

Abstract

The Road Freight Transport (RFT) sector is continuously evolving, being confronted with intense competition and growing pressure from customers and suppliers, tight delivery times, and many regulations. All these factors lead the RFT sector to face multiple economic, environmental, and safety challenges. In these companies, the planner has a strategic position. Planner's tasks include supervising the transport round, keeping contact with the drivers, providing new instructions when needed, ensuring order follow-up with customers, and many others. Such tasks are considered cognitively demanding. Therefore, planners have difficulties in assessing the consequences of their decisions in terms of economic, environmental, and safety dimensions. To help the planner and to anticipate the future computer-supported decision-making, we focused on the effect of system transparency and reliability. In this study, we confronted six planners with planning problems. We modified solutions and indicators to provide high and low levels of transparency. Starting from real issues, we also generated reliable and unreliable solutions (for these latter by adding invalid data). For each problem, planners were asked to assess the quality of the proposed solutions and indicators. The results showed that planners give a higher score to reliable scenarios displaying reliable information than to a less reliable scenario in which information is modified. This assessment of scenario is more adequate and easier with high transparency than with low transparency of information.

Keywords : transparency, reliability, planning, road freight transport, information

Introduction

The Road Freight Transport (RFT) is an important sector that groups several thousands of employees through no less than 37.200 companies in France. Due to this large number of companies, this sector faces constant competition and pressure from customers and suppliers (Govaere et al., 2018). The multiple challenges, such as economic, environmental, or safety/health, became a central issue for them. For example, the financial constraints impose tight delivery times and low flexibility that require optimization of delivery routes. The changes to on-line sales in the last few years increased these constraints.

In this context, the role of the planner is central to a road freight transport company. The planner must integrate the management strategies of the company in the delivery planning, which the drivers execute. During the delivery routes, the planner has to follow up and support the problems drivers may experience (e.g., road accidents, breakdown, or issues with customers, Khademi, 2016; Govaere & Wioland, 2018). The role of planners appears complex, and all of their missions involve several cognitive activities. Khademi et al. (2015) showed that planners spend about half of their working time (55%) to process information: taking information, managing data, and recovering crucial data from memory (Khademi, 2016).

Moreover, information processing is also involved in diagnosis and decision-making processes. Information management appears to be important when planning. The planner must manage some uncertainties related to the late or even missing information about the freight to transport (the exact number of pallets, for example). Moreover, as Larco et al. (2018) noted, the planning task embedded in the organization leads the planner to continually adapt to the situations and make decisions that depend on the context.

To relieve the planner, an automated system could be a solution. New technologies facilitate the creation of multiple automated systems that support humans in different tasks/functions. More precisely, Parasuraman et al. (2000) noted that an automated system supports four broad classes of functions: 1) information acquisition, 2) information analysis, 3) decision and action selection, and 4) action implementation. In their daily practice, planners use a word processor or spreadsheet. They can also use sophisticated tools as vehicle on-board computer, transport management system (TMS), or electronic data interchange (EDI). Finally, in the RFT sector, automated systems may take the form of decision support systems.

These decision support systems are challenging to design due to the strategic and tactical processes of planners (i.e., the expertise of planners) that may be difficult to consider in the design (Cegarra & Van Wezel, 2012; Guerin et al., 2012). Cegarra and Van Wezel (2012)

argued that the planning task could be poorly documented, and solution criteria can be poorly defined. Individual differences must also be considered, mostly because humans use different thinking processes when they plan (Cegarra, 2008). Moreover, establishing a criterion of performance in planning is sometimes tricky because other objectives may need to be satisfied (Gary et al., 1995). For these reasons, a decision support system should not be based on one single observed strategy but algorithms simplifying the socio-technical process. The variables considered in the algorithm must be determined beforehand (e.g., constraints, objectives, resources available), but in the sector of planning, more variables are unstable (Cegarra, 2012; Khademi, 2016). This design difficulty, coupled with the fact that decisions of the planners are context-dependent, influence the perception of the decision support systems: the solutions proposed by these decision support systems are often considered unreliable (Kerr, 1992) and sometimes planners leave them out (Van Wezel & Jorna, 2009; Van Wezel et al., 2010).

Human factors specialists have widely studied the relationship between automated aid reliability and its use by human operators (Oakley et al., 2003; Parasuraman et al., 1993; Parasuraman & Manzey, 2010). The research highlighted that reliability of an automated system is directly related to the trust of the human operator in the automated system (Lee & See 2004). Other factors, like transparency, are implied in the development of faith in automated aid. Chen et al. (2014) defined transparency as “the descriptive quality of an interface in terms of its ability to allow an operator to understand the intent, performance, plans and reasoning process of an intelligent agent” (p. 2). Automation transparency refers to the automation’s openness to information communicated through the interface to the operator: *what* the automation is doing, *which* information is being used, *how* it is being processed, and *when* it is provided (Bass & Pritchett, 2008). According to Chen et al. (2014), there are three levels of transparency. Level 1 is transparency support perception of the current state, goals, planning, and progress. Level 2 refers to transparency support comprehension of the reasoning behind the robot's behavior and limitations. Level 3 includes transparency support projection, predictions, and probabilities of failure/success based on the history of performance. For these authors, level 1 is the most prone to errors of perception when the information is not provided. In this case, failures are caused due to the lack of situation awareness (Endsley, 2017), as information is more or less transmitted to the human operator. According to Westin et al. (2016), for decision support systems, design decisions have to be made about how much information should be provided and in what ways. In their taxonomy, Parasuraman et al. (2000) noted that in the information function of an automated system, the information was more or less available. With a high level of transparency, human operators

have access to clear information that helps them make their decisions (e.g., raw information, that is to say, a low level of information). A low level of transparency pertains to less clear information (e.g., aggregated information, that is to say, a high level of information).

However, studies on the users' perceptions of the decision support system confronted with a high and low level of transparency (i.e., involving the display of clear or aggregated information) are lacking. Based on the previous studies (Parasuraman et al., 2000; Sarter & Schroeder, 2001), we decided to study the reliability of a decision support system across low and high transparency levels. This study was a part of the multidisciplinary project ANR "Smart-planning," which aimed to design a decision support system focusing on the challenges in the sector (i.e., integrated economic, environmental, or safety/health dimensions in their planning). This decision system (1) supports planning generation in which information is displayed to the human planner (number of kilometers, delivery times...) and (2) informs, through indicators, about the cost of each planning in terms of economic, environmental, or safety/health dimensions. First, the information on the proposed planning is highly transparent "low-level information" on kilometers, delivery times, and the like. Second, the planning is less transparent with a "high level of information" displayed through indicators, which aggregate elements related to the cost of planning for each dimension. Thus, the interface provides two levels of information transparency: a high level of information on the delivery route and a low level of information with the indicators.

This study aimed to increase the understanding of the operators' activity with this decision support system. In the sector of road freight transport, some companies still plan their delivery route manually. The switch to an automated system is a significant challenge. To guarantee efficient use in the right conditions, it is necessary to understand the operators' representation of automation and its effects.

The study aimed to understand how expert planners understand, perceive, and evaluate automated information (more and less reliable) according to two levels of transparency (low: raw information, high: aggregated information).

Hypothesis 1: Reliable information, whether high or low transparency (raw or aggregated information), should be perceived and assessed as higher than information with a lower level of reliability.

Hypothesis 2: Information with a high level of transparency should be better evaluated than information with a low level of transparency.

Method

As a part of the ANR Smart Planning project, two road freight transport companies participated in the consortium. The study took place in both companies where we collected real data from the daily practice (their customers, their constraints...). In this sector, planning consists of building a plan to deliver all the customers' orders during a specific day. This implies anticipating the availability of human resources (drivers) and material resources (vehicle). Thus, the result of the activity is daily planning, which contains a set of delivery routes. For a driver, delivery routes include the point of departure from the warehouse (pick-up time), the vehicle moving to a location (driving activity), delivery time, a delivery point (arrival time to a site, loading and unloading time), and the end of delivery (see Figure 1).

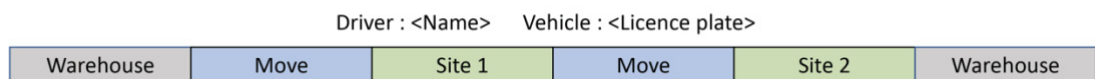


Figure 1. Example of one delivery route assigned to a driver and a vehicle

Participants

Six male expert planners aged from 40 to 45 years old participated in this study. Three participants were planners in the same road freight transport company in the North of France, and three others were planners in the same road freight transport company in Paris region. Planners have been with the companies for a long time: five to seventeen years for planners in the first company and two to five years in the second company. Five out of the six planners also had experience as drivers, which allowed them to know the constraints of the drivers.

Apparatus

To test our hypotheses, we used information with a high level of transparency, i.e., with raw information (condition 1). Second, we used information with a low level of transparency, i.e., with aggregate indicators (condition 2). For each condition, we constructed scenarios where we manipulated the reliability of the information presented (reliable or less reliable). A reliable scenario was named “reference scenario,” and a less reliable scenario was

named “modified scenario.” These two conditions were applied in both companies within their specific context.

Hereafter, the reliable scenario, i.e., reference scenario, reflects a delivery route usually carried out in the participating company, i.e., it is performed without automated aid, and the route is regularly practiced through its execution. Thus, we considered it reliable because it corresponds to the practice (i.e., the field) that planners usually implement. Reliable information was provided and validated by experts.

Based on these reliable scenarios, the information was modified to change its reliability. These modified routes are called modified scenarios (see description below). The reference and modified scenarios were, in each phase, presented on paper to the participants as if a decision support system had created them.

The first condition – Scenario with one delivery route: High level of transparency

The visible information in the scenarios in this condition is the name of the customer, the travel time to get from one customer to another, the different sites of loading and unloading, the time spent at these loading or unloading sites, hours of delivery required by customers, the number of kilometers between each site (see Figure 2 for an example).

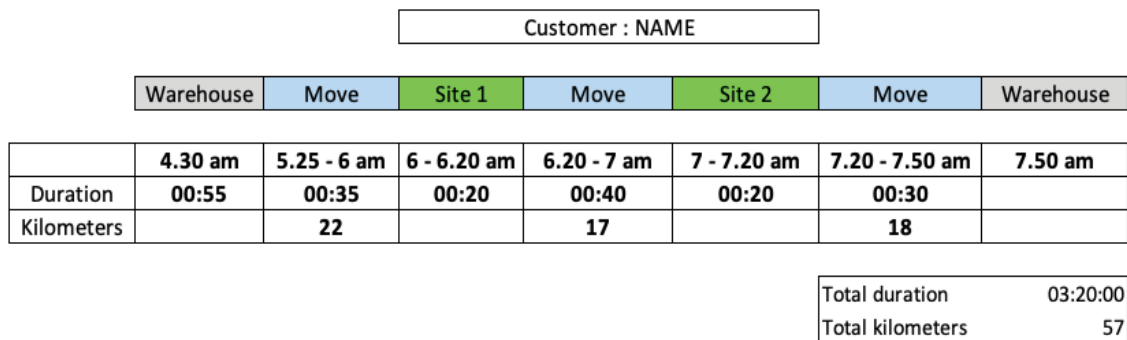


Figure 2. An example of reference route scenario in the first condition.

For example, in Figure 2, the driver's arrival at the *warehouse* is 4.30 am; after loading the truck, the driver is leaving the warehouse at 5.25 to the *site 1* located 22 kilometers away and 35 minutes of *drive*; the time spent at *site 1* corresponds to 20 minutes of unloading; the driver then moves to *site 2* (40 min and 17 kilometers); the time spent at *site 2* corresponds to 20 minutes of unloading; finally, the driver returns to the *warehouse* (30 min and 18 kilometers). The total duration of the delivery route (three hours and 20 min) and the total kilometers (57 km) are also given.

In Figure 3, we display the same information. However, the number of kilometers is modified to deviate from the reference scenario (information displayed is less reliable, corresponding to the modified scenario).

Customer : NAME							
Warehouse	Move	Site 1	Move	Site 2	Move	Warehouse	
	4.30 am	5.25 - 6 am	6 - 6.20 am	6.20 - 7 am	7 - 7.20 am	7.20 - 7.50 am	7.50 am
Duration	00:55	00:35	00:20	00:40	00:20	00:30	
Kilometers		13		10		30	

Total duration	03:20:00
Total kilometers	53

Figure 3. An example of modified route scenario in condition 1.

In this phase, ten scenarios were proposed. Five scenarios represented 5 reference routes (R-R), and 5 corresponded to the reference scenario in which information was modified (R-M). For example, a reference scenario (i.e., delivery route based on reliable information, R1-R) was modified only for the “number of kilometers to each site.” This scenario was called R1-M (modified). Table 1 present the information modified for each of the five modified scenarios.

Table 1. Several modified information in each modified scenario.

Modified scenarios (R-M)	Modified information
R1-M	The number of kilometers traveled between each site
R2-M	The number of kilometers traveled between each site The time spent at the unloading site
R3-M	The number of kilometers traveled between each site The time spent at the unloading site The time spent at the loading site
R4-M	The number of kilometers traveled between each site The time spent at the unloading site The time spent at the loading site The hours of delivery explicitly required by customers
R5-M	The number of kilometers traveled between each site The time spent at the unloading site The time spent at the loading site The hours of delivery explicitly required by customers The travel time to get from one site to another

A planner was confronted with five comparisons. A comparison was composed of a reference scenario (R-R) presented systematically with the same scenario in which information was modified (R-M).

The experts had to rate each scenario a scale from “1” (low-quality proposition of the decision support system) to 5 (very good quality proposition of the decision support system) and explain their choice for each scenario in each comparison. The explanations were recorded with a voice recorder during the experiment.

The second condition – Scenario with one delivery route and indicators: A low level of transparency

In the second phase (Figure 1), scenarios corresponded to the previous reference delivery route from the first condition. We added three indicators (I) that qualified the planning regarding economic, environmental, or health and safety costs. These costs were calculated from various sources and data gathered from the participating companies. Thus, these indicators corresponded to aggregate information. Each indicator was presented through a number that corresponded to the cost of the delivery route for each of the dimensions (safety/health, environmental, and economic). The higher the number, the higher the cost of the delivery route. The level of transparency of this information was low (Figure 4). In the example shown in Figure 4, the cost of the route was 456 for health and safety, 57 for economic, and 15 for environmental dimensions.

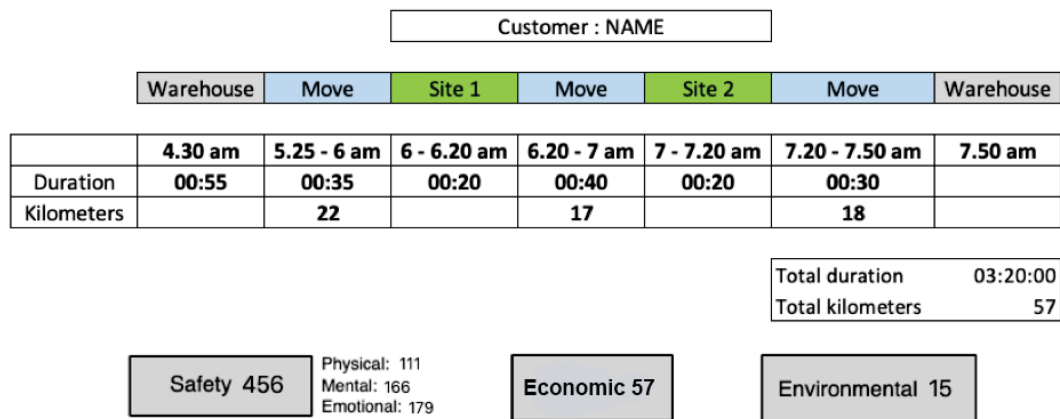


Figure 4. An example of reference scenario with the three indicators.

As in the previous phase, ten scenarios were created per company: five of these scenarios corresponded to the reference route (RI-R). The other five scenarios corresponded to the modified tour (RI-M). The scenarios were modified the same way for the three indicators, through an underestimation of the three indicators, which can be 20%, 30%, 40%, 50%, or 60% depending on the scenario. Thus, Figure 5 shows the reference scenario (RI-R) presented in Figure 4 when the underestimation of 20% was applied, i.e., a health and safety cost of 365, an economic cost of 46, and an environmental cost of 12.

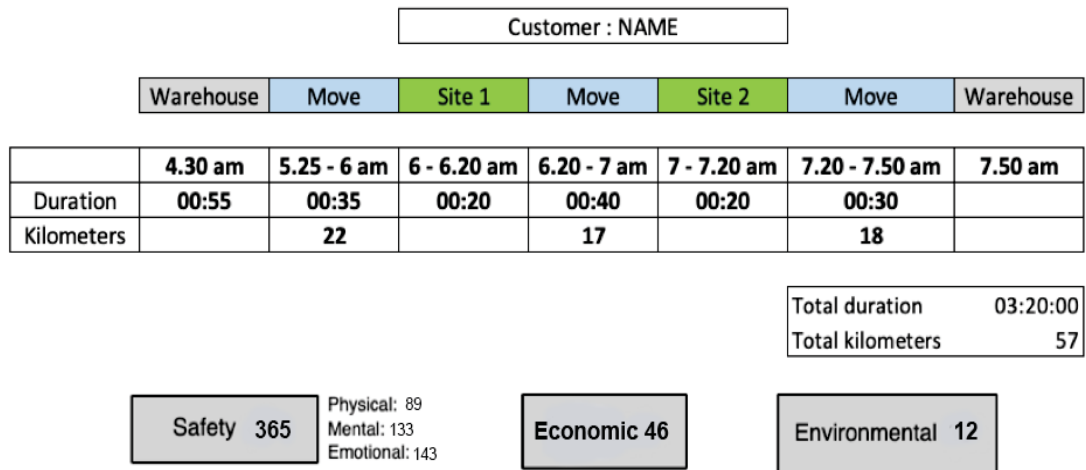


Figure 5. An example of a modified scenario with the three indicators.

In our study, for one of both the company (first company), we applied overestimation in indicators to explore the effect of this modification. In addition, we chose to apply overestimates and underestimates that could be significant in the second company. In Table 2, the modifications applied to the reference scenarios are presented for the second company. They range from an underestimate of 40% to an overestimate of 227%.

Table 2. Overestimates and underestimates applied to the scenario for the second company.

Scenarios	Indicators underestimated or overestimated
1	Overestimation of 1%
2	Underestimation of 40%
3	Underestimation of 32%
4	Overestimation of 35%
5	Underestimation of 20%
6	Overestimation of 100%

7	Overestimation of 52%
8	Overestimation of 118%
9	Overestimation of 161%
10	Overestimation of 227%

The task to be carried out and the instructions were the same as in the first condition; a comparison of scenario and an assessment with a score between 1 and 5. Similarly, the assessment of each scenario and comparison was explained and recorded.

Results

Condition 1– High Transparency Scenarios

In the first company, the reference scenarios were assessed at a mean of 3.47 ($SD=0.64$) and the modified scenarios at 2.67 ($SD=0.62$). In company 2, the reference scenarios had a mean of 3.73 ($SD=0.88$) and the modified scenarios of 2.2 ($SD=1.01$). The differences in the means of the scenarios between these two companies cannot be interpreted strictly because the scenarios were based on company data, which were not the same (different customers, different departure times, and the like). The modifications realized in the scenarios were based on the same information and "perimeter."

Table 3. The score given by each planner in both companies for each reference scenario (R-R) and modified scenario (R-M)

Scenario	Planner (Company)	Reference scenario	Modified scenario
R1-R / R1-M	1 (1)	4	3
	2 (1)	4	2
	3 (1)	4	3
	4 (2)	4	2
	5 (2)	4	2
	6 (2)	4	3
R2-R / R2-M	1 (1)	3	3
	2 (1)	3	3
	3 (1)	2	3
	4 (2)	5	4
	5 (2)	4	3
	6 (2)	4	3

R3-R / R3-M	1 (1)	4	3
	2 (1)	3	3
	3 (1)	3	3
	4 (2)	3	3
	5 (2)	3	3
	6 (2)	3	3
R4-R / R4-M	1 (1)	4	3
	2 (1)	3	3
	3 (1)	4	2
	4 (2)	3	2
	5 (2)	3	1
	6 (2)	2	1
R5-R / R5-M	1 (1)	4	3
	2 (1)	3	2
	3 (1)	4	1
	4 (2)	5	1
	5 (2)	5	1
	6 (2)	4	1

Note. The name of the scenario corresponds to the number of scenarios R” X” for reference “R” or modified “M” scenario. For each comparison, the scenario with the high score is colored in orange when a high score is given to the referential scenario and red when a high score is given to the modified scenario. Reference and modified scenarios that received the same score are colored in purple.

First, in Table 3, we note that only 3 evaluations of the quality of the "referent" scenario (out of 30) received the maximal score “5,” which corresponds to the quality of the scenario evaluated as "very good proposal" (i.e., 10% of the evaluations of the referent scenarios). This result was unexpected because these scenarios are usually proposed by the planners and executed by the drivers. Nevertheless, despite their uses and operational efficiencies, the planners affirmed that a route is determined based on a compromise between different constraints (economic constraints specific to the company and constraints in terms of the quality of service provided to the customer). Finally, a delivery route is not considered a "really good" quality but is somewhat acceptable in a given context.

The second result on referent/modified comparison, 70% (21/30) rated the referent scenario higher, 27% rated the scenarios as equivalent, and only one modified scenario was considered better than the referent scenarios. These ratings do not seem to depend on the amount of information modified (see Table 4).

Table 4. The number of reference scenarios rated higher than the modified scenario, and the number of reference scenarios considered equal to scenario modified.

	Evaluation of reference scenario > evaluation of the modified scenario	Evaluation of reference scenario = evaluation of the modified scenario
R1-R / R1-M 1 modified information	6 scenario /6 scenario	0 scenario /6 scenario
R2-R / R2-M 2 modified informations	3 scenario /6 scenario	2 scenario /6 scenario
R3-R/ R3-M 3 modified informations	1 scenario /6 scenario	5 scenario /6 scenario
R4-R/ R4-M 4 modified informations	5 scenario /6 scenario	1 scenario /6 scenario
R5-R / R5-M 5 modified informations	6 scenario /6 scenario	0 scenario /6 scenario
Total	21 scenario /30 scenario	14 scenario /30 scenario

Overall, 70% of reference scenarios were assessed as better than the modified scenario. This result was expected as the same routes were presented simultaneously, leaving the planner the possibility to compare each piece of information directly. The reference scenario that the drivers executed was known, controlled, and operational since.

Furthermore, 27% of the reference scenarios and modified scenarios were rated equally. This result was unexpected. Nevertheless, in these situations, the given score was “3”, that is to say, these scenarios were rated as of medium quality, regardless of whether they were reference or modified scenarios. Moreover, as previously mentioned, the reference scenarios are the result of compromises. For this reason, when the quality is considered as "medium" (i.e., acceptable), the information modified in terms of the number of kilometers traveled between each site, the time spent at unloading sites, or the time spent at the loading site did not lead to a different evaluation of the scenario.

Table 5 shows that the planner noted the modified information during the assessment of the scenario. The information is considered unreliable in the modified scenario compared to the reference scenario.

Table 5. Examples of explanations given by planners during their evaluation

Planner (Company)	
Planner 1 (1)	“The number of kilometers between “Genevilliers” (<i>name of the warehouse location</i>) and “Versailles” (<i>name of the location of the site 1</i>) is better in this scenario with 22 kilometers (i.e., reference scenario) than with 13 kilometers (i.e., modified scenario).”
Planner 2 (1)	“The time of move between Taverny (<i>name of the location of the site 2</i>) and the warehouse is too short on this delivery route (i.e., modified scenario). This one (i.e., reference scenario) is more coherent.”
Planner 3 (1)	“20 kilometers between the warehouse and “l’isle adam” (<i>name of the location of the site 1</i>)... It is too low. Forty kilometers (<i>number of kilometers in the reference scenario</i>) is more coherent... The time of unloading is also too long in “Taverny” (<i>name of the location of the site 2</i>)”
Planner 1 (2)	“There is an inconsistency in kilometers for this delivery route (i.e., <i>modified scenario</i>). There are too many kilometers between the warehouse and “Lompret” (<i>name of the location of the site 1</i>)”.
Planner 2 (2)	“The kilometers are too high in this scenario (i.e., modified scenario). There are too many kilometers between the warehouse and “Lompret” and between “Campus” and the return to the warehouse.”
Planner 3 (2)	“In this scenario (i.e., modified scenario), the time of unloading is too short at “Petite forêt” (<i>name of the location of the site 2</i>)... 10 minutes only is too short for the driver.”

Thus, at the end of these scenario comparisons and this first analysis, we observed that (1) the planners noted the modifications, (2) the number of modified information did not affect the assessment of the reliability of the scenario, and (3) reliable information was qualitatively better evaluated.

The second condition – Scenario with a low level of transparency

As a reminder, in one of the companies (first company), we proposed to apply overestimation to explore the effect of modifying information.

The reference scenario included three indicators (safety/ health, economic, and environmental), resulting from several sources (this explained the low transparency). There are cost indicators. Thus, only the indicators contained modified information corresponding to:

- (1) A decrease in the cost of the delivery route when the modification was negative.
- (2) An increase in the cost when the modification was positive.

First company

The score given by each planner for each reference and modified scenario (RI-R and RI-M) in the first company is displayed in the following table (Table 6).

Table 6. The score given by each planner in the first company for each reference scenario (RI-R) and modified scenario (RI-M)

Scenario	Planner	Reference scenario	Modified scenario
RI1-R / RI1-M (-20%)	1	2	3
	2	3	3
	3	1	3
RI2-R / RI2-M (-30%)	1	4	4
	2	4	3
	3	4	2
RI3-R / RI3-M (-40%)	1	3	3
	2	3	4
	3	2	4
RI4-R / RI4-M (-50%)	1	4	3
	2	3	5
	3	3	2
RI5-R / RI5-M (-60%)	1	3	1
	2	4	2
	3	4	2

Note. The name of the scene corresponds to the number of scenarios R (route) I (indicator) “X” for reference “R” or modified “M” scenario. For each comparison, the scenario with the high score is colored in orange when a high score is given to the referential scenario and red when a high score is given to the modified scenario. Reference and modified scenarios that received the same score are colored in purple.

Our results showed that with low transparency (i.e., indicators), the modified scenario was “assessed as better” than the reference scenario 5/15. One interpretation is that the planner favored this delivery route because it was identical but cheaper. However, we noted

that the indicator could be considered reliable as long as it remains close to the planner's representation.

When the indicator modification was -20% (lower costs allocated to the modified route), the planners assigned a higher score to the modified scenario, except for the second planner who rated the scenarios as identical in terms of their qualities. Overall, the assessment of the quality of the reference scenario was medium (between 1 and 3), while the modified scenario was assessed with a score of 3. It appears that in this scenario, the planners viewed the indicators (with a low level of transparency) presented in the reference scenario as having low reliability, regardless of the comparison.

Planner 1: «The indicators in both scenarios are too high for the safety/health cost. However, this delivery route (i.e., modified scenario) is better because the safety/healthy indicator is low. It's an easy delivery route because there is no handling; the location of the store is accessible. This indicator should be even lower ».
« The economic indicator is better in this scenario (i.e., modified scenario) and environment indicator is better in this scenario (i.e., reference scenario) given the vehicle used... »

Planner 2: « This scenario (i.e., reference scenario) is overestimated, and the other (i.e., modified scenario) is not good too... I give the same score to both scenarios. »

Planner 3: « The safety/health indicator is too high for both scenarios. Concerning the economic and environmental indicators, I don't have an opinion. It's difficult to judge similar numbers in comparison with the difference between the health/safety indicators.»

For the second comparison, the reference scenario was rated as good quality (score of 4). Overall, the modification of indicators of -30% was considered to be of lower quality. Based on the explanations, the quality assigned to the scenarios depended on the planners' perceived reliability of the indicators. A 30% reduction in indicator values was detected by all the planners and considered significant. In addition, the -30% modification was applied consistently across all three indicators. However, in the previous explanation, we detected compromises in decision-making: a compromise for planner 3 that ignored the values that the economic and environmental indicators.

Planner 1: « This scenario (i.e., reference scenario) is closer to the safety/health indicator than this one (i.e., modified scenario), it (i.e., reference scenario) is higher and closer to reality. On the other hand, this scenario (i.e., reference scenario) is not good in terms of economic and environmental indicators. The indicator is higher than the other scenario (i.e., modified scenario), although it is the same delivery route. Moreover, this delivery route is not very long.

Planner 2: « This scenario (i.e., reference scenario) is more coherent in terms of health/safety indicators than the other scenario. The economic and environmental indicators are acceptable for this scenario (i.e., reference scenario) »

Planner 3: « This scenario (i.e., reference scenario) is good. The other scenario had a health/safety indicator that is too low, although it is the same delivery route in both scenarios. It's difficult to judge economic and environmental indicators ».

With a reference scenario assessed as having a medium to the medium-low level of quality (score of 3 and 2), the change in indicators of -40% is considered to provide overall higher quality. However, as with the -20% modification, it appears that the indicators for the reference route are not considered reliable. Moreover, a compromise in "balance/cancellation" of deficiencies emerged for planner 1.

Planner 1: « For this scenario (i.e., reference scenario), the health/safety indicator is too high because the delivery route is medium, not difficult. The health/safety indicator of the other scenario (i.e., modified scenario) should be higher because the delivery route is located very away and performed with a thermal vehicle. I preferred this scenario (i.e., reference scenario) for economic and environmental indicators. Both scenarios are imperfect, so I gave them the same score ».

Planner 2: « For this scenario (i.e., reference tour), the safety/health indicator is too high. It's a medium tour. The environment and economic indicators are better in this scenario (i.e., modified scenario) than the other because for the other (i.e., reference scenario), the indicators are too high.»

Planner 3: « The health/safety indicator is too high for this scenario (i.e., reference scenario), especially when I compare this one with all the scenarios that I have seen previously.»

With a reference scenario in which scores corresponded to a good quality level (scores of 4 and 3), the modification of indicators of -50% was considered to be of somewhat poorer quality, except for one of the planners.

Planners saw the differences between the scenarios and seemed to make compromises seem in favor of an indicator that they perceived as reliable.

Planner 1: « For this scenario (i.e., reference scenario), the safety/health indicator is too high. The economic and environmental indicators are coherent. For this scenario (i.e., modified scenario), economic and environmental indicators should be higher due to the distance, the time, and the energy of the vehicle (gas).»

Planner 2: « For this scenario (i.e., modified scenario), according to the safety/health indicator, the delivery tour is medium (in terms of difficulty), so it's good. Concerning environment and economic indicators, they seem to be acceptable. For the other scenario (i.e., reference scenario), all the indicators are too high. »

Planner 3: « The safety/health indicator is better in this scenario (i.e., reference scenario) than in the other, even if I found the safety/health too high.»

With a reference scenario assessed as having a medium to good quality level (scores of 3 and 4), all planners considered the modification in indicators of -60% to be of lower quality. Moreover, the difference between reference and the modified scenario was the largest of all the comparisons in this case, which can be explained by inconsistency in the cost representations of the planners:

Planner 1: « This scenario (i.e., modified scenario) is not acceptable at all. The indicators are too high for the three dimensions. Nothing is coherent. »

Planner 2: « In this scenario (i.e., modified scenario), the health/safety indicator is not high enough. It's similar to economic and environmental indicators. It's a thermal vehicle. The indicators should be higher. »

Planner 3: « This scenario (i.e., modified scenario) is not high enough for all the indicators. It's a difficult delivery route. It's the same driver for each delivery tour, we don't recommend this delivery route to anyone, it's away, and the location is specific and sensitive... »

Finally, it should be noted that (1) the planners' evaluations were the result of a compromise, (2) one of the planners used only a part of the indicators to construct evaluation, and (3) reliable information (i.e., reference) was not always considered as reliable.

Second company

As a reminder, we chose to apply overestimates and underestimates that could be significant in the second company. For this reason, we called reference scenario RI-R, the scenario in which indicators were closer to the "reference indicator" based on various elements related to the cost of planning for each dimension. We called the "modified scenario" RI-M, which displayed indicators that were further away from the reference scenario.

Table 7 displays the score given by the planner for each reference and modified scenario. In the second company, the modified scenario was "assessed better" than the reference scenario, 6 times out of 15. Concerning the first company, one interpretation is that the modified scenario was "assessed better" because the delivery route was identical but cheaper. Hence, the planner favored the indicator that decreased the cost of the delivery route.

Table 7. The score given by each planner in the second company for each reference scenario (RI-R) and modified scenario (RI-M)

Scenario	Planner	Reference scenario	Modified scenario
RI1-R (1%) / RI1-M (-40%)	1	3	4
	2	3	4
	3	2	2
RI2-R (-32%) / RI2-M (35%)	1	3	4
	2	1	3
	3	3	4
RI3-R (-20%) / RI3-M(100%)	1	3	4
	2	3	5
	3	4	3
RI4-R (50%) / RI4-M (118%)	1	4	3
	2	4	3
	3	4	3
RI5-R (161%) / RI5-M (227%)	1	4	3
	2	4	3
	3	4	3

Note. The name of the scene corresponds to the number of scenario R (routes) I (indicator) "X" for reference "R" or modified "M" scenario. For each comparison, the scenario with the high score is colored in orange when a high score is given to the referential scenario and red when a high score is given to the modified scenario. Reference and modified scenarios that received the same score are colored in purple.

When the indicator was close to the referential scenario (1%), and the modified scenario was underestimated by 40%, planners gave a higher score (score of 4) to the modified scenario. Only planner 3 did not give a higher score (score of 2 for both scenarios). However, according to their explanations, all the planners found the modified scenario better than the reference scenario. Here, the delivery tour was considered to be "easy," so the planners favored the scenario with underestimation.

Planner 1: "If I compare the two scenarios, this scenario (i.e., reference scenario) is higher than the other scenario (i.e., modified scenario) even though it's an easy delivery route... so this scenario (i.e., modified scenario) is more reliable."

Planner 2: "This scenario (i.e., modified scenario) is better physically and emotionally because indicators are lower, and it's an easy delivery route ... so it takes the delivery route with the lowest numbers."

Planner 3: "This scenario (i.e., modified scenario) is more reliable because it's an easy delivery route, so the lower number is better. concerning economic and environmental indicators, they are complicated to define without a reference... but rather this scenario (i.e., modified scenario) is lower."

When modifications of information were proportionately the same (30%) but with underestimation compared to overestimation, planners considered the scenario in which indicators were underestimated as better (score of 3 and 4). According to their explanations, the delivery route of the present scenario was challenging, so the planners favored the scenario with overestimated information.

Planner 1: "For this scenario (i.e., referential scenario), the health/safety indicator is low."

Planner 2: "This scenario (i.e., modified scenario) is better because the indicator is higher. The other scenario (i.e., reference scenario) is too low. It's difficult to judge environmental and economic indicators because the delivery tour is the same, and I don't have a reference for the number."

Planner 3: "This scenario (i.e., reference scenario) should be higher. This delivery tour is more difficult; it's early in the morning, a tailgate is needed, and the orders are fragile."

For the third comparison, all planners except for one assessed the reference scenario as a low level of quality (score of 3) compared to the modified scenario of a good quality level (score of 3 and 4). However, explanations showed that this score was also based on the planners' representation of the delivery routes. Two planners favored the overestimation because they considered the delivery route as more difficult. One planner also noted the underestimation of indicators but favored the scenario with indicators close to the reference.

Planner 1: "This scenario (i.e., modified scenario) is more reliable than the other. It's a route where you have to get up early, so you have to hold on all day."

Planner 2: "This scenario (i.e., reference scenario) is too low. It's a night delivery route, so it's more reliable with high indicators. For economic and environmental indicators, I favored the scenario with low indicators; it's a 3-hour delivery route and 57 kilometers."

Planner 3: "This scenario (i.e., reference scenario) is better even if the safety/health cost is too low. All indicators should be increased."

For the fourth comparison with two overestimations (50 and 118%), planners gave a higher score to the scenario in which the indicators were closer to the reference (a score of 4

compared to 3). According to their explanations, we can notice that planners acknowledged the overestimation of the modified scenario. We also noted that planner 2 explained that the difference between environmental and economic indicators is not discriminating than safety/health indicators. Indeed, the number of safety/health is significant, which implied more differences with overestimated/ underestimation.

Planner 1: “I favored this scenario (i.e., reference scenario) because indicators are low compared to the other scenario.”

Planner 2: “this scenario (i.e., reference scenario) is better because indicators are low. The difference between environmental indicators is low on both scenarios.”

Planner 3: “314... (i.e., modified scenario) is too high for safety/healthy indicator.”

When two overestimations are applied to information, but with a more difference (161 and 227%), planners give a higher score to the referential scenario. According to their explanations, all planners noted the overestimation of the modified scenario. Moreover, the planners’ representation of the difficulty of the delivery route influenced the score given to each scenario.

Planner 1: “This scenario (i.e., reference scenario) is more reliable than the other because it is less high on economic and environmental indicators: it is a short route of only a few kilometers and a gas vehicle. For the healthy/safety indicator, it’s similar: it’s not a difficult route, so if I compare it to this scenario (i.e., modified scenario), which is higher, I prefer the lower one.”

Planner 2: “It’s not a difficult delivery route, so I favor the scenario that is low.”

Planner 3: “This scenario (i.e., reference scenario) is better because all indicators are low. It’s an easy delivery tour. For economic indicators, it depends on the reference, but this scenario (i.e., reference scenario) is better because it is lower.”

Finally, it should be noted that as in condition 1, (1) the indicators were considered reliable as long as they resembled the planners’ representation of the perceived difficulty of the delivery route and (2) this representation influenced the score given, regardless of underestimation or overestimation.

Discussion

The purpose of this study was to examine the reliability and transparency of a decision support system in the road freight transport sector. We confronted participants with different solutions proposed by the decision support system containing high transparency, condition 1, and low transparency, condition 2. To investigate the behavior of the participants reliably, we

confronted expert planners in each phase with two systematic propositions of the decision support system in which one scenario displayed less reliable information compared to the other scenario.

We proposed two hypotheses. (1) Reliable information, whether high or low transparency (raw or aggregated information), should be better perceived and assessed compared to information with a lower level of reliability. (2) Information with a high level of transparency should be evaluated better than information with a low level of transparency.

Our first hypothesis was confirmed. In both companies, the planners gave a higher score to the reference scenario than the modified scenario. In the first phase, the planners in both companies seldom gave a higher score to a modified scenario. This result implied that planners are sensitive to the reliability of the decision support system, which is in line with studies on reliability in the human-machine interaction domain (Bagheri & Jamieson, 2004; Parasuraman et al., 1993).

Lee and Moray (1994) found that participants used the automated system only when their trust in themselves was lower than their trust in automation. The planners never gave the maximum score to the reference scenario, even when they had no inconsistencies to cite in the scenario. Sheridan (2002) and Parasuraman and Riley (1997) explained that operators tend to distrust new alarms or decision support systems until such automated systems have proven themselves. Madhavan and Wiegmann (2007) highlighted the need to conduct studies where participants are explicitly informed that the automated system is “new or unproven” in terms of its reliability. In our study, planners knew that the decision support system was new and in a prototype phase, so potentially unreliable. We can suppose that this information may have influenced their evaluation of the system and their trust in it. It would be interesting to study the real use of scenarios longitudinally when implementing such software. It would be interesting to assess differences in the use of such programs (depending on the reliability of the system) based on the expertise of the operators. Cegarra and Van Wezel (2012) highlighted that the perceived usefulness of the system influences the use of a decision support system.

For this reason, comparing the results of the evaluations of the scenarios and their real use will be interesting, especially when using unreliable scenarios. However, it should be noted that whether utilizing a reliable or unreliable scenario, both must also integrate the context of the situation. As explained in the introduction, planning involves many factors (Khademi, 2016). For instance, planners can use an incorrect scenario because they recognize a “margin

of time” on this specific day. Therefore it does not matter whether the time between sites is incorrect, for example. In this study, and with the explanations of planners, we showed that the assessment of a scenario is based on a compromise, especially in condition 2. In the presence of three indicators, planners tried to balance the deficiencies of each indicator.

Our second hypothesis was also confirmed. We found that information with a high level of transparency is better evaluated than information with low transparency (indicators). In the first company, the modified scenario with a high level of transparency received a higher score only once of fifteen comparisons compared to five out of 15 comparisons when the transparency was low. Similarly, in the second company, a modified scenario with high transparency did not receive a higher score at all compared to 7/15 when the transparency was low. However, we can also notice that with high transparency, more scenarios were considered equal with the same score (5 for the first company and 3 for the second company) compared to low transparency (3 for the first company and 1 for the second company). These results showed that the score is more discriminating with the high transparency: planners give a higher score to the modified scenario more often with the low transparency. We can suppose that the aggregation of the information in the low transparency condition leads the operators to score based on their representation. The compromises planners made in this representation led to frequently assigning a higher score to a less reliable scenario. In condition 1, with reference indicators, the planners did not give a higher score systematically to the reference indicators calculated based on various elements. This suggests that for a decision support system in the RFT sector, displayed information aggregated leads to more difficulty to recognize unreliable information.

We also noted that in their explanations, participants were more likely to use the health/safety indicator to evaluate the scenario. According to the calculation of the healthy/safety cost, the indicator is a more bigger number compared to economic and environmental indicator. So, with underestimation or overestimation, the variation of the number (i.e., the cost) of the healthy/safety indicator leads to greater differences for the cost compared to the other indicators (i.e., environmental and economic). It would be interesting to repeat the same experiment with two modified and one reference scenario by systematically presenting underestimated, reference, and overestimated indicators to see if planners will evaluate scenarios with less expensive indicators better if they think the route of the scenario is straightforward and evaluate routes with expensive high indicators better if they think the route is challenging.

Research has shown that increasing the transparency of automation might help participants assess the reliability of automation. For example, providing operators with explicit reliability information leads to faster and more accurate trust calibration (Lee & Moray, 1994; Wang, Jamieson, & Hollands, 2009). In our study, we can suppose that indicators' lack of transparency induced operators to evaluate the reliability of the indicators based on their perceptions. Even if the planners had participated in the calculation of these indicators, in practice, the indicators need to be displayed with a scale which would serve as a reference. That is to say, the position of the indicator on a scale should include a minimum and maximum based on the calculation of all others indicators of the others planning proposed by the decision support system. This is an essential point because multiple studies have stated the positive effect of transparency on trust and operators' perceived reliability of an automated system (Hussein et al., 2020; Roundtree et al., 2019;).

Overall, this study showed that planners give a higher score to scenarios displaying reliable information compared to less reliable scenarios in which information is modified. Their assessments were more adequate and easier when given information with high transparency than with low transparency. With low transparency, planners based their assessment of the scenario on their representation. Additionally, planners make a compromise when they assessed different scenarios, which explains differences in scores between planners.

This study highlighted the importance to better understand the reliability of information according to the different levels of transparency, especially in the planning task. Future studies on the use of such automation of information are needed.

Reference

- Bagheri, N., & Jamieson, G. A. (2004). The impact of context-related reliability on automation failure detection and scanning behaviour. *IEEE International Conference on Systems, Man and Cybernetics*, 1, 212–217.
- Bass, E. J., & Pritchett, A. R. (2008). Human-automated judge learning: A methodology for examining human interaction with information analysis automation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 38(4), 759-776.
- Cegarra, J. (2008). A cognitive typology of scheduling situations: A contribution to laboratory and field studies. *Theoretical Issues in Ergonomics Science*, 9(3), 201-222.
- Cegarra, J. (2012). *De la gestion de la complexité à son assistance: Contributions en psychologie ergonomique*. (Unpublished manuscript, Habilitation à diriger des recherches, Toulouse II-Le Mirail University)

- Cegarra, J., & van Wezel, W. (2012). Revisiting decision support systems for cognitive readiness: A contribution to unstructured and complex scheduling situations. *Journal of Cognitive Engineering and Decision Making*, 6(3), 299-324.
- Chen, J. Y., Procci, K., Boyce, M., Wright, J., Garcia, A., & Barnes, M. (2014). *Situation awareness-based agent transparency* (No. ARL-TR-6905). Army research lab aberdeen proving ground md human research and engineering directorate.
- Endsley, M. R. (2017). From here to autonomy: lessons learned from human–automation research. *Human factors*, 59(1), 5-27.
- Gary, K., Uzsoy, R., Smith, S. P., & Kempf, K. (1995). Measuring the quality of manufacturing schedules. In *Intelligent scheduling systems* (pp. 129-154). Springer, Boston, MA.
- Govaere, V., & Wioland, L. (2018b). Modélisation des contraintes et des compromis spécifiques au domaine du transport routier de marchandises (TRM). Livrable 2 du projet ANR Smart Planning ANR-16-CE26-0017.
- Govaere, V., Wioland, L., Cegarra, J., Gourc, D., & Clément, A. (2018, August). Smart Planning-Approaching the Characteristics of a Valid, Balanced Transport Round. In *Congress of the International Ergonomics Association* (pp. 387-396). Springer, Cham.
- Guerin, C., Hoc, J. M., & Mebarki, N. (2012). The nature of expertise in industrial scheduling: Strategic and tactical processes, constraint and object management. *International Journal of Industrial Ergonomics*, 42(5), 457-468
- Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors*, 57(3), 407-434.
- Hussein, A., Elsayah, S., & Abbass, H. A. (2020). The reliability and transparency bases of trust in human-swarm interaction: principles and implications. *Ergonomics*, 1-19.
- Kerr, R. M. (1992). Expert systems in production scheduling: Lessons from a failed implementation. *Journal of Systems and Software*, 19(2), 123-130.
- Khademi, K. (2016). *Les processus cognitifs dans les activités d'ordonnement en environnement incertain*. (Doctoral dissertation, Toulouse Jean Jaures II University).
- Khademi, K., Cegarra, J., Govaere, V., & Wioland, L. (2015). *La planification par l'exploitant dans le transport routier de marchandises: méthode d'analyse et perspectives d'application inter-domaines. Partie 2*. (Report). Institut National de Recherche et de Sécurité (INRS). <https://hal-lara.archives-ouvertes.fr/hal-01428640/document>
- Larco, J. A., Fransoo, J. C., & Wiers, V. C. (2018). Scheduling the scheduling task: a time-management perspective on scheduling. *Cognition, Technology & Work*, 20(1), 1-10.

- 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*, 46(1), 50-80.
- Lerch, F. J., Prietula, M. J., & Kulik, C. T. (1997). The Turing effect: The nature of trust in expert systems advice. In *Expertise in context: Human and machine* (pp. 417-448).
- Madhavan, P., & Wiegmann, D. A. (2007). Effects of information source, pedigree, and reliability on operator interaction with decision support systems. *Human Factors*, 49(5), 773-785
- Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history based trust in human-automation interactions. *Human Factors*, 50, 194- 210.
- Oakley, B., Mouloua, M., & Hancock, P. (2003, October). Effects of automation reliability on human monitoring performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 47, No. 1, pp. 188-190). Sage CA: Los Angeles, CA: SAGE Publications.
- Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human factors*, 56(3), 476-488.
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation : An attentional integration. *Human factors*, 52(3), 381-410.
- 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*, 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
- Roundtree, K. A., Goodrich, M. A., & Adams, J. A. (2019). Transparency: Transitioning from human–machine systems to human-swarm systems. *Journal of Cognitive Engineering and Decision Making*, 13(3), 171-195.
- Rovira, E., McGarry, K., & Parasuraman, R. (2007). Effects of imperfect automation on decision making in a simulated command and control task. *Human factors*, 49(1), 76-87.
- Sarter, N. B., & Schroeder, B. (2001). Supporting decision making and action selection under time pressure and uncertainty: The case of in-flight icing. *Human factors*, 43(4), 573-583

- Sheridan TB (2002). *Humans and automation: system design and research issues*. Wiley, New York
- Van Wezel, W., Cegarra, J., & Hoc, J.-M. (2010). *Allocating Functions to Human and Algorithm in Scheduling*. *Behavioral Operations in Planning and Scheduling*, 339–370. doi:10.1007/978-3-642-13382-4_14
- Van Wezel, W., & Jorna, R. (2009). Cognition, tasks and planning: supporting the planning of shunting operations at the Netherlands Railways. *Cognition, Technology & Work*, 11(2), 165-176.
- Wang, L., Jamieson, G. A., & Hollands, J. G. (2009). Trust and reliance on an automated combat identification system. *Human factors*, 51(3), 281-291.
- Westin, C., Borst, C., & Hilburn, B. (2016). Automation transparency and personalized decision support: air traffic controller interaction with a resolution advisory system. *IFAC-PapersOnLine*, 49(19), 201-206.
- Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science*, 8(3), 201-212.
- Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomics Science*, 2(4), 352-367.

DISCUSSION GENERALE

La psychologie ergonomique est parfois vue comme une discipline appliquée. Or, en développant nos connaissances sur les effets de la fiabilité des systèmes automatisés nous montrons qu'il existe tout un corps de connaissance qu'il est essentiel de développer. Ce travail de doctorat avait pour objectif d'approfondir les connaissances sur les effets de la fiabilité des systèmes automatisés en se concentrant sur l'approche des fonctions et niveaux d'automatisation proposée par Parasuraman, Sheridan et Wickens (2000). Plus précisément, il s'agissait de mettre en évidence les effets de la fiabilité appliquée à la fonction information des systèmes automatisés sur les performances humaines et le comportement humain. Les études focalisées sur la fonction information en cas de problèmes de fiabilité étant moins nombreuses dans la littérature, nous avons décidé de combler ce manque de connaissances en concentrant majoritairement cette thèse sur cette première fonction du modèle. Par ailleurs, nos travaux soulignent que développer des connaissances fondamentales peut se faire en combinant des études en laboratoire et sur le terrain car la psychologie ergonomique met plus en avant l'aspect naturel des situations que la réalisation d'études uniquement sur le terrain.

Résumé des résultats expérimentaux

Les résultats sont discutés à travers le modèle intégré du contentement et des biais d'automatisation proposé par Parasuraman et Manzey (2010) et modifié dans la figure 15.

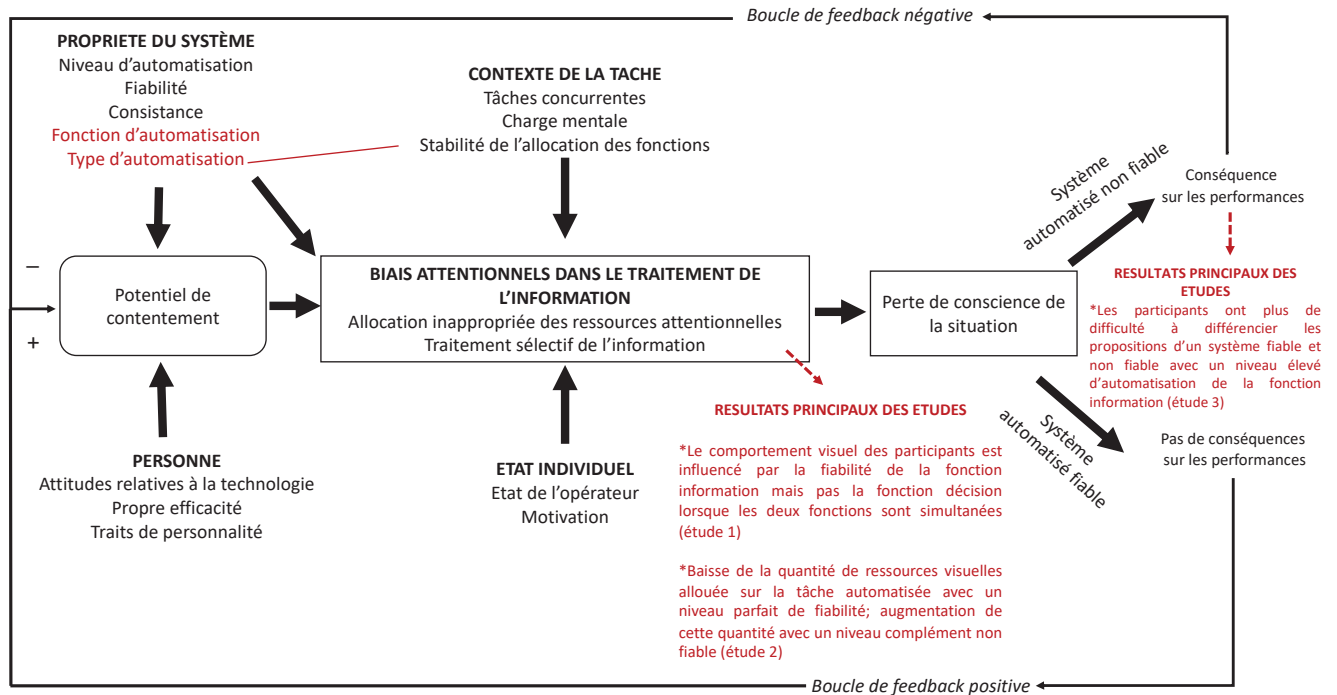


Figure 15. Principaux résultats appliqués sur le modèle adapté de Parasuraman et Manzey (2010).

Ce modèle est construit autour d'une notion centrale : l'attention. Dans nos études, nous avons considéré que la quantité de ressources visuelles, évaluée par la somme des durées de fixations étaient directement liées à l'attention visuelle des opérateurs humains. Ce modèle est donc particulièrement intéressant pour discuter de nos résultats, d'autant plus qu'il intègre deux notions associés à l'interaction homme-machine souvent considérées comme différentes alors que des points communs les réunissent : les biais d'automatisation et le contentement. Ces deux phénomènes résultent d'une interaction complexe entre des facteurs individuels, situationnels et liés au système automatisé. Ces phénomènes se traduisent par un retrait ou une réaffectation des ressources attentionnelles. Ces effets attentionnels (conscients ou inconscients) de l'opérateur humain sont liés à une confiance excessive envers un système automatisé considéré comme non défaillant (Parasuraman & Manzey, 2010). Plus précisément, le modèle suppose que la conséquence immédiate sur les performances d'un traitement sélectif ou moins attentif de l'information est une perte de conscience de la situation. Tant que le

système automatisé n'est pas défaillant, il n'y a pas de conséquences. Cependant, en cas de défaillances (non fiabilité du système) cela entraîne des conséquences sur la performance (tels que des erreurs d'omission et/ou de commission⁷). Ces deux types d'erreurs sont présentes dans la littérature sur les biais d'automatisation et sont étroitement liés au phénomène de contentement. Par exemple, l'erreur d'omission qui correspond au fait que l'opérateur humain ne réagisse pas à une situation nécessitant pourtant une réaction car le système ne l'a pas indiqué ressemble définitivement à la définition du phénomène de contentement : une insuffisance de surveillance du système automatisé par l'opérateur humain. Cette insuffisance de surveillance a été un résultat important de notre première étude :

Le premier chapitre empirique avait pour objectif de tester les effets de la fiabilité d'un système automatisé mettant en œuvre simultanément la fonction information et la fonction décision. En effet, la plupart des études présentes dans la littérature testent les conséquences de la fiabilité de chaque fonction mais non leur interaction directe (Sarter & Schroeder, 2001 ; Rovira, McGarry, & Parasuraman, 2007 ; Rovira, Cross, Leitch, & Bonaceto, 2014). Le manque flagrant d'études testant les effets de la fiabilité avec l'interaction de deux fonctions, nous a amené à tester l'influence possible d'une fonction sur l'autre. Ce choix s'explique par le fait qu'un système peut mettre en place simultanément les deux fonctions et n'est pas limité qu'à la prise en charge d'une seule fonction à la fois. Connaître les effets de différents niveaux de fiabilité sur les deux fonctions croisées simultanément était donc important. Dans cette étude, les participants ont dû gérer une batterie de tâche composée de trois tâches dont une tâche était automatisée selon deux fonctions (information vs décision) et selon quatre niveaux

⁷ Les erreurs d'omission correspondent à la non réaction d'un opérateur humain à une situation critique en lien avec un système d'alarme. Par exemple, un opérateur humain manque la sortie sur l'autoroute car le système de navigation ne l'a pas averti. Les erreurs de commission correspondent au fait que l'opérateur humain suit une indication du système automatisé qui est incorrecte. Par exemple, s'engager dans une rue à sens unique du mauvais côté car le système automatisé lui dit de le faire (Parasuraman & Manzey, 2010).

de fiabilité pour chaque fonction. Le peu d'études étudiant spécifiquement plusieurs fonctions plutôt liées à l'information et plutôt liées à la décision avec des niveaux de fiabilité différentes sont généralement effectuées sur des environnements de simulation très spécifiques à différents terrains avec des experts de ces mêmes terrains (Cf. des pilotes confrontés à un simulateur de givrage en plein vol ; Sarter et Schroeder, 2001 ; des cadets (grade militaire) face à une simulation de reconnaissance et de renseignement militaire). Notre choix d'utiliser le MATB permettait une généralisation des résultats dans un environnement contrôlé. La nature même de la simulation composée de plusieurs tâches distinctes permettait de mettre en lumière de possibles stratégies de priorisation des participants. Même si cette tâche est avant tout un micro-monde du pilotage d'avion, sa simplification permettait une utilisation par tout type de sujets. Cet avantage peut tout de même être aussi considéré comme une limite dans le sens où la simplification de la tâche de pilotage et les sujets non experts peuvent questionner une généralisation à tout type d'expertise.

Nous avons observé que le niveau de fiabilité de la fonction information a principalement influencé le comportement visuel des participants, à l'inverse de la fonction décision : la quantité de ressources visuelles des participants sur la tâche automatisée est largement influencée par la fonction information; et en cas de non fiabilité totale de ce système automatisé de l'information (défaillances non signalées), les participants ont tendance à ne pas détecter les corrections à effectuer. Notre étude, en simplifiant les deux premières fonctions information et les deux dernières de décision a permis de mettre en lumière l'importance des premières fonctions par rapport aux dernières dans les choix stratégiques visuelles des participants. Au final, notre première étude nous a amené à rajouter la notion de *fonction d'automatisation* au modèle proposé par Parasuraman et Manzey (2010) puisque nous avons montré que ces fonctions jouaient un rôle sur la quantité de ressources visuelles allouées par les participants (figure 15).

Nous avons également rajouté la notion de *type d'automatisation* dans la partie propriété du système automatisé suite aux résultats de notre deuxième étude (figure 15). Cette notion est déjà présente dans le modèle de Parasuraman et Manzey (2010) à travers la stabilité de *l'allocation des fonctions* dans la partie contexte de la tâche. Cependant, *le type d'automatisation* réfère directement à un choix de conception du système automatisé alors que la notion de stabilité évoque déjà une description plus spécifique de distribution des fonctions entre l'opérateur humain et le système automatisé. Cette notion de *type d'automatisation* a été au centre de notre deuxième étude.

Dans le deuxième chapitre empirique, nous avons testé les effets deux types de système automatisé : adaptatif ou statique, caractérisé par la présence ou non d'une reprise en main temporaire de la tâche automatisée par les participants mais cette fois-ci sur une plus longue durée de supervision du système automatisé. De façon générale, notre étude a montré des résultats similaires entre la présence d'une reprise en main ou non : la quantité de ressources visuelles allouée à la tâche automatisée à diminuer au cours du temps avec un système automatisé parfaitement fiable et a augmenté avec un système automatisé dont la fiabilité était nulle. De même que Parasuraman, Mouloua et Molloy (1996), la reprise en main a influencé le comportement des participants qui ont continué à maintenir leur quantité d'attention visuelle durant les sessions automatisées suivant directement la session temporaire de reprise en main manuelle. Un autre résultat important de cette étude a montré que la quantité de ressources visuelles accordées lors des sessions manuelles était plus longue après des sessions automatisées que durant des sessions non automatisées. Pour le niveau de intermédiaire, le temps passé à examiner la tâche automatisée est resté élevé après la session manuelle dans les sessions automatisées suivantes. Cela indique qu'avec un niveau intermédiaire de fiabilité, surveiller la tâche automatisée nécessitait la même implication visuelle qu'effectuer entièrement manuellement la tâche durant une longue période.

Après avoir testé les effets de la fiabilité sur un système automatisé mettant en œuvre la fonction information en laboratoire de manière contrôlée, nous avons décidé d’approfondir les connaissances sur une situation plus écologique dans le troisième chapitre empirique. L’activité de planification offrait la possibilité d’en apprendre davantage sur un système automatisé mettant en œuvre la fonction information puisque dans cette activité, le planificateur s’appuie sur de multiples informations pour prendre des décisions (Khademi, 2016). Ces informations qui sont à l’origine des choix effectués par l’exploitant peuvent être perçues de façons différentes par plusieurs planificateurs. Étudier ces informations lorsqu’elles étaient dispensées par un système automatisé dans une activité de planification et dans un contexte de fiabilité remise en question constituait une nouveauté dans la littérature. Nous avons observé que lorsqu’ils étaient confrontés à des propositions fiables et moins fiables du système automatisé, les exploitants jugeaient de meilleure qualité les propositions fiables que celles moins fiables. Nous avons également montré qu’un niveau plus élevé d’agrégation de l’information entraînait plus de difficultés à distinguer les propositions fiables et non fiables du système automatisé par les exploitants. Ce dernier résultat souligne l’importance de la transparence des informations pour une meilleure évaluation de la fiabilité du système. Il vient donc s’ajouter à d’autres études montrant l’importance des niveaux de transparence (Chen et al., 2014 ; Alonso & De la Puente, 2018) et fournit une première étude exploratoire sur la transparence des informations du système automatisé dans le secteur particulier du transport routier de marchandises et de l’activité de planification.

Apports de cette thèse

La fiabilité appliquée à la fonction information

Dans cette thèse, nous étudions la fiabilité à travers la prise en charge par le système automatisé de deux types d’informations : une information sous forme de signe que nous

illustrons à travers nos deux premières études, et une information sous forme de symbole à travers la troisième étude. Dans le MATB, l'information correspondait à une information donnée par le biais d'un signe présent ou non : une alarme rouge visuelle indiquait une défaillance du système. Cette alarme était fiable ou pas selon son état (allumé ou non). Dans la situation de planification, l'information était présente sous forme textuelle. Les informations étaient fiables si le renseignement qu'elle fournissait était valide. Cette validité pouvait être perçue plus ou moins différemment par les exploitants.

Même si les deux types d'informations étaient quoi qu'il arrive bien relatif à la fonction information tel que proposé par le modèle de Parasuraman, Sheridan et Wickens (2000), la différence de nature de l'information entre les études a engendré des méthodes différentes dans les études impliquant du MATB et impliquant l'activité de planification. En fin de compte, nous englobons ces deux visions sous le chapeau de la notion de « la fiabilité de l'automatisation de l'information » du système automatisé dans l'interaction homme-machine. Ces deux formes de l'information sont directement en lien avec la nature même des tâches effectuées. En effet, dans notre troisième étude la fiabilité du système automatisé était étudié à travers de l'information textuelle (symbole). La particularité de l'activité de planification fait que ce qui est considéré comme une défaillance du système (affichage d'une information erronée) peut être considérée comme non défaillant par un autre, notamment parce que l'information manipulée est d'ordre qualitatif : derrière l'information, il y a un sens subjectif pour l'opérateur expert.

A noter que dans les études d'interaction homme-machine, la perception de la fiabilité d'une aide automatisée peut être demandée aux participants. Cette fiabilité perçue se réfère à la perception d'un niveau de fiabilité basé sur un nombre de défaillances. Par exemple, Onnasch (2015) module sa fiabilité par le nombre de défaillances en utilisant aussi le MATB. Elle demande aux participants à la fin de l'expérimentation d'évaluer la fiabilité perçue du système automatisé.

Il aurait été intéressant pour nous de poser la même question à nos participants pour les deux études avec le MATB. Nous aurions pu étudier la fiabilité perçue des participants sur une fiabilité définie : sur la tâche de MATB, la fiabilité est définie par un nombre strict de défaillances et ne peut pas être remise en cause. Nous aurions pu mettre en lien cette fiabilité perçue et nos résultats : si la fiabilité perçue du système automatisé de l'information était vue comme élevée, alors nous pouvons supposer que les participants allaient moins surveiller la tâche automatisée.

Dans l'activité de planification, la fiabilité perçue du système automatisé se construit aussi sur le nombre d'informations erronées présentées dans le planning dans la première phase. Cependant, chaque information est considérée comme plus ou moins valide. Ce prorata d'informations considérées comme valides ou non permet de créer une fiabilité perçue de chaque proposition de planning. Plus les informations étaient considérées comme valides, plus la fiabilité perçue de l'aide était meilleure, meilleur était le score attribué au scénario.

Dans la seconde phase, la fiabilité perçue du système s'est construit à travers les indicateurs. Un indicateur prenait la forme d'un nombre. La fiabilité perçue pouvait donc être construite sur une quantité, un écart entre un nombre considéré comme valide et non valide. Plus cet écart était considéré comme grand, plus il était non valide, plus la fiabilité perçue du système était mauvaise.

Deux approches méthodologiques complémentaires

Cette thèse implique deux approches méthodologiques différentes : la première s'applique à des études plutôt contrôlées en laboratoire, effectuée à travers un micro-monde de pilotage aéronautique et impliquant des participants novices. La seconde s'applique à une étude plus spécifique à l'activité de planification avec des participants experts dans une tâche d'évaluation de scénarios. Le choix de ces deux approches dans deux activités différentes peut poser question mais elles sont pourtant complémentaires, notamment dans une thèse en ergonomie. L'ergonomie est un domaine qui trouve son origine à travers l'étude des situations

de travail. Le champ de la psychologie lui est lié, ce qui lui vaut parfois le nom de psychologie ergonomique. La psychologie ergonomique est donc la partie de l'ergonomie qui cherche à étudier les composantes mentales de l'activité humaine (Navarro, 2018). Elle défend l'intérêt de l'étude des situations naturelles mais n'en oublie pas pour autant la nécessité de connaissances fondamentales (Cegarra, 2012). Elle cherche des réponses concrètes à des problèmes sociétaux avec des applications directes, mais cherche également à produire des connaissances scientifiques sur le fonctionnement cognitif des opérateurs humains face à ces mêmes problèmes. Deux préoccupations sont au centre de la psychologie ergonomique : la nécessité pour le scientifique de respecter une validité interne en garantissant que les résultats observés soient explicables par la théorie, sous les contraintes de méthodologie expérimentale très contrôlée. Mais aussi, à la fois, de préserver une validité externe, en permettant de garder la possibilité d'une généralisation de ces résultats de laboratoire à un plus large éventail de situations (Cegarra, 2012). Or, il y a parfois des situations de travail où la généralisation de ces résultats relève du cas particulier⁸.

Nos deux premières études nous ont permis d'obtenir des résultats sur le comportement d'opérateurs humains face à un système automatisé prenant en charge la fonction information de façon plus ou moins fiable. Ces résultats ont été obtenus en laboratoire dans une situation contrôlée. Suite à cela, il était intéressant pour nous de comprendre ce que pouvait engendrer une fiabilité remise en question de la fonction information d'un système automatisé dans une situation écologique sur le terrain, et notamment dans le cas de l'activité de planification qui est particulière comme le montre les

⁸ Cegarra (2012) cite le cas des apprentissages moteur comme exemple. Dans son exemple, une théorie robuste peut être remise en question pour certaines situations particulières où elle ne sera pas forcément applicable.

difficultés rencontrées dans la conception de système automatisé à cause de l'expertise des opérateurs (Cegarra & Van Wezel, 2012).

Afin de concevoir un système automatisé tel qu'un système d'aide à la planification, il est nécessaire d'explorer les processus cognitifs étant impliqués dans son implémentation. L'intégration d'un système d'aide à la planification implique l'interaction d'un opérateur humain avec un nouveau système automatisé. Nous avons décidé d'interroger les processus cognitifs existants avec deux premières études impliquant le MATB en nous focalisant sur les la fonction information. Puis, nous avons décidé d'approfondir nos connaissances par le biais d'évaluation de scénario dans une dernière étude mettant en pratique les principales fonctions envisagées pour l'outil d'aide à la planification. La figure 16 représente la démarche que nous avons entreprise dans ce processus de conception. Dans nos premières études (partie gauche), nous avons exploré la notion de fiabilité appliquée à la fonction information. Ces études nous ont permis d'émettre des hypothèses sur le comportement humain interagissant avec un système automatisé (cf. tel qu'un prototype d'aide à la planification) prenant en charge la fonction information. Ensuite, nous avons exploré ces processus en mettant en place une étude utilisant les principales fonctions envisagées de l'outil (partie droite). Dans un processus de conception de système et de réflexions sur les processus cognitifs futurs impliqués dans une situation, l'alternance d'études permet d'effectuer et de préciser des hypothèses sur l'interaction des opérateurs avec ce système. Une fois ces hypothèses obtenues, elles vont être testées avec l'outil afin de pouvoir les valider (ou non), et d'en retirer des préconisations pour la conception du système final.

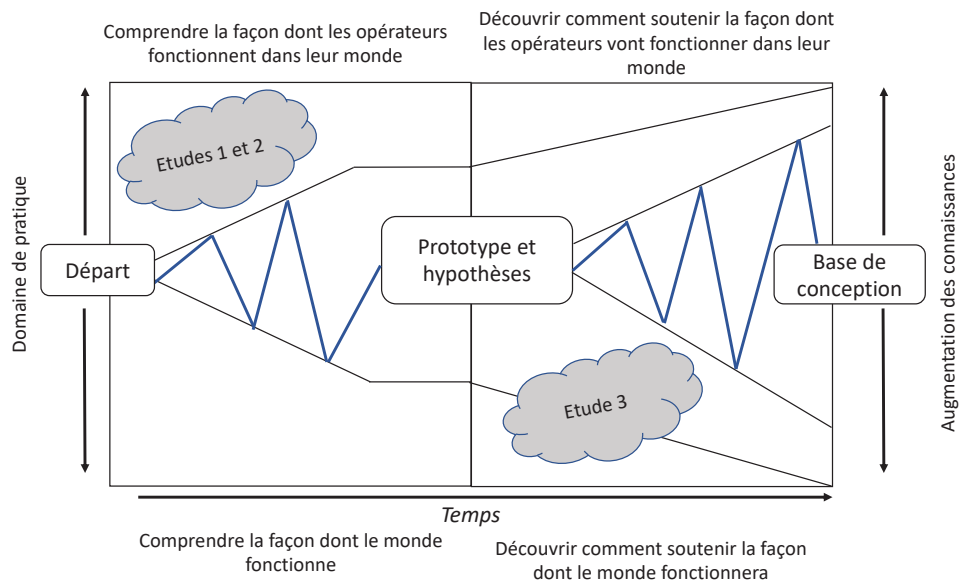


Figure 16. Vue générale d'un processus de développement de système automatisé inspiré de Potter, Roth, Woods et Helms (2000).

L'utilisation du micro-monde MATB était pour nous évidente puisqu'un grand nombre d'études sur les problèmes de fiabilité utilisent cette tâche expérimentale pour étudier l'interaction homme-machine, ce qui nous permettait la comparaison de nos résultats avec une littérature robuste sur les systèmes automatisés et la fiabilité. La possibilité et facilité technique de manipuler la fiabilité et d'ajouter l'idée des fonctions prises en charge par le système automatisé dans ces tâches a renforcé ce choix.

Le second choix d'une méthode d'évaluation de scénarios dans l'activité de planification était pour nous surtout vue comme un avantage. Nous avons considéré plusieurs éléments : un manque de connaissances flagrant dans l'étude des problèmes de fiabilité des systèmes automatisés, notamment liée à l'information, dans cette activité ; le fait indéniable que cette activité sera vouée un jour ou l'autre à s'automatiser et donc l'intérêt pour l'ergonomie prospective d'anticiper le comportement des opérateurs dans des situations d'interaction homme-machine dans cette activité ; la possibilité d'approfondir encore plus la fonction d'information à travers un vrai système automatisé prenant en charge cette fonction

de différentes façons ; la possibilité d'avoir accès à ce terrain – même pour le temps d'une seule étude – et de pouvoir échanger avec des opérateurs humains experts en planification. En effet, il peut être compliqué pour les chercheurs en ergonomie d'accéder au terrain, puisque cela consiste à extraire des opérateurs humains de leur situation de travail pendant un certain temps. Il était donc particulièrement enrichissant dans le cadre d'une thèse d'avoir cette opportunité.

- **Le logiciel OpenMATB**

Cette thèse a été l'occasion de revisiter le Multi Attribute Task Battery à travers la création et l'utilisation du OpenMATB (Cegarra, Valéry, Avril, Calmettes, & Navarro, 2020). Le OpenMATB est une version open source du Multi Attribute Task Battery qui permet la création d'expérimentations diverses grâce à la gratuité et l'accessibilité directe du logiciel. La personnalisation des différentes tâches permet la création de multiples scénarios d'expérimentations pouvant être utilisés pour un large éventail de sujets (cf. la charge mentale pouvant être modulée par le nombre de tâches utilisées et leur difficulté, différentes expérimentations utilisant des systèmes automatisés de différents types...). Il permet notamment un accès généralisé pour tous à l'intégration de dispositifs psychophysiologiques directement associés au logiciel tels que l'oculométrie. En effet, jusque-là chaque chercheur devait se « débrouiller » pour intégrer de tels dispositifs aux précédentes versions du MATB. Aujourd'hui, celui-ci a directement la possibilité d'intégrer de tels dispositifs, le OpenMATB ayant été pensé pour des expérimentations futures nécessitant de tels dispositifs. Ce dernier point est important puisqu'il a permis la collecte de mesures oculométriques durant nos deux études associées à ce logiciel ; collecte qui n'était pas possible quand on se réfère aux premières études incluant le MATB sur la fiabilité des systèmes automatisés où le comportement visuel était enregistré à travers des enregistrements vidéos par le biais de caméra (Parasuraman, Molloy, & Singh, 1993).

- **Une première étude exploratoire incluant fiabilité et prise en charge de la fonction information par un système automatisé dans l'activité de planification**

Comme expliqué dans la partie théorique de cette thèse, l'activité de planification dans le secteur du transport routier de marchandises est spécifique. L'intégration de systèmes de support à la décision dans cette activité est complexe, à cause de l'incertitude présente dans ce métier et de la complexité de la tâche même de planification dont les choix sont construits selon un certain nombre de facteurs (Cegarra & Van Wezel, 2012). Avec l'évolution des systèmes automatisés et leur expansion dans les activités professionnelles, il est fort probable que les systèmes de support à la décision finissent par être implantés dans les activités, même si ceux-ci restent à ce jour difficile à concevoir. Étant difficile et complexe à concevoir, des travaux sont encore nécessaires sur la fiabilité liée à cette fonction d'automatisation dans cette activité. Cela en fait un sujet intéressant pour des recherches sur l'interaction homme-machine. A travers notre troisième étude, nous avons voulu mener une première exploration de la mise en place de tels systèmes dans l'activité de planification. Nous nous sommes donc intéressés à la fiabilité des propositions d'un système d'aide à la planification prenant en charge la fonction information. Dans notre première vision de l'étude, les exploitants pouvaient manipuler un prototype de système d'aide à la planification en temps réel et la fiabilité des informations étaient manipulées de la même façon qu'elles le sont actuellement : à travers les indicateurs et à travers les informations présentées dans le planning. Les résultats obtenus n'en sont pas pour autant moins intéressants puisqu'ils permettent de montrer que les exploitants ont tendance à mieux évaluer les propositions fiables que celles moins fiables et que cette évaluation est moins évidente à effectuer avec des informations agrégées sous forme d'indicateurs. Ce point de départ ouvre la possibilité à d'autres recherches.

Réflexions sur les limites de chaque étude et perspectives

Dans notre premier chapitre empirique, l'interaction entre un système prenant en charge deux fonctions du système automatisé a permis de mettre en lumière l'influence

prédominante sur le comportement des participants de la fonction de l'information. Ce résultat pourrait être discuté au regard du design expérimental de l'expérience. Dans cette étude, la fiabilité de la fonction information était stable et ne changeait pas pour un même participant. A l'inverse, la fiabilité de la fonction décision changeait au cours du temps. Les participants se sont fiés à la fonction information qui était constante et ont donc réduit leur quantité de ressources visuelles sur la tâche automatisée. Ce résultat est d'autant plus intéressant qu'il met en interaction la mise en œuvre de deux fonctions prises en charge par le système automatisé dont une est variable dans sa fiabilité alors que l'autre est constante, ce qui change des études précédentes impliquant de manière indépendante une fiabilité constante ou variable de la fonction liée à la décision (Parasuraman, Molloy, & Singh, 1993, Bagheri & Jamieson, 2004 ; Parasuraman & Manzey, 2010). Ce résultat souligne que lorsque le système automatisé - prenant en charge la décision - est présent et variable, la fiabilité constante de la fonction information engendre une baisse de la quantité de ressources visuelles chez les participants au fur et à mesure que la fonction information augmentait en fiabilité. Ce pattern a également été confirmé par les résultats de notre deuxième étude puisqu'avec une fiabilité parfaite de la fonction information, la quantité de ressources visuelle accordée à la tâche automatisée a également baissé. Par ailleurs, une des limites de cette étude est le fait qu'un participant ait été confronté à toutes les conditions de la fiabilité liées à l'action (variable) et une seule de la fiabilité liée à l'information (constante). Afin de conforter nos résultats et l'implication de la fonction information, il serait donc intéressant d'approfondir les résultats de cette première étude avec une autre étude changeant cette fois-ci la variabilité des deux fonctions : le système automatisé prenant en charge l'information deviendrait variable et la fonction liée à la décision serait constant. Une limite de cette étude relève de sa validité externe dans le sens où il est rare qu'un opérateur humain dans une situation de travail soit confronté à quatre niveaux de fiabilité différents en une heure de temps.

Le deuxième chapitre empirique a permis de confirmer qu'un système automatisé fiable à niveau intermédiaire (56.25%) n'était pas forcément plus avantageux qu'exécuter la tâche manuellement. Ce résultat vient consolider les études précédentes en se focalisant cette-fois-ci sur l'information (Onnasch, Wickens, Li & Manzey, 2014 ; Rovira, McGarry, & Parasuraman, 2007 ; Wickens & Dixon, 2007). Rovira, McGarry et Parasuraman (2007) ont montré qu'un système automatisé prenant en charge la fonction information fiable à 80% améliorait les performances par rapport à une condition manuelle, contrairement à une automatisation moins fiable (60%). Wickens et Dixon (2007) ont effectué une méta-analyse des études examinant la fiabilité de l'automatisation et les performances humaines et ont constaté que les performances humaines s'amélioraient avec la fiabilité du système automatisé et que celui avec un niveau de fiabilité inférieur à environ 70% n'était pas meilleure qu'exécuter la tâche manuellement. Cependant, nos résultats ont été trouvés sur l'implication visuelle et non sur la performance en comparant les périodes manuelles aux périodes suivantes sous système automatisé. Les niveaux intermédiaire (56.25% et 87.5%) n'ont pas montré d'effets des niveaux de fiabilité sur la performance et sur l'implication visuelle pour chaque niveau au cours du temps. On peut donc supposer que la fonction information a permis de maintenir de façon générale un niveau constant de performance et d'implication visuelle au cours du temps, sans l'améliorer, ni la dégrader. Ce résultat est à confronter à des études comme Parasuraman, Molloy et Singh (1993) ou encore Bagheri et Jamieson (2004b) dans lesquelles des niveaux constants de fiabilité ont généré une dégradation de la surveillance et de la performance. Pour le niveau 56.25%, le retour de l'opérateur après une session manuelle a nécessité plus d'implication visuelle chez les participants. La *récupération adaptative* (Farrell & Lewandowsky, 2000), c'est-à-dire, le temps après l'attribution temporaire du contrôle à l'opérateur pourrait être plus coûteuse avec un système automatisé prenant en charge la fonction de l'information. Cela souligne qu'implanter des systèmes automatisés adaptatifs n'a pas les mêmes conséquences sur les performances humaines selon le niveau de fiabilité du système et qu'avec

un niveau de 56.25% par exemple, l'opérateur a plus de quantité visuelle à fournir pour maintenir la surveillance du système lors de la mise en œuvre de la fonction information. Plusieurs études ont souligné la nécessité de trouver le meilleur moment d'une reprise en main par l'opérateur humain (Kaber & Endsley, 2004 ; Jones, 2007), voire même combien de temps cette reprise en main devait être implantée pour montrer des avantages (Hilburn, Parasuraman, & Mouloua, 1996) ou encore permettre à l'opérateur humain de reprendre pleinement conscience de la situation (Gold, Dambock, Lorenz, & Bengler, 2013 ; Vogelpohl, Kühn, Hummel, Gehlert, & Vollrath, 2018). Cependant ces études ne comparent pas différents niveaux de fiabilité du système automatisé, or nous avons montré que différents niveaux de fiabilité n'ont pas les mêmes conséquences sur le comportement humain. Nous pensons donc que de nouvelles études impliquant une reprise en main et différents niveaux d'automatisation seraient enrichissantes, notamment en intégrant d'autres mesures physiologiques permettant d'évaluer, par exemple, la charge mentale et l'effort nécessaire à l'opérateur lors de la transition manuel-système automatisé pour maintenir une surveillance adéquate du système automatisé (Prinzel, Pope, Freeman, Scerbo, & Mikulka, 2001 ; Prinzel et al., 2003).

Dans le domaine de la conduite automobile, beaucoup d'études se sont intéressées à la reprise en main temporaire par le conducteur durant la tâche de conduite afin de répondre à un imprévu. Plusieurs d'entre elles évoquent la nécessité d'une période suffisante de temps pour l'opérateur humain afin de reprendre conscience de la situation après une période automatisée et de pouvoir réagir à l'imprévu (Merah, Jamson, Lai, Dali, & Carsten, 2014; Vogelpohl et al. 2018). Beaucoup d'études ont été faites sur le *Take-over-request*, c'est-à-dire le temps qu'il est nécessaire de laisser au conducteur entre la demande de reprise en main par le véhicule et la reprise en main effective pour que l'opérateur humain puisse reprendre le contrôle de son véhicule et répondre comme il se doit à l'évènement. De notre côté, nous sommes intéressés à des défaillances multiples arrivant avant et après la période de reprise en main manuelle. Avec 56.25% de fiabilité avant la reprise en main, l'opérateur a investi autant

de temps dans la reprise en main manuelle qu'après ; mais pas avec des niveaux plus élevés. Cela veut dire que la période automatisée avant la période de reprise en main est importante selon le nombre de défaillances auxquels l'opérateur humain est confronté car elle a directement un effet sur ce qu'il se passe durant cette période de reprise en main. Les études sur le *take-over* en conduite automobile se focalisent souvent sur la reprise en main avant un évènement critique et le temps qu'il faut entre le moment où la requête reprise en main est faite et la reprise réel (Zhang et al., 2019). Nous pensons qu'il faudrait aussi s'intéresser aux potentielles défaillances qu'il y a eu pendant cette période automatisée, avant de reprendre la main et de voir si cela a un impact sur ce temps nécessaire avant la reprise en main effective.

Dans notre troisième article, nous avons décidé d'étudier la fiabilité de la fonction information à travers l'activité de planification. L'objectif était de voir si les planificateurs étaient capables de reconnaître des propositions fiables et moins fiables proposées par une aide à la planification mettant en œuvre la fonction information. L'application était que si les planificateurs considéraient des propositions moins fiables comme meilleures, cela supposait que s'ils étaient amenés à les utiliser, ils utiliseraient ces mêmes propositions moins fiables plutôt que les fiables. Cela aurait un impact sur la tournée effectuée sur le terrain : baisse de la satisfaction client car ceux-ci seraient livrés en retard, perte de rentabilité pour l'entreprise... Cependant, la mise en pratique d'une expérimentation sur le terrain s'est confrontée à différents problèmes : les difficultés techniques rencontrés le jour des expérimentations nous ont amené à modifier l'expérimentation et à concevoir un nouveau protocole en urgence dans un temps relativement court ce qui a parfois engendré un manque de recul dans les choix méthodologiques (quel scénario spécifique est présenté en même temps qu'un autre scénario...). Bien que les tâches expérimentales soient différentes entre la tâche spécifique du micro-monde *Multi Attribute Task Battery* et l'activité de planification, des parallèles sont envisageables entre les résultats des études. L'aide à la planification décrit dans le troisième article est présenté comme mise en œuvre avec la prise en charge de la fonction information.

Une telle aide pourrait être envisagée avec la mise en œuvre de la fonction décision en plus, comme présentée dans le premier article. Par exemple, l'aide à la planification pourrait choisir automatiquement des plannings, tout en délivrant à travers les indicateurs des informations sur ces mêmes plannings. Dans un tel cas, il est intéressant de comprendre comment les opérateurs se comportent face à deux fonctions de mise en œuvre du système automatisé qui interagissent en même temps. Notre première étude montre que les opérateurs ont tendance à se focaliser sur la fonction information et non sur la décision. Dans une mise en application, on pourrait imaginer une aide à la planification qui proposerait des indicateurs et qui validerait en même temps automatiquement des plannings. Imaginons qu'il y ait un problème de fiabilité avec les indicateurs et que ceux-ci soient moins fiables, l'opérateur aurait tendance à se fier à ceux-ci plus qu'à vérifier le choix automatique des plannings. Cette situation serait d'autant plus compliquée que la troisième expérience nous a montré que les planificateurs ont plus de difficultés à déceler des informations agrégées à travers des indicateurs fiables et moins fiables que des informations clairement présentées dans une tournée. Ces résultats soulignent donc la nécessité de travailler sur la transparence de la fiabilité des systèmes automatisés afin d'éviter des biais d'automatisation (Bagheri & Jamieson, 2004b ; Wang, Jamieson, & Hollands, 2009 ; Beller, Heesen, & Vollrath, 2013). En fournissant des informations telles que les incertitudes inhérentes à un système, la confiance des opérateurs peut être affectée et devenir plus appropriée (Lee & See, 2004 ; Hoff & Bashir, 2015). Si nous appliquons ce conseil à l'aide à la planification, la mise en place d'alarmes indiquant les informations n'ayant pas pu être prises en compte lors du calcul des indicateurs pourrait permettre cette meilleure calibration de la confiance et donc de la fiabilité perçue du système automatisé.

Dans une autre perspective et au vu de nos résultats dans la troisième étude, il serait intéressant de voir le comportement des opérateurs avec la manipulation en temps réel sur prototype de l'aide à la planification. En effet, dans cette dernière étude ceux-ci ont évalué des propositions effectuées par une aide automatisée mais ils n'ont pas manipulé eux-mêmes l'aide

à la planification par manque de moyen technique, le prototype de l'aide à la planification n'étant pas techniquement opérationnel au moment des expérimentations. La simulation d'une activité de planification avec le prototype aurait permis de garder une plus grande validité écologique (Hoc, 2001). Cette simulation par utilisation du prototype permettrait également de mettre en relation la qualité perçue des propositions et l'utilisation réelle faite par les opérateurs utilisant l'aide à la planification. Cela serait l'occasion de dépasser la simple évaluation de propositions du système automatisé en pouvant constater l'utilisation effective de celui-ci. Des méthodes telles que l'analyse des protocoles verbaux durant l'utilisation du prototype pourrait permettre de comprendre l'utilisation faite du système et les possibles difficultés rencontrées lors de son utilisation. Dans cette méthode, l'opérateur humain doit verbaliser toutes ses réflexions durant l'accomplissement de sa tâche (Ericsson & Simon, 1993). En parallèle, il serait également intéressant d'étudier plus précisément l'acceptabilité et l'acceptation d'une telle aide à la planification par le biais de questionnaires et d'entretiens. Selon Bobillier-Chaumon et Dubois (2009), la qualité, la fiabilité et la pertinence des informations (cf. fiabilité technique) jouent un rôle sur l'acceptation ou le rejet d'un système automatisé. Ce type de questionnaires pourrait permettre de mieux comprendre le comportement des exploitants lors de l'utilisation du prototype. Lorsque l'on se réfère à notre troisième étude, on peut voir qu'il est très rare que les exploitants aient donné la note maximum de 5 aux scénarios fiables présentés comme des propositions du système automatisé. Même si les exploitants n'avaient pas de défauts à donner aux scénarios, ils ne donnaient jamais la note maximale.

Réflexions générales

La technologie évolue constamment, entraînant des changements organisationnels de plus en plus fréquents dans nos activités. Ces changements sont aussi accompagnés de nouveaux processus cognitifs. A l'heure où ce document est écrit, des milliers d'ingénieurs, concepteurs ou encore techniciens travaillent et inventent les prochains systèmes automatisés

qui viendront partager notre quotidien dans les prochaines années. Le président Emmanuel Macron annonçait en 2019 à l'occasion du mondial de l'automobile, une avancée dans la réglementation française afin de permettre la mise en circulation de voitures autonomes d'ici 2022 ; l'évolution de la domotique connectée dans nos intérieurs ; les projets de livraison par drones menés par de grands groupes multinationales ; le développement de lunettes dans les hôpitaux permettant aux chirurgiens d'opérer en réalité augmentée avec des hologrammes se superposant au champ de vision pendant les opérations et permettant de donner des informations... Les exemples sont nombreux pour témoigner de la nécessité de travaux en ergonomie sur l'interaction homme-machine. Les questions qui en découlent sont nombreuses : quels sont les processus psychologiques mises en jeu dans l'interaction avec ces systèmes ? Quels sont les impacts de ces systèmes sur l'activité des opérateurs ? Quels sont les déterminants qui favorisent l'usage réel et l'acceptation de ces systèmes ? Quels sont les difficultés engendrées par l'utilisation de ces systèmes ? ...

L'arrivée des voitures autonomes n'est pas un exemple anodin. Elle aura des conséquences sur les passagers du véhicule mais aussi sur les autres conducteurs. On peut supposer par exemple, que les autres conducteurs auront leur propre fiabilité perçue de ces voitures et que cela pourrait engendrer des effets sur l'acceptation de ces véhicules. Par exemple, prendre des risques en voulant doubler une voiture autonome pour ne pas rester derrière car celle-ci est considérée comme ayant un comportement imprévisible et faucher un piéton qui traversait. Les comportements qui s'en suivront seront donc importants à comprendre pour la sécurité de tous.

Références

- Alonso, V., & De La Puente, P. (2018). System transparency in shared autonomy: A mini review. *Frontiers in neurorobotics*, 12, 83.
- Bailey, N. R., & Scerbo, M. W. (2007). Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science*, 8(4), 321-348.

- Bailey, N. R., Scerbo, M. W., Freeman, F. G., Mikulka, P. J., & Scott, L. A. (2006). Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. *Human factors*, 48(4), 693-709.
- Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19, 775-779. [https://doi.org/10.1016/0005-1098\(83\)90046-8](https://doi.org/10.1016/0005-1098(83)90046-8)
- Baldwin, J. R., & Sabourin, D. (2004). *Effet de l'évolution de l'utilisation des technologies sur le rendement des établissements dans le secteur de la fabrication au Canada* (Report No11F0027MIF). Division de l'analyse microéconomique. <https://www150.statcan.gc.ca/n1/pub/11f0027m/11f0027m2004020-fra.pdf>
- Bellet, T., Hoc, J.M., Boverie, S., & Boy, G.A. (2011). From human-machine interaction to cooperation: Towards the integrated copilot. *Human-Computer Interaction in Transport*, Ashgate, 129-156.
- Billings, C. E. (1991). *Human-centered aircraft automation: A Human Factors concept and guidelines* (Report No103885). National Aeronautics and Space Administration Ames Research Center. <https://ntrs.nasa.gov/citations/19910022821>
- Billings, C. E., Lauber, J. K., Funkhouser, H., Lyman, G., & Huff, E. M. (1976). *Aviation Safety Reporting System* (Report NoTM-X-3445). National Aeronautics and Space Administration Ames Research Center. <https://ntrs.nasa.gov/citations/19760026757>
- Billings, C. E., & Woods, D. D. (1994). Concerns about adaptive automation in aviation systems. In *Human performance in automated systems: Current research and trends* (pp. 24-29). Hillsdale, NJ: Erlbaum.
- Bobillier-Chaumon, M. E., & Dubois, M. (2009). L'adoption des technologies en situation professionnelle: quelles articulations possibles entre acceptabilité et acceptation?. *Le travail humain*, 72(4), 355-382.
- Boukerma, Z. (1998). Notions d'ergonomie et de fiabilité. *Journal de la médecine du travail*. 1(1). 24-27.
- Brangier, E., & Robert, J. M. (2010, September). Manifeste pour l'ergonomie prospective: Anticiper de futures activités humaines en vue de concevoir de nouveaux artefacts. In *Proceedings of the 22nd Conference on l'Interaction Homme-Machine* (pp. 57-64).
- Burns, C. M., Skraaning Jr, G., Jamieson, G. A., Lau, N., Kwok, J., Welch, R., & Andresen, G. (2008). Evaluation of ecological interface design for nuclear process control: situation awareness effects. *Human factors*, 50(4), 663-679.

- Calhoun, G. L., Ruff, H. A., Draper, M. H., & Wright, E. J. (2011). Automation-level transference effects in simulated multiple unmanned aerial vehicle control. *Journal of Cognitive Engineering and Decision Making*, 5(1), 55-82.
- Cegarra, J. (2004). *La gestion de la complexité dans la planification : le cas de l'ordonnancement* (Doctoral dissertation, University Paris 8).
- Cegarra, J. (2008). A cognitive typology of scheduling situations: A contribution to laboratory and field studies. *Theoretical Issues in Ergonomics Science*, 9(3), 201-222.
- Cegarra, J. (2012). *De la gestion de la complexité à son assistance: Contributions en psychologie ergonomique*. (Unpublished manuscript, Habilitation à diriger des recherches, Toulouse II-Le Mirail University)
- Cegarra, J., Valéry, B., Avril, E., Calmettes, C., & Navarro, J. (2020). OpenMATB: A Multi-Attribute Task Battery promoting task customization, software extensibility and experiment replicability. *Behavior Research Methods*, 1-11.
- Cegarra, J., & van Wezel, W. (2012). Revisiting decision support systems for cognitive readiness: A contribution to unstructured and complex scheduling situations. *Journal of Cognitive Engineering and Decision Making*, 6(3), 299-324.
- Chen, J. Y., Procci, K., Boyce, M., Wright, J., Garcia, A., & Barnes, M. (2014). *Situation awareness-based agent transparency* (Report NoARL-TR-6905). U.S. Army research laboratory. <https://apps.dtic.mil/dtic/tr/fulltext/u2/a600351.pdf>
- Crocoll, W. M., & Coury, B. G. (1990). Status or recommendation: Selecting the type of information for decision aiding. In *Proceedings of the human factors society annual meeting* (Vol. 34, No. 19, pp. 1524-1528). Sage CA: Los Angeles.
- Dekker, S. W. A., & Woods, D. D. (2002). MABA-MABA or Abracadabra? Progress on Human–Automation Co-ordination. *Cognition, Technology & Work*, 4(4), 240-244. <https://doi.org/10.1007/s101110200022>
- De Visser, E., & Parasuraman, R. (2011). Adaptive aiding of human-robot teaming: Effects of imperfect automation on performance, trust, and workload. *Journal of Cognitive Engineering and Decision Making*, 5(2), 209-231.
- De Vries, P., & Midden, C. (2008). Effect of indirect information on system trust and control allocation. *Behaviour & information technology*, 27(1), 17-29.
- Dzindolet, M. T., Pierce, L. G., Beck, H. P., Dawe, L. A., & Anderson, B. W. (2001). Predicting misuse and disuse of combat identification systems. *Military Psychology*, 13(3), 147-164.

- Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International journal of human-computer studies*, 58(6), 697-718.
- Endsley, M. R. (1996). Automation and situation awareness. *Automation and human performance: Theory and applications*, 20, 163-181.
- Endsley, M. R. (2018). Level of automation forms a key aspect of autonomy design. *Journal of Cognitive Engineering and Decision Making*, 12(1), 29-34.
- Endsley, M. R., & Kaber, E. (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*, 37(2), 381-394.
- Ericsson, K.A. & Simon, H.A. (1993). Protocol analysis: Verbal reports as data. Cambridge, MIT Press.
- Falkland, E. C., & Wiggins, M. W. (2019). Cross-task cue utilisation and situational awareness in simulated air traffic control. *Applied ergonomics*, 74, 24-30.
- Fitts, P.M. (1951). *Human engineering for an effective air navigation and traffic control system*. Washington, DC: National Research Council.
- Fuld, R.B. (2000). The fiction of function allocation, revisited. *International Journal of Human Computer Studies*, 52(2), 217-233.
- Galster, S. M., Bolia, R. S., & Parasuraman, R. (2002). Effects of information automation and decision-aiding cueing on action implementation in a visual search task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 46, No. 3, pp. 438-442). Sage CA: Los Angeles.
- Gandil, P. (2018). *Rapport sur la sécurité aérienne 2018* (Report No2018). Direction générale de l'Aviation civile. <https://www.ecologique-solidaire.gouv.fr/rapports-annuels-sur-securite-aerienne>
- Gary, K., Uzsoy, R., Smith, S. P., & Kempf, K. (1995). Measuring the quality of manufacturing schedules. In D.E Brown & W.T. Scherer (Eds.), *Intelligent scheduling systems* (pp. 129-154). Boston, MA: Kluwer Academy.
- Govaere, V. & Wioland, L. (2018a). Propagir : une démarche de prévention pour les organisations en réseaux de logistique et de transport. *Hygiène et sécurité du travail*, 252.
- Govaere, V., & Wioland, L. (2018b). Modélisation des contraintes et des compromis spécifiques au domaine du transport routier de marchandises (TRM). Livrable 2 du projet ANR Smart Planning ANR-16-CE26-0017.

- Govaere, V., Wioland, L., Cegarra, J., Gourc, D., & Clément, A. (2018). Smart Planning- Approaching the Characteristics of a Valid, Balanced Transport Round. In *Congress of the International Ergonomics Association* (pp. 387-396). Springer, Cham.
- Hainley Jr, C. J., Duda, K. R., Oman, C. M., & Natapoff, A. (2013). Pilot performance, workload, and situation awareness during lunar landing mode transitions. *Journal of Spacecraft and Rockets*, 50(4), 793-801.
- Hilburn, B., Molloy, R., Wong, D., & Parasuraman, R. (1993). Operator versus computer control of adaptive automation. In *Proceedings of the Seventh International Symposium on Aviation Psychology*, 161-166.
- Hoc, J. M. (1993). Conditions méthodologiques d'une recherche fondamentale en psychologie ergonomique et validité des résultats. *Le travail humain*, 171-184.
- 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. & Debernard, S. (2002). Respective demands of task and function allocation on human-machine co-operation design: a psychological approach. *Connection science*, 14(4), 283-295.
- Hoc, J.M., Young, M. S., & Blosseville, J.-M. (2009). Cooperation between drivers and automation : Implications for safety. *Theoretical Issues in Ergonomics Science*, 10(2), 135-160.
- Hoc, J.M., & Chauvin, C. (2011). *Cooperative implications of the allocation of functions to humans and machines*. IRCCyN, Nantes, France.
- Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors*, 57(3), 407-434.
- Hollnagel, E. (1999). From function allocation to function congruence. In S.A. Dekker & E. Hollnagel (Eds.), *Coping with computers in the cockpit* (pp29-53). Ashgate: London.
- Jipp, M., & Ackerman, P. L. (2016). The impact of higher levels of automation on performance and situation awareness: a function of information-processing ability and working-memory capacity. *Journal of Cognitive Engineering and Decision Making*, 10(2), 138-166.
- Jones, L. M. (2007). Effect of Repeated Function Allocation and Reliability on Automation-Induced Monitoring Inefficiency (Doctoral dissertation, University of Central Florida).
- Jones, P. M., & Mitchell, C. M. (1994). Model-based communicative acts: human-computer collaboration in supervisory control. *International Journal of Human-Computer Studies*, 41(4), 527-551.

- Jüngen, F. J., & Kowalczyk, W. (1995). An intelligent interactive project management support system. *European Journal of Operational Research*, 84(1), 60-81.
- 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.
- Kerr, R. M. (1992). Expert systems in production scheduling: Lessons from a failed implementation. *Journal of Systems and Software*, 19(2), 123-130.
- Khademi, K. (2016). *Les processus cognitifs dans les activités d'ordonnement en environnement incertain*. (Doctoral dissertation, Toulouse Jean Jaures II University).
- Khademi, K., Cegarra, J., Govaere, V., & Wioland, L. (2015). *La planification par l'exploitant dans le transport routier de marchandises: méthode d'analyse et perspectives d'application inter-domaines. Partie 2*. (Report). Institut National de Recherche et de Sécurité (INRS). <https://hal-lara.archives-ouvertes.fr/hal-01428640/document>
- Layton, C., Smith, P. J., & McCoy, C. E. (1994). Design of a cooperative problem-solving system for en-route flight planning: An empirical evaluation. *Human Factors*, 36(1), 94-119.
- 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*, 46(1), 50-80.
- Li, H., Wickens, C. D., Sarter, N., & Sebok, A. (2014). Stages and levels of automation in support of space teleoperations. *Human Factors*, 56(6), 1050-1061.
- Lorenz, B., Di Nocera, F., Röttger, S., & Parasuraman, R. (2001). The effects of level of automation on the out-of-the-loop unfamiliarity in a complex dynamic fault-management task during simulated spaceflight operations. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 45(2), 44–48.
- Madhavan, P., Wiegmann, D. A., & Lacson, F. C. (2006). Automation failures on tasks easily performed by operators undermine trust in automated aids. *Human factors*, 48(2), 241-256.
- Marsh, S., & Dibben, M. R. (2003). The role of trust in information science and technology. *Annual Review of Information Science and Technology (ARIST)*, 37, 465-98.

- 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
- Mouloua, M., & Hancock, P. A. (2019). *Human Performance in Automated and Autonomous Systems: Emerging Issues and Practical Perspectives*. CRC Press.
- Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of experimental psychology: Applied*, 6(1), 44.
- Navarro, J. (2018). *De l'analyse de l'activité de conduite automobile à son assistance* (Unpublished manuscript, habilitation à diriger des recherches, Université Lumière Lyon 2)
- Navarro, J. (2017). Human–Machine interaction theories and lane departure warnings. *Theoretical issues in ergonomics science*, 18(6), 519–547.
- Navarro, J., Heuveline, L., Avril, E., & Cegarra, J. (2018). Influence of human-machine interactions and task demand on automation selection and use. *Ergonomics*, 61(12), 1601-1612.
- Navarro, J., & Osiurak, F. (2015). When do we use automatic tools rather than doing a task manually? Influence of automatic tool speed. *The American journal of psychology*, 128(1), 77–88.
- Nyssen, A. S., Aunac, S. Faymonville, M.E., & Lutte, I. (2004). Reporting systems in healthcare from a case-by-case experience to a general framework: an example in anaesthesia. *European Journal of Anaesthesiology* (EJA), 21(10), 757-765.
- Oakley, B., Mouloua, M., & Hancock, P. (2003). Effects of automation reliability on human monitoring performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 1(47), 188-190.
- Onnasch, L. (2015). Crossing the boundaries of automation—Function allocation and reliability. *International Journal of Human-Computer Studies*, 76, 12-21. <https://doi.org/10.1016/j.ijhcs.2014.12.004>
- Onnasch, L., Ruff, S., & Manzey, D. (2014). Operators' adaptation to imperfect automation – Impact of miss-prone alarm systems on attention allocation and performance. *International Journal of Human-Computer Studies*, 72(10), 772-782. <https://doi.org/10.1016/j.ijhcs.2014.05.001>
- Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human factors*, 56(3), 476-488.

- Parasuraman, R., & Riley, V. (1997). Humans and automation : Use, misuse, disuse, abuse. *Human factors*, 39(2), 230-253.
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation : An attentional integration. *Human factors*, 52(3), 381-410.
- 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., & Hilburn, B. (1993). Adaptive function allocation reduces performance cost of static automation. *Proceedings of 7th international symposium on aviation psychology*, 37-42.
- 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*, 50(3), 511-520.
- Parasuraman, R., & Wilson, G. F. (2008). Putting the brain to work : Neuroergonomics past, present, and future. *Human factors*, 50(3), 468-474.
- Payre, W. (2015). *Conduite complètement automatisée : Acceptabilité, confiance et apprentissage de la reprise de contrôle manuel* (Doctoral dissertation, Paris 8 University). <https://www.theses.fr/2015PA080115>
- Rigas, G., Carling, E., & Brehmer, B. (2002). Reliability and validity of performance measures in microworlds. *Intelligence*, 30(5), 463-480
- Riley, J. M., Endsley, M. R., Bolstad, C. A., & Cuevas, H. M. (2006). Collaborative planning and situation awareness in Army command and control. *Ergonomics*, 49(12-13), 1139-1153.
- Rovira, E., Cross, A., Leitch, E., & Bonaceto, C. (2014). Displaying contextual information reduces the costs of imperfect decision automation in rapid retasking of ISR assets. *Human factors*, 56(6), 1036-1049.
- Rovira, E., McGarry, K., & Parasuraman, R. (2007). Effects of imperfect automation on decision making in a simulated command and control task. *Human factors*, 49(1), 76-87.
- Rovira, E., Zinni, M., & Parasuraman, R. (2002). Effects of information and decision automation on multi-task performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 46(3), 327-331.

- Sanchez, J., Rogers, W. A., Fisk, A. D., & Rovira, E. (2014). Understanding reliance on automation: effects of error type, error distribution, age and experience. *Theoretical issues in ergonomics science*, *15*(2), 134-160.
- Santiago-Espada, Y., Myer, R. R., Latorella, K. A., & Comstock Jr, 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., & Schroeder, B. (2001). Supporting decision making and action selection under time pressure and uncertainty: The case of in-flight icing. *Human factors*, *43*(4), 573-583.
- Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. *Handbook of human factors and ergonomics*, *2*, 1926-1943.
- Seong, Y., & Bisantz, A. M. (2008). The impact of cognitive feedback on judgment performance and trust with decision aids. *International Journal of Industrial Ergonomics*, *38*(7-8), 608-625.
- Singh, A. L., Tiwari, T., & Singh, I. L. (2009). Effects of automation reliability and training on automation-induced complacency and perceived mental workload. *Journal of the Indian Academy of Applied Psychology*, *35*(2009), 9-22.
- Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab.
- Sheridan, T. B., & Parasuraman, R. (2005). *Human-Automation Interaction. Reviews of Human Factors and Ergonomics*, *1*(1), 89-129. doi :10.1518/155723405783703082
- Skrypchuk, L., Langdon, P., Sawyer, B. D., Mouzakitis, A., & Clarkson, P. J. (2019). Enabling multitasking by designing for situation awareness within the vehicle environment. *Theoretical issues in ergonomics science*, *20*(2), 105-128
- Strauch, B. (2017). *Investigating human error : Incidents, accidents, and complex systems*. CRC Press.
- Tatineni, V. & Demetsky, M. (2005). *Supply chain models for freight transportation planning*. Research report UVACTS-14-0-85, University of Virginia, USA.
- Vagia, M., Transeth, A. A., & Fjerdingen, S. A. (2016). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? *Applied ergonomics*, *53*, 190-202.
- Van Wezel, W., Cegarra, J., & Hoc, J.-M. (2010). *Allocating Functions to Human and Algorithm in Scheduling. Behavioral Operations in Planning and Scheduling*, 339–370. doi:10.1007/978-3-642-13382-4_14

- Van Wezel, W., & Jorna, R. (2009). Cognition, tasks and planning: supporting the planning of shunting operations at the Netherlands Railways. *Cognition, Technology & Work*, 11(2), 165-176.
- Vogelpohl, T., Kühn, M., Hummel, T., Gehlert, T., & Vollrath, M. (2018). Transitioning to manual driving requires additional time after automation deactivation. *Transportation research part F: traffic psychology and behaviour*, 55, 464-482.
- Wang, L., Jamieson, G. A., & Hollands, J. G. (2009). Trust and reliance on an automated combat identification system. *Human factors*, 51(3), 281-291.
- Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation : A synthesis of the literature. *Theoretical Issues in Ergonomics Science*, 8(3), 201-212.
- Wickens, C.D., Hollands, J., Banbury, S. & Parasuraman, R. (2013). *Engineering psychology and human performance* (4th ed.). Upper Saddle River, NJ: Pearson.
- Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomics Science*, 2(4), 352-367.
- Wiener, Earl L., & Curry, R. E. (1980). Flight-deck automation : Promises and problems. *Ergonomics*, 23(10), 995–1011.
- Wiener, E. L. (1981). Complacency: is the term useful for air safety? *Proceedings of the 26th corporate aviation safety seminar*, 116–125.
- Wioland, L. (2013) Ergonomic Analyses in the Transport and Logistics Sector. Reflection on developing a new Prevention Approach: “Act Elsewhere”. *Accident Analysis & Prevention*, 59, 213-220
- Wioland, L., & Govaere, V. (2014). Comparative Study of Two Organisations Implementing Information and Communication Technology: Effects on the Health of Goods Transport Operators. *Journal of Traffic and Logistics Engineering Vol*, 2(1).
- Yuviler-Gavish, N., & Gopher, D. (2011). Effect of descriptive information and experience on automation reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53(3), 230-244.
- Zhang, B., de Winter, J., Varotto, S., Happee, R., & Martens, M. (2019). Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transportation research part F: traffic psychology and behaviour*, 64, 285-307.